PHILOLOGICA 1/2017

Editor
LIBUŠE DUŠKOVÁ

Editor: prof. PhDr. Libuše Dušková, DrSc.

CONTENTS

CODING ANGER IN FICTION

ALEŠ KLÉGR, PAVLÍNA ŠALDOVÁ

ABSTRACT

The paper examines the fictional representation of anger in English. The emotion is identified by its lexical indication in the reporting clause (*said angrily*) accompanied by the direct speech which is assumed to verbalize the emotion. The reporting clause and the direct speech are viewed as the components of an anger-coding unit which is examined in terms of their mutual position and the syntactic and speech-act properties of the direct speech component. Analysis and correlation of these variables (and the length of the direct speech) suggest that anger-coding units display recognizable patterns with characteristic structure and illocutionary profile.

Keywords: anger-coding patterns, English, fiction, direct speech, syntactic features, speech act analysis

1. Introduction

Emotions in their various guises and reflections in language have attracted a great deal of attention and been the subject of many studies from a number of perspectives, including their categorization and crosslinguistic differences (Wierzbicka, 1999; Harkins and Wierzbicka, 2001; Mikolajczuk, 2004; Pütz and Neff-van Aertselaer, 2008; Goddard and Ye, 2016), from lexical, grammatical and contrastive points of view (Trosborg. 1995; Weigand, 1998; Durst, 2001; Klégr, 2005, 2007; Wiklund, 2009; Zhang, 2014), a diachronic perspective (Geeraerts, Gevaert, and Speelman, 2012), in discourse (Glynn, 2014; Blumenthal, Novakova and Siepmann, 2014) or from a gender angle (Galasinski, 2004). Some of them deal specifically with anger.

The question of how an emotion gets expressed by linguistic means is an intriguing one. For the purposes of this paper a very distinct type of emotion, anger, was chosen and a seemingly easy way to explore it: in fiction where it is vocalized by direct speech (stylized dialogue). When we are angry it is manifested by our mien, voice or gestures and we articulate the reasons for being angry by what we say. The problem is that the written language does not have extralinguistic and prosodic means of expression at its disposal. Hence the symptoms of anger must be graphically described or the writer may simply state that the fictional character is angry. The study of anger in language may thus examine how the phenomenal aspect of anger is verbally described or it may examine the

verbal articulation of (the reasons for) anger by a given person, which in fiction is done through direct speech. The present study attempts the latter while recognizing that it is the writer's stylized literary approximation of what people actually say.

2. Anger-coding units: definition and data collection

The selection of angry utterances for analysis had to deal with two problems, their location in text and finding criteria for their description. The problem of locating an angry utterance was resolved by looking for the sequence "said angrily" in a corpus (other adverbs, such as "furiously", were excluded for the sake of homogeneity and a manageable size of the sample). This led to instances of direct speech that due to this sequence could be safely considered "angry utterances" as conceptualized by the authors. Since both the reporting clause including the lexical anger marker ("said angrily") and the direct speech (the angry utterance proper) are part and parcel of coding anger in fiction, they were both included in the analysis and treated as one unit.

The investigation thus focused on anger-coding units made up of two components: a reporting clause ("rc" for short) containing the sequence *said angrily* and a direct speech component which presumably expresses the speaker's anger. Both the reporting clause and the direct speech are clearly distinguishable from the previous and the subsequent text and relatively self-contained. The hypothesis was that direct speech components conveying the speaker's anger will primarily include (a) emotion-charged expressive illocutions, and (b) sentence types likewise associated with emotions (exclamative and imperative sentences).

The source of the material for analysis was William Fletcher's database PIE (Phrases in English) based on the British National Corpus. After weeding out hits that did not conform to the specification – the sequence *said angrily* must be part of a reporting clause introducing direct speech – the sample included 152 anger-coding units.

3. The position of the reporting clause
and the length of the direct speech

On closer examination it turned out that the anger-coding units are characterized by two basic interrelated features which distinguish between them: the position of the reporting clause and the length of the direct speech. The reporting clauses with *said angrily* occur in all three possible positions (rcp): before direct speech (rcp1), inside direct speech, bisecting it into two parts (rcp2), and after the direct speech (rcp3):

initial (rcp1): Lambert said angrily. "Not in this squadron, never."
medial (rcp2): "It's not a proposal," she said angrily. "It's – it's an attempt to buy me."
final (rcp3): "Come on, tell me!" Gog said angrily.

Table 1 shows that anger-coding units with mid-position reporting clauses are the commonest, occurring in almost 60%; those with initial reporting clauses are the least frequent (12.5%), and those with final reporting clauses are somewhere in between. Table 1 also indicates that the direct speech component (utterance) may consist of

more than one sentence, especially with rcp2 utterances in which one sentence typically precedes and another sentence follows the reporting clause (rather than there being just one sentence divided by the reporting clause). The last column in the table shows the difference between the number of utterances and the number of sentences the utterances consist of.

Table 1. The reporting clause position and direct speech length figures (* i.e., one sentence per utterance ratio or one sentence per rcp2 utterance part 1 or 2 ratio)

position / Σ	utterances	sentences within utterances	sentences in excess of 1/1 ratio*
rcp1	19 12.5%	22	3 15.8%
rcp3	45 29.6%	53	8 17.7%
rcp2 (part 1, 2)	88 57.9%	89/94	1/6 1.1%/6.8%
total	152 100.0%	258	

There seems to be a direct relation between the reporting clause position and the total length of the direct speech component expressed in the number of sentences. In the initial and final reporting clause utterances the direct speech component is largely composed of a single sentence. In both rcp1 and rcp3 utterances more than one sentence occurred in only 15.8% and 17.7% respectively, although there is no obvious reason why it could not be otherwise. The rcp2 utterances are bisected into two separate parts which again tend to consist of only one sentence just like rcp1 and recp3 utterances (part 1 departs from this tendency in 1.1%, part 2 in 6.8%). Hence the number of sentences in anger-coding units with rcp2 is more than twice the number of the rcp2 utterances (88 utterances, 183 sentences). In the few cases of rcp2 utterances where it was not clear whether part 1 and part 2 should be seen as two sentences or two clauses of one sentence, the decision was made by reference to illocution. Where clauses in what was formally one compound sentence had distinctly different force, they were treated as separate sentences.

Since the position of the reporting clause appears to correlate with the length of the direct speech, and the length of the direct speech component seems to be connected with its pragmatic makeup, as we shall see, the reporting clause position is apparently not a negligible factor. So, even at this point, rcp1, rcp2 and rcp3 anger-coding units can be seen as different anger presentation patterns.

4. Syntactic analysis of the direct speech component

In the next step, the analysis concentrated on the key component of the anger-coding units, direct speech. It was subject first to syntactic analysis, then to the pragmatic analysis of the speech acts (illocutions) that realize it, and finally to the analysis focused on the correlation between the syntax and illocutions.

Syntactic analysis looks at the distribution of the basic sentence types in the direct speech components. The hypothesis was that they would be typically realized by sentence types with interactive and emotion-prone functions, i.e. exclamative and imper-

ative sentences, rather than by the primarily neutral declarative sentences. It turned out to be convenient to distinguish one specific formal type in addition to imperative, interrogative and declarative sentences, namely sentences (called here "other") consisting of verbless, non-finite and formulaic structures, characteristic of the spoken language. Their functions range from expletives to exclamations, and they are generally emotion-laden.

Sentence type examples in the direct speech of anger-coding units:
- **Declarative**: "It's not a proposal," she said angrily.
- **Imperative**: He said angrily, dismissively: "Don't be stupid." / "Shut up!" said Ray Shepherd angrily.
- **Interrogative**: "What do you think I am, anyway?" Gazzer said, angrily / … he said angrily. "Did you expect a present, Miss Eyre?"
- **Other**: "Of course not!" she said angrily, reddening. / "For God's sake," I said angrily, / "Tomorrow?" Gog said angrily. / "Silly young fool," Dickie said angrily.

Table 2. Distribution of the sentence types

sentence type		rcp1	rcp3	rcp2	total	%
declarative		10	27	121	158	
	%	45.5	50.9	66.1		**61.2**
other (verbless. non-finite, etc.)		3	11	26	40	
	%	13.6	20.8	14.2		**15.5**
imperative		4	9	21	34	
	%	18.2	17.2	11.5		**13.2**
interrogative		5	6	15	26	
	%	22.7	11.3	8.2		**10.1**
total		22	53	183	258	**100.0**

However, quite contrary to expectations, the predominant sentence type in the whole sample is, as Table 2 shows, the declarative sentence (61.2%), while the remaining three types, the category other, imperative sentences and interrogative sentences are roughly equally represented and form slightly more than one third. Interestingly, this ratio appears to be connected with the length of the direct speech component and the reporting-clause position. In the two-part/two-sentence angry utterances with mid-position reporting clauses the dominance of declarative sentences is very strong (two thirds), while in the initial reporting clause utterances this ratio is reversed, the directly interactive types, interrogative, imperative and other sentences, prevail here over the declarative sentences.

In the rcp3 utterances the interactive types and the declarative sentences are balanced. The findings suggest that the shorter the direct speech component (with rc positions 1 and 3), the greater the tendency to use sentence types associated with direct, emotion-prone impact. Still, in all three rcp patterns of angry utterances the proportion of supposedly neutral declarative sentences is very high.

5. Pragmatic analysis of the direct speech component

The most interesting, but at the same time, the most difficult part was the pragmatic analysis striving to describe the direct speech component in terms of speech acts, i.e. its illocutionary force. Searle's (1979) taxonomy of speech acts, which provided the basic guidelines, is unfortunately very general and, as far as we know, there are very few detailed taxonomies of illocutions that could be adapted for our purposes. Ronan's (2015) interesting categorization of expressive speech acts which as she claims are particularly underresearched partly overlaps with the categories used in this research.

Examination of the sentences in the direct speech components of the sample revealed as many as 26 distinct secondary illocutionary forces. In order to bring the number to a manageable level while retaining useful semantic distinctions, the 26 illocutionary concepts were subsumed under nine categories that correspond to the following four types of the Searlian classification: (i) *representatives* – (1) belief, (2) explanation; (ii) *directives* – (3) direct coercion, (4) indirect coercion, (5) query; (iii) *commissives*: – (6) self-coercion (promise), (7) disagreement; (iv) *expressives* – (8) expressivity, (9) dissatisfaction. The **belief** category comprises illocutions such as assertion, claim, confirmation and conviction, which become anger-filled in circumstances when the speaker is questioned, doubted or contradicted. The category **explanation** appears – with only four exceptions – in the second part of the bisected rcp2 utterances. Explanation serves as a complement to some of the other illocutions as will be shown later. The three **coercion** categories describe cases when the speakers emphatically enforce their will either directly (order, demand, urging, forbidding) or indirectly (advice, warning, threat) or this enforcement is self-imposed as a promise. The association of the category **query** with anger is context-dependent and marginal. The category **disagreement** always involves some kind of negation, i.e., denial, refusal, rejection, or disagreement proper. The **expressivity** category covers all kinds of short utterances realized by verbless or non-finite clauses or formula-type structures. Their shortness makes the precise specification of the illocution difficult, their purpose is mainly and only to express a strong negative emotion. Comparison of this category with the other illocutions suggests that the computation of the specific nature of an illocution may depend on the length of the utterance – the longer the utterance, the easier it may be to identify it, and vice versa. **Dissatisfaction** subsumes complaints, accusations, criticisms, as well as reproaches and irony, in which the expression of an emotional load seems to be the main point.

The list of illocution categories in the sample with examples:

(1) **belief:** claim, assertion, conviction, confirmation – (assertion) "It was an accident," Monsieur Armand said angrily. / (confirmation) "Of course I am," said Hatch angrily. / (claim) "I'm twenty," she said angrily.
(2) **explanation:** explanation – Anna said angrily. "Because she didn't do it."
(3) **direct coercion:** order, demand, urging, forbidding – (order) "Get back!" he said angrily. / (demand) "We have to try, mother!" said Grace angrily. / (urging) "Come on, you dogs!" he said angrily. "Pull!" / (forbidding) Dickie said angrily. "That's not allowed."
(4) **indirect coercion:** advice, warning, threat – (advice) "Well if you see the bugger, run him off!" McIllvanney said angrily, / (warning) "You know what happened to Dai-

sy…" Lily said angrily, / (threat) "The next time they come, I'm going to kill them," I said angrily.

(5) **query:** "What d'you mean peculiar?" said Jinny angrily.

(6) **self-coercion:** promise – (promise) "There is no man or devil who will stop me from going to the home of my family," said Sir Henry angrily. / "I'll chuck them out," said Ricky angrily.

(7) **disagreement:** refusal, rejection, denial, disagreement – (refusal) "I don't want to know," **said** the chaplain angrily. / (denial) "I'm *not chicken*," said Gedanken angrily.

(8) **expressivity:** annoyance, disbelief, surprise – (annoyance) "Silly young fool," Dickie said angrily. / (disbelief) "Oh no?" said the bird angrily. / (surprise) "Tomorrow?" Gog said angrily.

(9) **dissatisfaction:** (i) complaint, accusation, criticism; (ii) reproach, irony – (complaint) "Typical!" said Rose angrily. / (accusation), "You've put me in an impossible position," Modigliani said angrily. / (criticism) "That is typical of a man," Mrs Theobald said angrily. / (irony) "Nice of you to tell me after I get back,…," he said angrily.

6. The distribution of illocutionary categories

The overall distribution of illocutionary categories is given in Table 3. The most frequent illocution, accounting for a quarter of all instances, is dissatisfaction which can be thus viewed as the main cause of anger. The second most frequent illocution in the sample is remarkably explanation which in itself does not imply anger. Obviously the function of explanation is quite different from that of the other illocutions: it typically accompanies another illocution which is related to anger. The next three most frequent types of illocution associated with anger are direct coercion, disagreement and belief, closely followed by the expressivity category. The remaining three categories are marginal.

Table 3. Overall distribution of illocutions in the reported speech

type of illocution	rcp1	rcp3	rcp2		total
			1st-part	2nd-part	
dissatisfaction	8 36.4	13 24.5	30 33.7	14 14.9	65 25.2
explanation	– –	3 5.7	1 1.1	40 42.6	44 17.1
direct coercion	2 9.1	13 24.5	13 14.6	13 13.8	41 15.9
disagreement	8 36.4	5 9.4	16 18.0	5 5.3	34 13.2
belief	1 4.5	5 9.4	13 14.6	8 8.5	27 10.5
expressivity	2 9.1	6 11.3	9 10.1	3 3.2	20 7.8
indirect coercion	– –	4 7.5	5 5.6	3 3.2	12 4.6
self-coercion	– –	3 5.7	2 2.2	6 6.4	11 4.2
query	1 4.5	1 1.9	– –	2 2.1	4 1.5
	22 100.0	**53** 100.0	**89** 100.0	**94** 100.0	**258** 100.0

Comparison of the distribution of illocutions within the three presentation models, rcp1, rcp2 and rcp3 (see Table 4), indicates that the frequency ranking of the three most frequent illocutions is very similar in utterances with undivided one-sentence direct speech components: dissatisfaction in the first position, expressivity in the second and direct coercion alternating in the first and second position. The only difference is disagreement in the first place in rcp1 utterances which comes only third in rcp3 utterances. Table 4 also shows that relatively close to this frequency ranking of illocutions is that of the first part of the rcp2 utterances, with dissatisfaction and disagreement at the top. By contrast the distribution in the second part of these utterances is very different: here explanation is the most frequent illocution of all.

Table 4. Illocutions in the first four positions according to frequency

RCP / FORCE	1	2	3	4
rcp1 utterances	dissatisfaction, disagreement	direct coercion, expressivity	belief, query	
rcp3 utterances	dissatisfaction, direct coercion	expressivity	belief, disagreement	indirect coercion
1-rcp2 utterances	dissatisfaction	disagreement	direct coercion, belief	expressivity
2-rcp2 utterances	explanation	dissatisfaction	direct coercion	belief

This fact led us to correlate the illocutions of the first and the second part of the medial rcp2 utterances. The results of the correlation, given in Table 5, revealed interesting patterns of pairing between the first part illocutions and the second part illocutions: most frequently it is dissatisfaction or disagreement which is followed by explanation, less frequently it is direct coercion, belief or expressivity which is subsequently explained. The other frequently occurring patterns are those of the same illocution extended into the second part of the direct speech: dissatisfaction-dissatisfaction, direct coercion-direct coercion and belief-belief. Anger-related illocutions (in one or both parts) thus come either without explanation or in two parts the second one of which contains an explanation: the first illocution describes the source of anger (dissatisfaction, and so on) and the second gives explanation why the speaker feels that way.

Examples of the correlated illocutions in rcp2 utterances:

Dissatisfaction-explanation: "Giles is an idiot," he said angrily. "He could have contacted me if he had tried."

Disagreement-explanation: "No, you didn't," said Peter, angrily. "It was my idea."

Direct coercion-explanation: "Put some fucking clothes on, Scott," he said angrily. "You're under arrest."

Repeated illocutions: (irony-irony) "Thank you," he said angrily. "I hope you'll enjoy reading about this in tomorrow's papers." / (order-order) "Get back!" he said angrily. "Don't touch me!"

Table 5. Correlation between the illocutions of part 1 and part 2 of rcp2 utterances

PART1/PART2	explan.	dissatisf.	dir.coer.	self-coer.	belief	disagreem.	express.	indir.co.	query	total
dissatisfaction	15	10	3		1	2	1			32/30
disagreement	10	1	1			2			2	16/16
direct coercion	6		5		1		1	1		14/13
belief	5	1		1	5	1				13/13
expressivity	3	1	2	3	1					10/9
indirect coercion		1		i			1	2		5/5
self-coercion			2	i						3/2
explanation	1									1/1
	40	14	13	6	8	5	3	3	2	94/89

7. Correlation between the sentence type and illocution

Finally the interesting question was whether there is any correlation between the sentence type and illocution. Table 6 shows the findings. Given the number and frequency of the sentence types and the number of illocutions in the sample, they are not entirely unpredictable. Although typically a neutral, emotionally unmarked type, declarative sentences which form over 60% of the direct speech text in the sample realize eight of the nine categories of illocutions. Not only do they express dissatisfaction (25.3%), explanation (24.7%) and belief (16.5%), they also signal disagreement, indirect coercion and promises.

Table 6. Correlation between sentence type and illocution

FORCE/TYPE	declar	other	imper	interr	total
dissatisfaction	40	8	2	15	65
explanation	39	2	1	2	44
direct coercion	9	3	26	3	41
disagreement	20	11	2	1	34
belief	26		1		27
expressivity	3	15	1	1	20
indirect coercion	11		1		12
self-coercion	10	1			11
query				4	4
total	158	40	34	26	258

The remaining sentence types are more specialized: the category other most frequently conveys expressives (37.5%) and disagreements (27.5%), imperative sentences as might be expected express direct coercion (73.5%) and interrogative sentences are mostly used to show dissatisfaction (57.7%). The fact that a single sentence type, declarative sentences, is used to realize all but one type of illocution and that 5 out of the 9 categories of illocu-

tions in the sample are realized by all four sentence types suggests that formal expression, namely the sentence type and its concomitant features such as the mood, word order, etc., are not the decisive or main indicators of illocution, i.e. IFIDs. Obviously, anger in direct speech is conveyed primarily by pragmatic means, drawing on the semantic content of the proposition, rather than by sentence type. Anger is thus expressed implicitly unlike in reporting clauses containing *angrily*.

8. Conclusions

The paper deals with one selected signal introducing an anger coding unit. Although in the whole of the BNC there were "only" 152 instances of this particular kind ("said angrily"), there is no reason to believe that it is in any respect an unusual way of expressing anger in literary text. It seems justified to speak of a pattern (cf. Hunston, 2010) because the anger-conveying stretch of text forms a unit clearly delimited and recognizable in text; it has a predictable and fairly fixed structure, its parts exhibit distinct tendencies in their syntactic and pragmatic makeup and length. In other words anger-coding units are fairly uniform in spite of being gathered from many books by many different authors.

The anger-coding unit is made up of two components: the reporting clause and the anger-bearing direct speech. The unit is identified by the sequence *said angrily* in the reporting clause. Anger-coding units divide into three groups or patterns according to the position of the reporting clause which either precedes or follows the direct speech or divides it into two. The position of the reporting clause and the length of the direct speech appear to be rather constant: undivided direct speech components are typically one sentence long, bisected direct speech components are made up of two sentences.

The paper focused especially on the direct speech component and its syntactic and pragmatic properties. Contrary to the initial hypothesis, by far the most common sentence type in the direct speech components is the declarative sentence accounting for almost two thirds of the sample. There was no formally distinct exclamative sentence in the sample. The remaining one third is comprised of "other" sentences (non-finite, verbless or formulaic), imperative sentences and interrogative sentences. The sentences were found to express 26 different kinds of illocutions which were divided into nine categories. Declarative sentences realize eight of the nine categories of illocutions identified in the text. The other three sentence types, on the other hand, did show a high degree of specialization.

All the same, it is possible to say that there is a weak correlation between the illocutionary profile of the direct speech component and its sentence type structure. The fact that declarative sentences may express most of the anger-related illocutions means that the sentence type is in itself not a sufficient IFID. When coding anger in fiction, the author does need the lexical trigger (*said) angrily* in the reporting clause to help the reader correctly calculate the illocutions in the direct speech component in the absence of extralinguistic and prosodic indicators of anger.

The three groups or patterns of anger-coding units characterized by the position of the reporting clause and the length of the direct speech component (one- and two-sentence) display typical illocutionary profiles: the single-sentence initial reporting clause model is

associated primarily with dissatisfaction and disagreement to the same degree, the single-sentence final reporting clause pattern with dissatisfaction and direct coercion (i.e., order, demand, urging, forbidding), and the two-sentence medial reporting clause model typically combines two illocutions, dissatisfaction and explanation in the first and the second sentence respectively. Incidentally, the findings underscore the fact that Searle's distinction of five speech act types is not very helpful in the analysis of actual utterances as it is too broad and that there is a considerable asymmetry between speech act types and the communicative sentence types.

Naturally, it would be interesting to know whether similar patterns of coding anger in fiction occur in other languages. However, an attempt to collect a comparable sample of anger-coding units in Czech encountered several difficulties. Not only does the English verb *to say* have two equivalents in Czech, *říci* and *pravit*, but unlike English reporting clauses in which the verb *to say* is by far the most preferred one Czech reporting clauses use a much wider range of verba dicendi (some of which incorporate the semantic feature of anger). As a result it is difficult to find a sufficient number of instances with the reporting verbs *řekl(a)* or *pravil(a),* matching the English *said*. In order to get the same number of anger-coding units in Czech fiction at least ten more verbs (such as *zvolat, vykřiknout, prohlásit, odseknout, zavrčet, začít, zasyčet, zařvat, zasyčet, utrhnout se,* etc.) would be required. Similarly in a comparable Czech corpus it is impossible to find the same number of instances of one Czech adverb matching the English *angrily*. Instead it would again be necessary to include synonyms, more than fourteen of them, to get the same sample size (*zlostně, vztekle, rozzlobeně, nazlobeně, rozčileně, naštvaně, rozhněvaně, nahněvaně, rozezleně, dopáleně, dožraně, namíchaně, navztekaně, hněvivě,* etc.). The use of different verbs and adverbs can be expected to influence the illocutionary profile of the direct speech component and so influence the results of comparison with English. Also the different principles of word order in Czech which allow for a variable position of the adverb with regard to the verb complicate the compilation of a reasonably homogeneous corresponding Czech sample.

To conclude, judging from the sample it appears that the language of fiction has developed a relatively stable and conventionalized pattern of coding anger in English, with typical formal and pragmatic features. By a combination of direct indication (lexical trigger) and indirect means (a range of illocutions) it overcomes the absence of context with prosodic and extralinguistic (facial, gestural) signals which communicate anger in spoken language.

REFERENCES

Blumenthal, P., I. Novakova and D. Siepmann (eds.) (2014) *Les émotions dans le discours [Emotions in Discourse]*. Frankfurt am Main: Peter Lang.

Durst, U. (2001) Why Germans don't feel anger. In: Harkins, J. and A. Wierzbicka (eds.), *Emotions in Crosslinguistic Perspective*, 115–148.

Glynn, D. (2014) The social nature of ANGER: Multivariate corpus evidence for context effects upon contextual structure. In: Blumenthal, P., et al. (eds.), 69–82.

Harkins, J. and A. Wierzbicka (eds.) (2001) *Emotions in Crosslinguistic Perspective*. Berlin/Boston: Mouton de Gruyter.

Galasinski, D. (2004) *Men and the Language of Emotions*. Houndmills/New York: Palgrave Macmillan.

Geeraerts, D., C. Gevaert and D. Speelman (2012) How anger rose: Hypothesis testing in diachronic semantics. In: Allan, K. and J. A. Robinson (eds.), *Current Methods in Historical Semantics*, 109–131. Berlin/Boston: Walter de Gruyter.

Goddard, C. and Z. Ye (eds.) (2016) *"Happiness" and "Pain" across Languages and Cultures*. Amsterdam/Philadelphia: John Benjamins.

Gries, S. Th., S. Wulff and M. Davies (eds.) (2010) *Corpus-linguistic Applications. Current Studies, New Directions*. Amsterdam/New York: Rodopi B. V.

Hunston, S. (2010) How can a corpus be used to explore patterns. In: O'Keeffe, A. and M. McCarthy (eds.), *The Routledge Handbook of Corpus Linguistics*. London/New York: Routledge.

Kissine, M. (2012) Sentences, utterances and speech acts. In: Allan, K. and K. M. Jaszczolt (eds.), *The Cambridge Handbook of Pragmatics*, 169–190. Cambridge: Cambridge University Press.

Klégr, A. (2005) *Sadness/smutek*: a comparison of the verbal collocates. In: Čermák, J., A. Klégr, M. Malá, and P. Šaldová (eds.), *Patterns. A Festschrift for Libuše Dušková*, 91–105. Prague: Charles University, Faculty of Arts.

Klégr, A. (2007) Comparing Collocations across Languages: An English-Czech Sample. *Proceedings of the Corpus Linguistics Conference CL 2007*. Birmingham: University of Birmingham. Available at: http://www.birmingham.ac.uk/research/activity/corpus/publications/conference-archives/2007-birmingham.aspx.

Mikolajczuk, A. (2004) Anger in Polish and English. In: Kay, C. J. and J. J. Smith (eds.), *Categorization in the History of English*, 169–178. Amsterdam/Philadelphia: John Benjamins.

O'Keeffe, A. and M. McCarthy (eds.) (2010) *The Routledge Handbook of Corpus Linguistics*. London/New York: Routledge.

Pütz, M. and J. Neff-van Aertselaer (eds.) (2008) *Developing Contrastive Pragmatics. Interlanguage and Cross-Cultural Perspectives*. Berlin/Boston: Mouton de Gruyter.

Ronan, P. (2015) Categorizing expressive speech acts in the pragmatically annotated SPICE Ireland corpus. *ICAME Journal* 39, 2015, 25–45.

Searle, J. R. (1979) *Expression and Meaning. Studies in the Theory of Speech Acts*. Cambridge: Cambridge University Press.

Trosborg, A. (1995) *Interlanguage Pragmatics. Requests, Complaints and Apologies*. Berlin: Mouton de Gruyter.

Weigand, E. (1998) The vocabulary of emotion. A contrastive analysis of ANGER in German, English, and Italian. In: Weigand, E. (ed.), *Contrastive Lexical Semantics*, 45–66. Amsterdam/Philadelphia: John Benjamins.

Wierzbicka, A. (1999) *Emotions across Languages and Cultures: Diversity and Universals*. Cambridge: Cambridge University Press.

Wiklund, A.-L. (2009) The syntax of surprise: Unexpected event readings in complex predication. *Working Papers in Scandinavian Syntax* 84, 181–224.

Electronic sources

British National Corpus (BYU-BNC). Availale at: http://corpus.byu.edu/.

Český národní korpus – InterCorp. Praha: Ústav Českého národního korpusu FF UK. Available at: http://www.korpus.cz.

Český národní korpus – SYN2010 (2010). Praha: Ústav Českého národního korpusu FF UK. Available at: http://ucnk.ff.cuni.cz.

Phrases in English (PIE). Availale at: http://phrasesinenglish.org/.

VYJADŘOVÁNÍ HNĚVU V BELETRII

Resumé

Studie zkoumá, jakým způsobem je v angličtině jazykově zobrazován hněv literárních postav. Tato emoce je v textu signalizována lexikálně v uvozovací větě (*said angrily*) doprovázející přímou řeč, v níž je podle předpokladu tato emoce verbalizována. Uvozovací věta spolu s přímou řečí jsou chápány jako dva komponenty hněv kódující jednotky, která je zkoumána z hlediska jejich vzájemné lineární pozice a z hlediska syntaktických vlastností přímé řeči a její povahy jako mluvního aktu. Analýza a korelace těchto proměnných (včetně délky přímé řeči) naznačují, že existují tři základní modely hněv kódujících jednotek s charakteristickou strukturou a ilokučním profilem.

Aleš Klégr
Department of English Language and ELT Methodology
Faculty of Arts, Charles University
ales.klegr@ff.cuni.cz

Pavlína Šaldová
Department of English Language and ELT Methodology
Faculty of Arts, Charles University
pavlina.saldova@ff.cuni.cz

ON ENGLISH LOCATIVE SUBJECTS

GABRIELA BRŮHOVÁ, MARKÉTA MALÁ

ABSTRACT

The paper analyses English sentences with thematic locative subjects. These subjects were detected as translation counterparts of Czech sentence-initial locative adverbials realized by prepositional phrases with the prepositions *do* (into), *na* (on), *v/ve* (in), *z/ze* (from) complemented by a noun. In the corresponding English structure, the initial scene-setting adverbial is reflected in the thematic subject, which results in the locative semantics of the subject. The sentences are analysed from syntactic, semantic and FSP aspects. From the syntactic point of view, we found five syntactic patterns of the English sentences with a locative subject (SV, SVA, SVO, SV_pass A and SVCs) that correspond to Czech sentences with initial locative adverbials. On the FSP level the paper studies the potential of the sentences to implement the Presentation or Quality Scale. Since it is the "semantic content of the verb that actuates the presentation semantics of the sentence" (Dušková, 2015a: 260), major attention is paid to the syntactic-semantic structure of the verb. The analysis of the semantics of the English sentences results in the identification of two semantic classes of verbs which co-occur with the English locative subject.

Keywords: contrastive approach, locative subject, presentation scale, quality scale, presentative verbs

1. Introduction

The paper analyses English sentences with thematic locative subjects. These subjects were detected as translation counterparts of Czech sentence-initial locative adverbials realized by prepositional phrases with the prepositions *do* (into), *na* (on), *v/ve* (in), *z/ze* (from) complemented by a noun. In the corresponding English structure, the initial scene-setting adverbial is reflected in the thematic subject, which results in the locative semantics of the subject. The sentences are analysed from syntactic, semantic and FSP aspects. From the syntactic point of view, we focus on the various syntactic patterns of the English sentences with a locative subject that correspond to Czech sentences with initial locative adverbials. On the FSP level the paper studies the potential of the sentences to implement the Presentation or Quality Scale. Since it is the "semantic content of the verb

that actuates the presentation semantics of the sentence" (Dušková, 2015a: 260), major attention is paid to the syntactic-semantic structure of the verb.

2. Theoretical background

The English subject has received considerable attention from the contrastive point of view with respect to its syntactically divergent counterparts and FSP function as compared with Czech (cf. Dušková, 1999b, 2015b). Drawing on this direction of research the present study focuses on the sentence-initial thematic clause elements with locative semantics in Czech and English, which perform the syntactic functions of the adverbial and the subject, respectively.[1]

(1) a. *This road* carries a lot of traffic.
     b. *Na této silnici* je velký provoz. (Dušková, 1998: 40)

Example (1) represents a prototypical instance of this type of correspondence. Syntactically, the English sentence displays the pattern SVO (ex. 1a). With the subject in the initial position, the syntactic structure of the English sentence complies with the grammatical word order. The Czech corresponding sentence (ex. 1b) has an initial locative adverbial and the subject in postposition (clause pattern AVS).
Another construction with a locative subject is illustrated by example (2a).

(2) a. *The garden* is swarming with bees.[2]
     b. Bees are swarming *in the garden*.[3]

Sentences such as (2a) have been described as resulting from 'locative alternation' (Salkoff, 1983; Levin, 1993: 53–55; Dowty, 2000a and 2000b; Fried 2005). Locative alternation is a kind of syntactic-semantic alternation which involves both intransitive (ex. 2) and transitive verbs.[4] In ex. (2) the two constructions differ in the assignment of semantic roles to the syntactic clause elements, i.e. valency complements of the verb (Panevová et al., 2014: 122–123): the role of location is assigned alternatively either to the subject (ex. 2a) or to the adverbial (ex. 2b). The two constructions, however, are not semantically equivalent. While (2b) can be described as a simple action (with the subject performing the role of the agent), ex. (2a) refers to a state resulting from the movement. This is closely connected with the 'holistic' vs. 'partitive' interpretation of the locative element (cf. Anderson, 1971). Example (2a) illustrates the 'holistic' interpretation of the locative subject, i.e. the whole garden is affected by the action; the movement taking place in the whole area of the garden constitutes in fact a characteristic of the location. Example (2b)

---

[1] The primary concern is the English construction with a locative subject, the corresponding Czech sentences are commented on only marginally where relevant.
[2] Cf. Dowty's (2000a: 112) "LOCATION-SUBJECT FORM (or L-Subject Form)"
[3] Cf. Dowty's (2000a: 112) "AGENT-SUBJECT FORM (or A-Subject Form)"
[4] Cf. Dowty's (2000a: 112) examples *Mary sprayed paint on the wall.* vs. *Mary sprayed the wall with paint.* However, in the case of transitive verbs in the active voice ("spray/load" verbs) locative alternation does not involve the subject but the postverbal elements, namely the object and adverbial.

illustrates the 'partitive' interpretation, i.e. only a part of the garden is affected by the action. The intransitive verbs which co-occur with a locative subject in the construction arising from the locative alternation (ex. 2) can be grouped into the following semantic classes (Dowty, 2000a: 115):

a. physical movements visually recognizable readily and at a 'small scale', usually occurring repetitively (*crawl*, *drip*, *foam*, *heave*, *vibrate*, etc.);
b. animal sounds and other perceptually simple sounds (*hum*, *buzz*, *rustle*, *resound*, etc.);
c. conceptually simple visual perception of some kind of light emission (*beam*, *flame*, *glow*, *glisten*, etc.);
d. smells (*reek*, *smell*, *be fragrant*, etc.);
e. predicates indicating degree of occupancy or abundance (*abound*, *teem*, *be rich (with)*,[5] etc.).

They are all atelic, intransitive verbs which can be characterized as 'process verbs'.[6] "They designate perceptually simple activities usually recognizable from temporally and spatially limited input." (Dowty, 2000a: 116) The activity "fills the space denoted by the Location subject" (ibid.: 120). Another characteristic feature of the construction concerns the *with*-phrase. According to Dowty (2000a: 117) the complement of the preposition must be realized by a plural or mass noun (e.g. *bees* in ex. 2a), and not by a definite, indefinite or quantified noun (e.g. **The garden is swarming with a bee.*).

From the FSP point of view, both the Czech and the English sentence in example (1) implement linear modification, i.e. the sentence elements manifest a gradual rise in communicative dynamism (CD) in the direction from the beginning to the end of the sentence. In other words, the least dynamic elements (i.e. elements that carry the lowest amount of information) *this road / na této silnici* precede the most dynamic elements (i.e. elements that carry the greatest amount of information) *a lot of traffic / velký provoz*. Thus, the element towards which the communication is perspectived occupies the final position. It is "through a locative subject construction of the adverbial element" that the English sentence achieves the basic distribution of the degrees of CD (Adam, 2013: 148; cf. also Quirk et al., 1985: 747–748; and Dušková, 1999b: 248ff.). Both the English and Czech sentences display the same linear increase in the degrees of CD despite their different syntactic structure. "Here the role of FSP as the motivating factor of the syntactic divergence is self-evident." (Dušková, 2015b: 39)

In the theory of FSP every sentence can be seen to implement one of the dynamic semantic scales: Presentation Scale or Quality Scale (Firbas, 1992: 87). Since the focus of the present paper lies in studying the potential of the locative constructions to implement the Presentation or Quality Scale, it is necessary to provide a more detailed characterization of these scales. In the dynamic semantic scales, each element is ascribed one of the dynamic-semantic functions (DSFs). The scales functionally reflect the distribution of CD and operate irrespective of word order (Firbas, 1992; Adam, 2013: 14; Chamoniko-

---

5 The examples given by Dowty involve not only verbal predicates but also copular predications (*be fragrant*, *be rich/rife/rampant (with)*).
6 The verbs typically imply neither change nor intentionality, although some may involve a certain degree of volition and control (cf. Kettnerová, 2015: 215)

lasová, 2010: 87). In the Quality Scale the English verb ascribes a quality to the subject, the communication being perspectived away from the subject. As Adam (2013: 46) points out, "[s]omething new (Specification) is said about the subject (Bearer of Quality)". In the Presentation Scale the verb performs the dynamic semantic function of presentation (i.e. it presents something new on the scene), the communication being perspectived towards the subject, which is the most dynamic element (Phenomenon), the least dynamic element being the Setting of the action ("usually temporal or spatial items of when and where the action takes place"(Adam, 2013: 45)). "For a verb to perform the presentation function, the subject must be context-independent: it then has the DSF of a phenomenon presented on the scene." (Dušková, 2015a: 260)

In the English construction with a locative subject (ex. 3a) the subject is context-dependent, performing the DSF of a setting. Thus, the construction "seems to implement the Quality rather than the Presentation Scale, displaying a thematic subject and a rhematic object" (Adam 2013: 148).

(3) a. The letter bears your signature. (Dušková, 1999b: 248)
    b. Na tom dopise je váš podpis.
    c. There is your signature on the letter.

However, it has been pointed out by several authors (cf. Dušková, 1998; Adam, 2013: 148; Rohrauer, 2013) that the English construction with a locative subject should be regarded as one of the syntactic forms of the Presentation Scale. As mentioned above, the Presentation Scale is characterized by three basic DSFs (Adam, 2013: 14), and example (3a) displays all three features: i. the act of communication is set by the scene of the action (DSF of a setting – a spatial item, i.e. where the action takes place), syntactically constructed as the subject *the letter*; ii. the object *your signature* represents the most dynamic element (Phenomenon) ushered onto the scene; iii. the existence on the scene is conveyed by the verb *bears*.

Figure 1. DSFs in the Presentation Scale

Setting	Presentation of Phenomenon	Phenomenon Presented
theme	transition	rheme
The letter	*bears*	*your signature.*

That the sentence in example (3a) implements the Presentation Scale is revealed not only by the dynamic semantic functions of the clause elements (Figure 1), but also by the possible alternative expression of the same content using the existential construction (ex. 3c). The existential construction is regarded as the most frequent realization form of the Presentation Scale (cf. Adam, 2013: 62; Rohrauer, 2013: 147; Dušková, 1998: 36). In the existential construction the scene is set by the adverbial (*on the letter*), "while the phenomenon appearing on the scene (the rheme, *your signature*) assumes the syntactic function of a subject" (Dušková, 1998: 40). As the paraphrase with the existential construction demonstrates, the syntactic form SVO with a locative subject is a surface structure implementing in its deep structure the Presentation Scale.

To sum up the ordering of the clause elements in the English construction with the initial locative subject we may say that although the English structure is syntactically divergent from the Czech structure, the elements are ordered in accordance with both the basic distribution of CD and the grammatical word order: "Without the syntactic divergence the English sentence structure would deviate either from the grammatical word order or from the basic distribution of CD. As a result of the syntactic divergence it complies with both" (Dušková, 2015b: 44).

From the semantic point of view, a crucial role appears to be played by the English verb, as it is the "semantic content of the verb that actuates the presentation semantics of the sentence, the so-called presentation scale" (Dušková, 2015a: 260). Verbs that can express the presentation function are traditionally defined on the basis of their static semantics. They generally belong to two classes; they express existence or appearance either explicitly or implicitly (cf. Firbas, 1992: 60). The group of presentation verbs that express existence or appearance explicitly is relatively well-defined, prototypical verbs of this group being *come, arrive, enter, appear, occur, turn up*, etc. The other group of verbs that are also capable of expressing the existence or appearance on the scene, yet in an implicit way (i.e. "with sufficient implicitness", cf. Firbas (1992: 65)) is much more difficult to delimit. According to Adam this class includes a relatively large group of different verbs that "come from different semantic groups of verbs and do not have much in common" (Adam, 2010: 22), e.g. *strike, await, buzz, shine, seize, pour, feed, blow*. As regards their valency, the presentation verbs are typically intransitive (e.g. *come, arrive, enter, appear*). Nevertheless, previous studies (Firbas, 1992: 61; Adam, 2011: 22; Adam, 2013: 158; Dušková, 2015a: 262) have shown that the presentation function can also be ascribed to transitive verbs (e.g. *fill, flood, brim*).

As for the verbs occurring in the construction with a locative thematic subject, their repertoire appears to be rather limited. They were shown to be typically monotransitive verbs exhibiting similar semantic features. The verbs most frequently attested in the literature are *have* (ex. 4) and *bear* (ex. 3a). Adam (2013: 150) points out that these verbs "seem to be quite general and auxiliary-like, denoting existence/appearance on the scene". They may even be considered "quasi-copulative for they merely provide a link between the participants in verbal action without contributing any relevant lexico-semantic feature" (Dušková, 1999b: 250).

(4) Inside, the cave was dark and cold and *had* the damp feel and smell of a place that had not been lived in for several days. (Adam, 2013: 149)[7]

Example (5) illustrates another important feature of the semantic content of the English presentation verbs, namely semantic affinity between the predicative verb and its subject (cf. Firbas, 1992: 60; Dušková, 2015a: 262; Adam, 2011: 22 and 2013: 160).

(5) A bee buzzed across their path.

---

[7] The choice of the construction with *have* (rather than the *there*-existential construction) may be influenced by the tendency towards a constant subject in English.

Adam (2011: 22) describes semantic affinity as one of the most significant features "of what may be considered to express existence or appearance on the scene in an implicit way". Semantic affinity is typically found with one syntactic type of Presentation Scale, namely Rhematic subject in preverbal position (ex. 5). In ex. (5) the action of the verb *buzz* is "so semantically inherent to the subjects (subject-related) that it is the subject [*bee*] that takes over the communicative prominence at the expense of power of the verbal content… The static semantics of the verb then – even if expressing a specific type of action – is reduced to that of presentation" (Adam, 2011: 23). Thus, through this affinity the verb prepares the way for the presentation of a new phenomenon (cf. Firbas, 1992: 60). Although Dowty (2000a: 117) does not mention this explicitly, the examples he gives to illustrate the semantically 'unquantified' nature of the *with*-phrase object in the locative subject construction display a similar semantic affinity between the verb and the complement of the preposition *with* (e.g. *crawl – roach, buzz – flies, tinkle – bells, blaze – lights/stars/bonfires/flashbulbs* in ex. 6).

(6) a. The wall crawled with roaches.
    b. The bottle buzzed with flies.
    c. The city square tinkled (resounded) with the sound of many bells (on the horses and carts).
    d. The sky blazed with lights/stars/bonfires/flashbulbs.

3. Material and method

The present study takes as its starting point the assumption that "linguistic structure is language-specific while the cognitive and functional-communicative substance which constrains it is potentially universal" (Boye, 2012: 7, cf. Haspelmath, 2010; Martinková, 2014). While the same semantic structure of the sentence occurs both in Czech and in English, serving the same "needs of expression and communication" (Mathesius, 1936: 95), it is likely to be represented partly by different linguistic structures in each of the languages (cf. Gast, 2015).

The material used to study the correspondences between the two languages was drawn from the fiction core of the Czech-English sub-corpus of the parallel translation corpus InterCorp (version 8).[8] Czech originals were selected as source texts, and the Czech sentence-initial adverbials as 'markers' of the locative (adverbial) meaning (Malá, 2013), or 'methodological anchors', i.e. formal correlates of the functional operation of localization (Gast, 2015: 8).

The study focuses on syntactically divergent translation counterparts of Czech adverbials, namely on the correspondences between a Czech adverbial and an English subject. As pointed out by Dušková (1999b: 254), "if an adverbial is to be convertible into the subject it must be a potential participant in the verbal action and its form must be such as to allow the concomitant formal changes." To comply with the first requirement, only

---

[8] The subcorpus used for the present study comprises 19 Czech books and their English translations. The size of the Czech part of the subcorpus is 2,014,551 tokens, and that of the English translations 2,293,237 tokens.

adverbials integrated in the sentence structure were considered. The form of the adverbial was restricted to prepositional phrases with prepositions occurring frequently in locative uses, i.e. *do* (into), *na* (on), *v/ve* (in), *z/ze* (from) (cf. Klégr et al., 2011: 18). The use of the parallel corpus made it possible to excerpt a sufficient number (130) of translation pairs of sentences comprising a Czech sentence-initial locative adverbial[9] and a corresponding English subject to identify several syntactic patterns employed in English to convey the sentence-semantic concept in question. Where needed, additional excerption was performed from the English original fiction texts in the same corpus. The additional queries focussed on specific English patterns in order to either rule out the translation effects or to provide additional material, expanding on the rather limited number of instances in some of the categories. The correspondences will be described in terms of their syntactic structure, static semantics and distribution of CD.

Table 1. Numbers of translation pairs comprising a sentence-initial locative adverbial prepositional phrase in Czech and a corresponding English locative subject

preposition	sentence-initial A_{loc} – S correspondences
do (into)	3
na (on)	38
v/ve (in)	77
z/ze (from)	12
Total	130

4. Analysis

The following section focuses on the English translation counterparts of Czech sentences with sentence-initial locative adverbials realized by prepositional phrases. The dominant pattern in Czech sentences is the AVS pattern. All the patterns attested in our data are presented in Table 2. The analysis of the translation counterparts of the Czech sentences revealed five different English clause patterns (see Table 3). Two of these patterns comprise intransitive verbs (patterns SVA and SV), two are based on transitive verbs (patterns SVO and $SV_{pass}A$) and one on copular verbs (pattern SVCs). These patterns will be dealt with below in the order in which they are presented in Table 3.

---

[9] The correspondences between Czech adverbials realized by prepositional phrases and English subjects also occur at the beginning of clauses within a sentence, e.g. *Strážmistr vytáhl ze stolku dvě svíčky, na kterých byly stopy od pečetního vosku, jak pečetil úřední spisy ...* – *The State police Station Chief pulled out of the little table two candles which bore traces left by the sealing wax as he would seal official files ...* The dominant realization form of the prepositional complement / subject in these cases is pronominal, unlike in the sentence-initial position. These correspondences were therefore excluded from the present study, although patterns of correspondence parallel to those identified in sentence-initial position can be expected (cf. also *Auta... zmenšila chodníky, na kterých se tlačí chodci.* – *The cars... have narrowed the pavements, which are crowded with pedestrians.*)

Table 2. Czech clause patterns with initial locative adverbials

Czech pattern	Number of occurrences	Czech sentence	English translation counterpart
AVS	83	*Ve velkém sále bylo mnoho stolů a mnoho židlí…*	*The hall had lots of tables, lots of chairs…*
subjectless	24	*V nemocnici bylo čisto...*	*The hospital was clean…*
AVSO(O)	16	*Na tváři měl klidnej úsměv.*	*His face wore a peaceful smile.*
AVCsS / AV$_{pass}$S	7	*V její intonaci nebylo zatím přítomno žádné hodnocení.*	*Her intonation betrayed no hint of judgement as yet.*
Total	**130**		

Table 3. English clause patterns with locative subjects

English pattern		Number of occurrences	English examples	Czech source sentences
intransitive	SVA	13	*The restaurant was humming with life.*	*V restauraci bylo poměrně živo.*
	SV	9	*The room reeked.*	*V pokoji byl puch.*
transitive	SVO	59	*Every chair had an inscription.*	*Na každé židli je nápis.*
	SV$_{pass}$A	11	*The walls were hung with interesting pictures…*	*Na zdech byly zajímavé obrazy…*
copular	SVCs	38	*The flat was quiet.* *This realm was alive with all sorts of women…*	*V bytě bylo ticho…* *Na tomto území se pohybovaly různé ženy…*
Total		**130**		

4.1 Intransitive verb patterns (SVA and SV)

4.1.1 SVA

In twelve out of the 13 cases[10] the source of the English SVA pattern was the Czech pattern AVS. Thus, the Czech locative adverbial is mirrored in the English subject and the Czech subject finds an expression in the adverbial (as a complement of the preposition) in the English sentence. The sentences with this pattern can be characterized by a semantic frame that is associated with perception and experience, namely the 'sensory experience' frame. The frame "must contain minimally three elements: a perceiver, an entity that triggers the sensory effect (= stimulus), and a place in which the perception holds" (Fried, 2005: 489). Since the perceiver is usually the speaker, it remains unexpressed. The trigger of perception (stimulus) is linked with the syntactic function of the adverbial (*with peaceful joy* in ex. 7). The place in which the perception holds is syntactically realized as the subject (*her eyes* in ex. 7).

The verbs in this pattern fall into several semantic types which correspond to those listed as being typical of the construction with a locative subject (cf. Dowty, 2000a: 115;

---

[10] In one case the SVA pattern corresponds to a Czech subjectless verbonominal clause: *The restaurant was humming with life. – V restauraci bylo poměrně živo.*

Panevová et al., 2014: 123): light emission (*glow* in ex. 7), physical movement, typically occurring repetitively (*swarm* in ex. 8), animal and other sounds (*buzz* in ex. 9), smells and tastes (*reek* in ex. 10).

(7) Her eyes *glowed* with ... peaceful joy.
V očích jí zářila ... klidná radost.
(8) The country *was swarming* with innumerable doughty foes of Communism.
Ve vlasti se vyrojily mraky čackých bojovníků proti komunismu.
(9) My head stopped *buzzing* with words and I tried to speak to him kindly.
V hlavě mi přestala hučet slova a já k němu zkusil promluvit laskavě.
(10) The boy's breath *reeked* of alcohol.
Z chlapcova dechu odporně páchl alkohol.

All the subjects carry the semantic role of location, which "is attributed a distinctive property, namely a sensory effect generated by the stimulus and thoroughly affecting the place" (Fried, 2005: 492). The stimulus is expressed by the prepositional phrase (typically *with*-phrase). In compliance with Dowty's observations the complement of the preposition is a plural (e.g. *words* in ex. 9) or an uncountable noun (e.g. *alcohol* in ex. 10). As regards its semantics, the prepositional phrase denotes "an unquantifiable secondary agent (a mere instrument)" (Fried, 2005: 493).

Some of the examples were found to display close semantic ties between the verb and the noun occurring in the prepositional phrase. This is reminiscent of the semantic affinity between the subject and the verb mentioned already by Firbas (1992: 60) and Adam (2013: 160). In most of our examples, however, the affinity does not hold between the verb and its subject but between the verb and the noun in the prepositional phrase (e.g. exx. 9–10: *buzz – words, reek – alcohol*). It seems that semantic affinity can be found only with nouns used in their literal sense (ex. 10). This explains why no affinity can be observed between the verb *swarm* and the following noun *foes* (ex. 7).[11] It has to be pointed out that in some cases a certain degree of semantic affinity can also be detected between the verb and the locative subject (*ring – voice* in ex. 11, where *ring* is the counterpart of an implied Czech verb).

(11) The virtuoso's voice *rings* with uncounterfeited astonishment.
Ve virtuosově hlase nefalšované překvapení.

The sentences with the SVA conform to the grammatical word order. At the same time, they implement linear modification, i.e. there is a gradual rise in CD in the direction from the locative subject to the most dynamic element, namely the adverbial realized by the prepositional phrase. Based on the surface structure the communication seems to be perspectived away from the subject with the sentence therefore implementing the Quality Scale. However, in its deep structure it implements a Presentation Scale (ex. 7, repeated here as 12a; Figure 2). The subject (*her eyes*) is context-dependent and it

---

[11] In its literal meaning, the verb *swarm* evokes the movement of insect. Therefore, in examples like *The garden is swarming with bees* (Fried, 2005: 476) the affinity between the verb and the noun in the prepositional complement is retained.

performs the DSF of a setting. The DSF function of the verb *glowed* is that of presentation. It presents a new phenomenon (*with peaceful joy*) on the scene. The Presentation Scale interpretation of the SVA sentences with a locative subject can be supported by paraphrases using existential constructions with the verbs *be* and *have*[12] (exx. 12 b and c, respectively):

(12) a. Her eyes *glowed* with ... peaceful joy.
     b. There was peaceful joy in her eyes.
     c. Her eyes had peaceful joy in them.

Figure 2. DSFs in the SVA pattern

Setting	Presentation of Phenomenon	Phenomenon Presented
theme	transition	rheme
Her eyes	*glowed*	*with peaceful joy.*

4.1.2 SV

Locative subjects were also attested in the SV pattern (exx. 13–15) with the verbs pertaining to the same semantic classes as in the SVA pattern. These examples are translation counterparts of Czech sentences displaying the pattern AVS (5 instances, ex. 13) and subjectless patterns (4 instances, exx. 14–15). Syntactically, in both types the English subject corresponds to the Czech locative adverbial. The English verb corresponds to the Czech verb in the subjectless patterns (exx. 14, 15) and covers the meaning of both the verb and the Czech subject in the AVS pattern (ex. 13).[13]

(13) The room *reeked*.
     V pokoji byl puch.
(14) The scarred face *twitched*, but that was all.
     Ve zjizvené tváři zaškubalo, ale to bylo vše.
(15) Her eyes suddenly *sparkled*…
     V očích jí zajiskřilo…

From the semantic point of view, the SV sentences refer to the same semantic frame of 'sensory experience' as the SVA ones, but only one of the three essential elements of the frame is present, the location. Due to the absence of the stimulus (the trigger of perception linked with the prepositional phrase in SVA) there is no successful competitor of the verb in terms of CD. Therefore, "the notional component of the verb abandons the transitional layer to complete the message as the most dynamic element of the distributional field" (Adam, 2013: 50). Thus these sentences clearly implement the Quality Scale[14] (Figure 3).

---

[12] On *there-/have*-existentials, see Quirk et al., 1985: 1402–1414.
[13] In the Czech sentence the subject (*to*) may be semantically empty (Kettnerová, 2015: 229–231):
   V kuchyni to páchlo – The kitchen reeked.
[14] This is also supported by the impossibility of the *there* existential construction.

Figure 3. DSFs in the SV pattern

Bearer of Quality	Quality
theme	transition/rheme
The room	*reeked.*

4.2 Transitive verb patterns (SVO and SV$_{pass}$A)

4.2.1 SVO

As was expected on the basis of previous research, SVO represents the most frequent pattern (45 per cent of sentences with the locative subject). The patterns of the Czech source sentences which occurred more than once comprise: AVS (42 instances, exx. 16, 19), AVSO (10 instances, exx. 17, 18) and AVCsS / AV$_{pass}$S (5 instances, ex. 23). Syntactically, the Czech locative adverbial is reflected in the English subject. The Czech subject in AVS and AVCsS / AV$_{pass}$S, (as well as the object in AVSO, in ex. 18) find an expression in the object in the English sentence (exx. 16, 22), retaining the clause-final position. The Czech subject in the AVSO pattern may be reflected in the possessive determiner within the English subject (*his* in ex. 18) but in some translation pairs it may have a zero English counterpart, being merely implied in the English sentence.[15]

Semantically, the English verbs in the pattern SVO fall into two classes. The first class can be described as 'contain/be covered'[16] verbs. The most frequent verb in this class was the verb *have* (18 instances, ex. 16a). In this construction, *have* is "stripped of its possessive meaning and is actually semantically emptied in this sense" (Adam, 2013: 149).

(16) a. Every chair *had* an inscription.
      Na každé židli je nápis.
   b. There was an inscription on every chair.

Our examples also contain other verbs which have already been described in literature: *bear* (ex. 17), *wear* (ex. 18), *hold* and *contain* (ex. 19). The verb *house* (ex. 20) is similar to these verbs. While all these verbs fall in the 'contain/be covered' class, they form a cline from the most semantically empty verb *have* to the verbs of more specific meaning (*wear*, *house*).

(17) The fashionable T-shirt covering her well-formed chest *bore* a chartreuse message…
     Na módním tričku, pokrývajícím dobře vzrostlou hruď, měla jedovatě zelený nápis…
(18) His face *wore* a peaceful smile.
     Na tváři měl klidnej úsměv.

---

[15] This typically applies to the Czech unexpressed subject (general human agent): *The papers say:…* – *V novinách píší:…*

[16] Levin (1993: 119) classifies the verbs as "*fill* verbs". The verbs in our sample can be more accurately described as 'contain/be covered' verbs.

(19) The white envelope *contained* a brief letter from Kral, our local millionaire…
    V bílé obálce byl krátký dopis od místního milionáře Krále…
(20) A dusty display case used to *house* a set of fancy shot glasses from Moscow.
    V zaprášený vitríně stálo pár zdobenejch moskevskejch stakanů.

The second semantic class of verbs in the SVO pattern comprises verbs which may be described as "verbs of transfer of a message" (Levin, 1993: 202) or "manifestation verbs"[17]. All these verbs are related to sensory perception, since the message is perceived through the senses, e.g. *show* (9 instances, ex. 21), *display* (3 instances, ex. 22), *read* (2 instances, ex. 23), *say* (3 instances, ex. 24), *betray* (2 instances, ex. 25), *give off, radiate, echo.*

(21) The photograph *showed* a naked child on a sheepskin rug.
    Na fotce bylo děťátko pasoucí ovce.
(22) The next page *displayed* a large wedding picture…
    Na další stránce se zaskvěla veliká svatební fotka…
(23) The sign *read*: Do not enter, this door is self-entering.
    Na tabulce bylo napsáno: Nevstupovat, vstupuje samo!
(24) He sent … two letters for me. One *said* he was doing well…
    V jednom stálo, že se má dobře…
(25) His voice *betrayed* not the slightest regret.
    V jeho hlase nebyl sebemenší náznak pokání.

It seems that all the verbs in this pattern are semantically similar to *have*, yet their meaning is more specific and more closely related to the meaning of the subject. The verbs link the subject with the object which denotes an inalienable part of the subject (e.g. the *naked child* is a part of *the photograph* in ex. 21). This semantic affinity between the verb and the subject is clearly illustrated in exx. 21–23 (*show – photograph, display – page, read – sign*).

Since the meaning of the verb is already implied in the subject, the degree of CD of the verb is low and its DSF is reduced to that of presentation (cf. Adam, 2011: 23). The subject is inanimate, denoting the location. The object represents the phenomenon presented on the locative scene. The SVO sentences with the locative subject can therefore be regarded as one of the realization forms of the Presentation Scale (Figure 4).The presentational nature of the construction can be manifested by a paraphrase using an existential *there*-construction (ex. 16b).

Figure 4. DSFs in the SVO pattern

Setting	Presentation of Phenomenon	Phenomenon Presented
theme	transition	rheme
Every chair	*had*	*an inscription.*

---

[17] This category label was suggested to us by Martin Adam.

In three instances of the SVO pattern the locative reference is expressed twice (exx. 26–27), once in the locative subject and once in the clause-final adverbial. The adverbial is realized by a prepositional phrase with a complement coreferential with the subject. "The doubled expression of the locative reference" (Adam, 2013: 149) is found with the verb *have* only, specifying the location.

(26) The table *had* a lamp *on it*.
     Na stolku stála rozsvícená lampa…
(27) … the corrals didn't *have* any trees *in them*.
     V ohradách nebyly žádné stromy.

4.2.2 SV_{pass}A

The English constructions with a locative subject described in literature are always (to our knowledge) in the active voice. However, our analysis revealed that the locative subject is found not only in active, but also in passive constructions (11 instances, ex. 28a). The original Czech sentence patterns which occurred more than once are AVS (7 instances, ex. 28) and AVCsS / AV_{pass}S (2 instances[18]). Similarly to the AVS – SVA correspondence, the Czech locative adverbial is mirrored in the English subject and the Czech subject is expressed in the adverbial in the English sentence.

(28) a. The room *was filled* with a dim and oppressive twilight…
      V pokoji panovalo dusné a temné přítmí…
    b. A dim and oppressive twilight filled the room.
    c. There was a dim and oppressive twilight in the room.
(29) a. A cold blue light *filled* the window panes. (Firbas, 1992: 61)
    b. The window panes were filled with a cold blue light.

These passive constructions are analogous to Firbas's active voice example (29a), which can be demonstrated by the active transformation (28b). Firbas describes the DSFs of clause elements in ex. (29a) as follows: "the object [window panes] expresses a phenomenon that is filled, permeated or covered by another phenomenon [a cold blue light]. The latter [a cold blue light] appears within the space provided by the former [window panes]. The latter [a cold blue light] is the phenomenon to be presented; the former [window panes] serves as the setting (scene) for the presentation" (Firbas, 1992: 61). Despite the different syntactic arrangement of the elements in ex. (28a), their DSFs remain identical, i.e. the locative subject denotes the setting (*the room*), the prepositional phrase (*with a dim and oppressive twilight*) expresses the phenomenon which is presented within the space of the room through the presentation verb in the passive (*was filled*) (cf. Figure 5). This interpretation is illustrated by the possible passive transformation of Firbas's example in ex. (29b). The passive construction (28a) also allows the *there*-existential transformation (ex. 28c), which may serve as a test of the Presentation Scale.

---

[18] E.g. *The valleys were dotted with the cottages of widely scattered villages. – V jejich údolích byly roztroušeny ve značných vzdálenostech domky nadmíru roztáhlých vsí.*

Figure 5. DSFs in the SV$_{pass}$A pattern

Setting	Presentation of Phenomenon	Phenomenon Presented
theme	transition	rheme
The room	*was filled*	*with a dim and oppressive twilight.*

In addition, the analysis revealed that in all SV$_{pass}$A constructions the *by*-agent of the action is not expressed. Interestingly, the suppression of the agent can be seen as a relevant factor that is closely connected with the Presentation Scale. Previous research has shown that in some cases of the Presentation Scale the phenomenon appeared on the scene "through some (unexpressed) external agency" (cf. Dušková, 2015a: 264; Brůhová, 2015: 32).[19] All the passive constructions have a stative meaning, denoting a state which results from the verbal action. They attribute a certain property to the subject. In this respect they resemble the SVCs constructions.[20]

The verbs in the SV$_{pass}$A pattern can be semantically classified as 'contain/be covered' verbs. They include: *be bathed, coated, covered, decorated, dotted, filled, hung, inscribed, jammed, swathed* (*with/in*) (exx. 30–32).

(30) The walls *were hung with* interesting pictures, mostly photographs and posters.
     Na zdech byly zajímavé obrazy, mnoho fotografií a plakátů.
(31) The kitchen table *was covered with* baggies of powder.
     Na kuchyňským stole byly plastikový balíčky s řezanikem.
(32) Their greenish skins *were coated with* shiny insects and black flies.
     V tý nazelenalý žluti byly závoje zářivýho hmyzu i černejch much.

4.3 The copular verb pattern SVCs

The last English clause pattern which was revealed as a translation counterpart of the Czech locative constructions is the SVCs pattern (38 instances). The source patterns differ from the previous groups in that the subjectless (ex. 33) and AVS (ex. 34) patterns are equally represented (17 instances each). Syntactically, the Czech locative adverbial corresponds to the English locative subject as in all previous patterns. The semantics of the English subject may be described as a blend of locative and affected semantic roles. While according to Quirk et al. (1985: 743) the subject of copular verbs is generally assigned the affected role, in these examples it is a counterpart of the Czech locative adverbial (ex. 33).

(33) The room **was dim**.
     V pokoji bylo šero.

As regards the 'sensory experience' frame, the SVCs pattern may be divided into two subtypes according to the number of participants overtly expressed. In a majority of

---

[19] Cf. also Firbas (1992: 63): *Then a blind and dumb demoniac was brought to him.*
[20] Moreover, some of the passive constructions are non-agentive (ex. 28a), i.e. "there seems to be no agency involved" (Dušková, 1999a: 120).

examples only one participant, namely the place, is expressed; the stimulus remains unexpressed.

From the FSP point of view these sentences implement the Quality Scale. The copular verb ascribes a quality to the subject and the subject complement performs the function of Quality (see Figure 6).

Figure 6. DSFs in the SVCs pattern

Bearer of Quality	Ascription of Quality	Quality
theme	transition	rheme
The room	*was*	*dim.*

In a small group of SVCs sentences (5 instances) two participants of the 'sensory experience' frame are expressed – the place and the stimulus. The stimulus is realized by an *of-/with-* prepositional phrase (exx. 34a and 35) complementing the adjective.

On the level of FSP, the prepositional phrase may be seen as a competitor of the verbonominal predicate in terms of CD. This makes it possible to view the sentences of the $SVCs_{(+comp)}$ pattern as an implementation of Presentation Scale (see Figure 7). The subject (*their faces*) denotes the setting, the verbonominal predicate (*were full*) can be assigned the DSF of presentation (expressing the existence of a phenomenon) and *real anger* represents the phenomenon ushered on the scene. The existential construction, which we have used as the test for the Presentation Scale, is also acceptable with the $SVCs_{(+comp)}$ pattern (ex. 34b).

In this respect these examples are reminiscent of the $SV_{pass}A$ pattern (ex. 28a repeated here as ex. 36).[21] Both $SVCs_{(+comp)}$ and $SV_{pass}A$ patterns have a locative subject, a complex predicate (copular/auxiliary verb (*be*) plus a notional component – *full, filled*) and a clause-final prepositional phrase (usually *with*-prepositional phrase).

(34) a. Their faces *were full* <u>of</u> real anger.
　　　 V jejich tvářích byla skutečná zloba.
　　b. There was real anger in their faces.
(35) This realm *was alive* <u>with</u> all sorts of women.
　　 Na tomto území se pohybovaly různé ženy.
(36) The room *was filled* <u>with</u> a dim and oppressive twilight…
　　 V pokoji panovalo dusné a temné přítmí…

Figure 7. DSFs in the SVCs pattern with a prepositional phrase (potential presentational interpretation)

Setting	Presentation of Phenomenon	Phenomenon Presented
theme	transition	rheme
Their faces	*were full*	*of real anger.*

---

[21] The example *The room became filled with a stillness, which made him through some sort of mysterious power move closer to her.* represents a borderline case between SVCs and $SV_{pass}A$ patterns.

5. Conclusions

In the present study an attempt has been made to explore the English construction with a locative subject corresponding to a Czech initial locative adverbial from the syntactic, semantic and FSP aspects.

All the constructions in our sample (130 instances) share the same background semantic frame, i.e. the 'sensory experience' frame (Fried, 2005: 489). The frame comprises three participants: the place, the stimulus and the perceiver. Table 4 illustrates the relation between the semantic frame participants and the syntactic realization forms. The place is always overtly expressed in the locative subject. Semantically it denotes location which combines with the affected role in the SVCs and $SV_{pass}A$ patterns (*flat* and *walls* in exx. (ii) and (iv), respectively). The stimulus (the trigger of perception) has two realization forms, depending on the pattern: the object in the SVO pattern (*a naked child* in ex. (vi)), and *with/of/in* prepositional phrase in the SVA, $SV_{pass}A$ and $SVCs_{(+comp)}$ patterns (*alcohol*, *pictures* and *women* in exx. (iii), (iv), (v), respectively). In the patterns SV and SVCs (without the prepositional complement) the stimulus is unexpressed (exx. (i), (ii)). The perceiver (typically the speaker) always remains unexpressed.

Table 4. The relation between the semantic frame participants and the syntactic realization forms

semantic frame participants	syntactic realizations			example
	un/expressed	syntactic function	syntactic pattern	
place	expressed	subject	all patterns	–
stimulus	unexpressed	–	SV	(i) *The room reeked.*
			SVCs	(ii) *The flat was quiet.*
	expressed	*with/of/in-* prep. phrase	SVA	(iii) *The boy's breath reeked of alcohol.*
			$SV_{pass}A$	(iv) *The walls were hung with interesting pictures...*
			$SVCs_{(+comp)}$	(v) *This realm was alive with all sorts of women...*
		object	SVO	(vi) *The photograph showed a naked child on a sheepskin rug.*
perceiver	unexpressed	–	all patterns	–

The lexical verbs[22] in all the constructions with a locative subject seem to fall within two broad semantic classes. The first class may be called 'contain/be covered' verbs. These verbs occur in the SVO and $SV_{pass}A$ patterns: *have, bear, contain, wear, hold, house*; *be bathed, coated, covered, decorated, dotted, filled, hung, inscribed, jammed, swathed* (*with/in*). All the verbs in the second class seem to be related to sensory perception; the process denoted by the verb is perceptible through the senses. On the one hand, they include verbs connected with light, sound, smell emission or movement (cf. Dowty, 2000a): *buzz,*

---

[22] In the SVCs pattern the semantics of the verb is reduced to linking function.

hum, ring, glow, flash, sparkle, reek, bristle, drip, swarm, twitch. These verbs occur in the patterns SVA, SV. On the other hand, there are 'manifestation verbs', such as *show, display, radiate, read, say, echo*, which occur in the SVO pattern.

From the FSP point of view, the aim of our study was to investigate the potential of the English sentences with a locative subject to implement the Presentation or Quality Scale. Here, a crucial role appears to be played by the semantics of the verb. The analysis has shown that there is a difference between the SV and SVCs on the one hand and the other patterns (SVA, SVO, $SV_{pass}A$ and SVCs with a prepositional complement) on the other. In the patterns SV and SVCs without the prepositional complement no new phenomenon is introduced onto the scene, and the verb ascribes a quality to the subject. These patterns constitute the Quality Scale (see Table 5). In the latter group (comprising the patterns SVA, SVO, $SV_{pass}A$ and $SVCs_{+comp}$) all the verbs[23] seem to perform the presentation function in the sense that they introduce a new phenomenon on the locative scene (see Table 6). Thus, these verbs may be seen to express existence or appearance on the scene in an implicit way (Adam, 2010: 22). In previous research (Adam, 2013: 69; Rohrauer, 2013: 154) the sentences with locative thematic subjects have already been identified as a realization form of the Presentation Scale. It has been pointed out that they represent a "transitional case" of presentation sentences (Rohrauer, 2013: 154), since the subject in the locative subject construction is context-dependent and the communication is not perspectived towards the subject. Thus, in this respect the construction fails to fulfil one of the basic criteria of the Presentation Scale. However, our data suggest that what makes it possible to view these sentences as an implementation of the Presentation Scale is the "semantic content of the verb that actuates the presentation semantics of the sentence" (Dušková, 2015a: 260), as is supported by the possible paraphrase using the existential construction. Thus, the sentences with a locative subject have the potential to constitute the Presentation Scale.

So far the capability of the locative thematic subject constructions to implement the Presentation Scale due to the presentation function of the semantically empty verb (*have, bear, wear, contain*) has been described for the SVO pattern only. Nevertheless, the results of our analysis imply that the range of semantic types of verbs which can perform the presentation function may be much broader, and that they are found in various patterns. Thus, we assume that SVA, SVO, $SV_{pass}A$ and SVCs with a prepositional complement constitute a subtype of the transitional type of Presentation Scale with a context-dependent locative subject and a postverbal rhematic element introduced on the scene.

Table 5. Quality Scale patterns
Quality Scale

pattern	Bearer of Quality	Ascription of Quality	Quality
	theme	transition	transition/rheme
SV	*The room*		*reeked.*
SVCs	*The room*	*was*	*dim.*

---

[23] In the $SVCs_{+comp}$ pattern the presentation function is performed by the verbo-nominal predicate.

Table 6. Presentation Scale patterns
Presentation Scale

pattern	Setting	Presentation of Phenomenon	Phenomenon Presented
	theme	transition	rheme
SVA	*The restaurant*	*was humming*	*with life.*
SVO	*Every chair*	*had*	*an inscription.*
SV$_{pass}$A	*The room*	*was filled*	*with a dim and oppressive twilight.*
SVCs $_{(+comp)}$	*This realm*	*was alive*	*with all sorts of women...*

We hope to have shown that the potential of the sentence to convey the presentation idea is primarily based on the presentation function of the verb. In the English construction with a locative subject this function may be performed by a broad range of semantically diverse verbs that all imply the existence or appearance of a phenomenon on the locative scene. The more empty the semantics of the verb, the more likely is the verb to perform the presentation function in various contexts. Our data suggest that even verbs with a specific meaning (*glow, reek, ring, show, house*) have the potential to introduce a phenomenon in a particular context.

REFERENCES

Adam, M. (2011) "A bee buzzed across their path": Semantic affinity as a formative force in presentation scale sentences. *Ostrava Journal of Philology* 3/1, 19–32.

Adam, M. (2013) *Presentation Sentences (Syntax, Semantics and FSP)*. Brno: Masaryk University, Faculty of Education.

Anderson, S. R. (1971) On the role of deep structure in semantic interpretation. *Foundations of Language* 7/3, 387–396. Available at: http://www.jstor.org/stable/25000539.

Boye, K. (2012) *Epistemic Meaning. A Crosslinguistic and Functional-cognitive Study*. Berlin/Boston: De Gruyter Mouton.

Brůhová, G. (2015) Passivization of ditransitive verbs from the FSP point of view. *Linguistica Pragensia* 25/1, 27–36.

Chamonikolasová, J. (2010) Communicative perspectives in the theory of FSP. *Linguistica Pragensia* 20/2, 86–93.

Dowty, D. (2000a) "The garden swarms with bees" and the fallacy of "argument alternation". In: Ravin, Y. and C. Leacock (eds.), *Polysemy: Theoretical and Computational Approaches*, 111–128. Oxford University Press.

Dowty, D. (2000b) The semantic asymmetry of 'argument alternations' (and why it matters). *Groninger Arbeiten zur germanistischen Linguistik*. Available at: http://www.ling.ohiostate.edu/~dowty.

Dušková, L. (1998) Syntactic forms of the presentation scale and their differentiation. *Linguistica Pragensia* 8/1, 36–43.

Dušková, L. (1999a) On some functional and stylistic aspects of the passive voice in present-day English. In: *Studies in the English Language. Part 1*, 113–148. Prague: Karolinum.

Dušková, L. (1999b) On some semantic and FSP aspects of the subject in Czech and in English. In: *Studies in the English Language. Part 2*, 247–272. Prague: Karolinum.

Dušková, L. (2015a) The relations between semantics and FSP as seen by Anglicist members of the Prague Linguistic Circle. In: *From Syntax to Text: the Janus Face of Functional Sentence Perspective*, 256–268. Prague: Karolinum Press.

Dušková, L. (2015b) Syntactic constancy of adverbials between English and Czech. In: *From Syntax to Text: the Janus Face of Functional Sentence Perspective*, 30–45. Prague: Karolinum Press.

Firbas, J. (1992) *Functional Sentence Perspective in Written and Spoken Communication*. Cambridge: Cambridge University Press.

Fried, M. (2005) A frame-based approach to case alternations: the *swarm*-class verbs in Czech. *Cognitive Linguistics* 16/3, 475–512.

Gast, V. (2015) On the use of translation corpora in contrastive linguistics. A case study of impersonalization in English and German. *Languages in Contrast* 15/1, 4–33.

Haspelmath, M. (2010) Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86/3, 663–687.

Kettnerová, V. (2015) *Lexikálně-sémantické konverze ve valenčním slovníku*. Prague: Karolinum.

Klégr A., M. Malá and P. Šaldová (2011) Anglické nepředložkové ekvivalenty českých předložek. In: Čermák, F. *Korpusová lingvistika Praha 2011 – 1 InterCorp*, 10–28. Prague: Nakladatelství Lidové noviny.

Levin, B. (1993) *English Verb Classes and Alternations: a Preliminary Investigation*. Chicago/IL: The University of Chicago Press.

Malá, M. (2013) Translation counterparts as markers of meaning. *Languages in Contrast* 13/2, 170–192.

Martinková, M. (2014) K metodologii využití paralelních korpusů v kontrastivní lingvistice. *Naše řeč* 97/3–4, 270–285.

Mathesius, V. (1936) On some problems of the systematic analysis of grammar. *Travaux du Cercle Linguistique de Prague* VI, 95–107.

Panevová, J. et al. (2014) *Mluvnice současné češtiny 2: Syntax češtiny na základě anotovaného korpusu*. Prague: Karolinum.

Quirk, R. et al. (1985) *A Comprehensive Grammar of the English Language*. London: Longman.

Rohrauer, L. (2013) Syntactic realizations of the presentation scale in academic prose and in fiction. *AUC Philologica* 3. *Prague Studies in English* XXVI. Prague: Karolinum, 145–157.

Rohrauer, L. (2015) *Presentation Sentences in Fiction and Academic Prose: a Syntactico-semantic, FSP and textual view*. Ph.D. thesis. Prague: Charles University.

Sources:

Český národní korpus – InterCorp. Ústav Českého národního korpusu FF UK, Praha. Available at: http://www.korpus.cz [2016-07-22].

LOKATIVNÍ PODMĚTY V ANGLIČTINĚ

Resumé

Studie se zabývá anglickými větami s tematickým lokativním podmětem. Příklady byly získány pomocí paralelního korpusu *InterCorp* jako překladové protějšky českých vět s iniciálním příslovečným určením realizovaným pomocí předložkové fráze s předložkami *do, na, v/ve, z/ze*. Toto iniciální adverbiální určení odpovídá ve struktuře anglické věty tematickému podmětu s lokativní sémantikou. Studie zkoumá anglické věty ze syntaktického, sémantického a aktuálněčlenského hlediska. Ukazuje se, že české věty s iniciálním lokativním příslovečným určením odpovídají v angličtině pěti různým syntaktickým strukturám (SV, SVA, SVO, SV$_{pass}$A and SVCs). Z aktuálněčlenského aspektu se studie snaží zjistit, zda anglické věty s lokativním podmětem implementují prezentační nebo kvalifikační škálu. Práce ukazuje, že i anglické věty s lokativním podmětem lze považovat za prezentační věty. Zvláštní pozornost je věnována sémantice slovesa a vymezení dvou základních sémantických skupin sloves, která mohou ve spojení s lokativním podmětem vést k prezentační funkci.

Gabriela Brůhová
Department of English Language and ELT Methodology
Faculty of Arts, Charles University
gabriela.bruhova@ff.cuni.cz

Markéta Malá
Department of English Language and ELT Methodology
Faculty of Arts, Charles University
marketa.mala@ff.cuni.cz

ENGLISH *IT*-CLEFTS WITH FOCUSED SUBJECT FROM THE PERSPECTIVE OF FSP AND THEIR CZECH TRANSLATION EQUIVALENTS

ANNA KUDRNOVÁ

ABSTRACT

The paper offers a brief survey of relevant theory on English *it*-clefts, with an emphasis on information structure. In the practical part, it compares *it*-clefts with focused subject in original English written texts and their Czech translation counterparts, based on a sample research in the parallel multilingual corpus InterCorp. Due to the different means of expressing FSP in the two languages, Czech naturally uses other ways to convey the desired meaning apart from the cleft construction, such as word-order, focalizing particles, and other devices. To get a better idea about the extent to which the translation may be influenced by the surface structure of the original, another additional sample research was carried out using the same queries to find English translations containing *it*-clefts and their original Czech source sentences. The preliminary research suggests that at least in the chosen area, the source language does not seem to significantly interfere in the surface structure of the resultant translation.

Keywords: syntax, *it*-cleft, functional sentence perspective, translation

1. Introduction

The objective of this paper is to examine English *it*-clefts, in particular those with focused subject, in the light of their Czech equivalents, using a parallel corpus, and mainly with regard to information structure. The first part presents the theoretical issues concerning *it*-clefts and related FSP aspects, while the second half concentrates on the research: its methods, advantages and problems, and puts forward some preliminary findings. Beside the most obvious and necessary line of research, which involves finding Czech translation equivalents of English *it*-clefts, other supplementary strands of search have been executed (or are to be executed), in order to gain more insight into some translation-related issues, such as interference from the source language into the target language.

2. Motivation

There are several reasons that make a comparative, corpus-based approach a convenient method. First, as regards FSP, the *it*-cleft has one advantage: it has a relatively fixed, recognizable syntactic structure, thus it can be searched for in the corpus; whereas in most "unmarked" constructions the information structure is expressed by a combination of means, usually not overtly identifiable. Researchers are therefore often confined to manual search and comparison. The surface structure of *it*-clefts, on the other hand, enables more extensive search and a greater variety of results. Inevitably, the corpus method has its disadvantages, too, which will be discussed in more detail in the second part of the paper.

As regards the comparative approach, with two languages as typologically different as English and Czech (analytical and inflectional, respectively), translation can serve as a useful tool; rendering a sentence into another language with different FSP means often helps to see the information structure more clearly. At the same time it needs to be emphasized that the equivalents found in a parallel corpus are translations proper rather than linguistic tools, with all that this entails. In other words, the choices made in the translation may be influenced by other factors than the form of the original structure that is being analysed (for example by the tendency to avoid repeating words/constructions, maintain the cohesion of the resulting text, or other translation-related factors).

3. Theoretical part

3.1 *It*-clefts in English

The cleft construction in English is a fairly common syntactical means, one of whose primary functions is to highlight the element that carries the information focus. Since English word order is for the most part fixed, which limits the operational scope of the FSP factor of linearity, information structure must sometimes be made explicit by this secondary construction, which puts the most dynamic element in focus. Nevertheless, the *it*-cleft, like many other linguistic means, is not restrained to a single function; it often has a textual role, such as introducing a new topic, summarizing etc. (cf. Dušková, 2010: 42). In order to determine the information structure of a given *it*-cleft, it is always necessary to consider other FSP factors, too, most importantly the context (Greenbaum and Quirk, 1990: 412)

As regards the information structure of the two constituent clauses of the cleft-sentence, Prince (1978) distinguishes two types of *it*-clefts; the dependent clause of the first type contains "old" or "knowable" information, which can sometimes even be omitted. The second type, called informative-presupposition *it*-clefts, on the other hand, actually presents new information in the dependent clause (Prince, 1978: 896–898). As a result, either of the constituent clauses may contain the rheme.

3.2 Cleft sentences in Czech

Czech as an inflectional language with free word-order relies heavily on linearity to indicate FSP structure, much more than English. It follows that the occurrence of cleft constructions will be markedly lower in Czech. However, the difference is not absolute and cleft constructions are indeed used in Czech as well, although they are less frequent and likely to be perceived as more marked. Their status in the language system seems to be slightly peripheral, as they have been seen by some linguists as a syntactical borrowing from other languages, such as English or French, and its excessive and unjustified use has been frowned upon as a mark of stylistic awkwardness (Filipec, 1955). What is interesting is that in spite of the lower frequency of Czech *it*-clefts, both the types identified by Prince (stressed-focus and informative presupposition) occur; the latter type serves mainly as a stylistic device expressing attitude or involvement, or is used to emphasize a repeated structure, even though the rheme is in the dependent clause (Filipec, 1955).

3.3 Comparative perspective: advantages and challenges

Leaving purely abstract linguistic aspects aside – what are the grounds on which translation equivalents are decided on? We may quite safely assume that translators, given the amount of time that they spend close-reading the text, do in most cases determine the FSP structure of the original correctly. What is more difficult to answer is the question how in particular they choose from the several options that the target language offers, and whether the original syntactic structure affects this decision; the factors involved in the choice are numerous. In a simplified perspective: there will probably be instances in which the translator simply copies the syntactic structure of the source without considering other options. On the other hand, some translators who are aware of the fact that *it*-clefts are much rarer in Czech than in English may avoid the construction altogether in an effort to prevent undesirable interference. The tendency to opt for the more conventional or more central solutions in the target language is by some theoreticians considered to be one of the translation universals (cf. for example the discussion in Chlumská, 2015: 33). We are thus faced with two opposing trends. Needless to say that this is also a matter of personal preference and idiolect, as it is in writing. To find out the extent to which these factors influence the resulting translation, an analysis could be carried out, determining the frequency of *it*-clefts in original texts and their translations, as well as in different texts translated by the same translator, or, alternatively, in two different translations of one source text. Needless to add, the choice may also be arbitrary or influenced by the surrounding translated text rather than by the original.

4. Research

Since the subject matter is rather extensive, only a sample research was carried out. Its main goal, besides obtaining preliminary results, was to observe the main tendencies, so that the following, larger-scale research could be adjusted if necessary. Since the concordances found in the corpus are preceded by a limited number of words, it is not always easy

(or reliable) to determine the type of the *it*-cleft according to Prince's classification. Hence the strictly comparative nature of the present research. The absence of context remains the greatest obstacle to any qualitative corpus-based analysis.

4.1 Sources

The corpus used for the sample research was InterCorp, a Czech parallel multilingual corpus. In the main line of research, a subcorpus of original English texts was used, along with the relevant Czech translations. More specifically, the field was limited to fiction; the main reason for this choice being that there is a large amount of material which is at the same time very diverse (with respect to authors, the complexity of the structures, style etc.). As opposed to unpublished texts, translations of fiction are supposed to undergo some sort of editorial process, which should to some degree lower the number of mediocre translation equivalents not reflecting the difference between the two languages.

4.2 Queries and results

CQL queries were formulated with the aim of covering as many instances of *it*-clefts as possible, including interrogative sentences, modal structures, *it*-clefts whose dependent clause is introduced by zero relative (juxtaposed) instead of the more usual relatives such as *that* or *who*. Some spaces were allowed for optional elements like adverbs, predeterminers, determiners, adjectives or pronouns. In the end, three queries were formulated, as follows:

▶ Query 1: declarative *it*-clefts and polar interrogatives:
([][] containing [word="(?i)it"] containing [word="(?i)(is|was)"])[tag="RB"]? [tag="PDT"]? [tag="DT"]? ([tag="JJ"]*)? [tag="PP."]? ([tag="N.*"]+|[tag="PP"]) [word="that|who|which"]?[tag="MD"]?[tag="V.*"]+
*It is / it was / is it / was it + (adverb)(predeterminer)(determiner)(one or more adjectives)(possessive pronoun) + one or more **nouns/personal pronoun** + (that/who/which) + (modal verb)(one or more **verb** forms)*
▶ Query 2: *Wh*-questions:
[word="(i?)(what|who)"][word="is|was"][word="it"][word="that"]? [tag="V.*"]+
What / who is / was it that (one or more verb forms)
▶ Query 3: Declarative and interrogative *it*-clefts containing a modal verb:
([][] containing [word=" (?i)it"] containing [tag="MD"]) [word="have"]?[word=" (be|been)"][tag="RB"]? [tag="PDT"]? [tag="DT"]? ([tag="JJ"]*)? [tag="PP."]? ([tag="N.*"]+|[tag="PP"])[word="that|who|which"]?[tag="MD"]?[tag="V.*"]+
*It + modal + (have) be/been (including inversion) (adverb)(predeterminer)(determiner) (one or more adjectives)(possessive pronoun) + one or more **nouns/personal pronoun** + (that/who/which) + (modal verb)(one or more **verb** forms)*

Since the queries cover rather a broad spectrum of sentences, they had to be sorted manually, and many of them were found irrelevant. Most commonly, the "false positives" with an identical surface structure were relative clauses, e.g. *It was a trick swim-*

mers used to re-oxygenate their blood between tightly scheduled races. The table below illustrates the efficiency of the search in both lines of research, showing how many of the results turned out to be genuine *it*-clefts with a focused subject. Query 1 was the most productive, as could be expected: 1886 concordances were found, 200 of which were randomly selected for the analysis. The remaining two queries proved to be much less productive. In the second, additional research, Query 1 yielded 364 results (which were then limited to a random sample of 200 lines); Queries 2 and 3 displayed the same trend as in the first case.

	1st line of research (English > Czech)		2nd line of research (Czech > English)	
	total	relevant	total	relevant
Query 1	200	90	200	74
Query 2	7	4	4	0
Query 3	75	14	15	4

4.3 Control search

It can be assumed that in spite of all effort, the queries cannot and will not cover all *it*-clefts present in the texts. To find out how efficient the queries are, I decided to carry out a manual search in a fiction book, namely George Orwell's *Animal Farm*. This control search yielded exactly the same number of *it*-clefts as the combination of all three queries. However, *Animal Farm* is short, stylistically rather monotonous and consists mainly of shorter, not very complex sentences; the *it*-clefts found in the novel are, as a rule, less complex than many other from different sources. While this proves that the queries are probably quite well-suited for such simple structures, it does not necessarily mean they will be absolutely reliable in syntactically elaborate texts. A similar result of automatic vs. manual search is described in Dušková (2015). The automatic search thus does not seem to be significantly inferior to manual search.

It may be noted in passing that this additional sub-study exemplifies the advantages of a corpus-based research, as the frequency and type of *it*-clefts clearly differs from text to text, at least in a genre as diverse as fiction.

5. Lines of research

5.1 First line of research: English to Czech

The core of the analysis involves searching for English *it*-clefts and considering both them and their Czech counterparts. In the sample research, the three queries introduced in the previous section were used to obtain material for the analysis, with the number of results specified in section 4.2. The English *it*-clefts were consequently classified according to the criteria presented by Prince, and their Czech counterparts according to the means used to convey the given information structure. At this point it must be added that the chosen approach proved to be somewhat limited in the classification of English

it-clefts – the corpus often does not provide sufficient context to safely determine whether the content of the subordinate clause is given or new. For this reason, there is some degree of uncertainty in the less obvious cases.

As regards the typology of Czech translation equivalents of English *it*-clefts, as could be expected, Czech translations proved to use a variety of means to express the information structure indicated by an English cleft sentence. The following types of syntactic means were found in the sample analysis:

- Cleft construction
  (1) It was her voice that had betrayed them. (London)
      Byl to její hlas, který je prozradil.
The Czech syntactic equivalent of the English *it*-cleft; arguably perceived as more marked than in the original.

- Word-order (the rhematic element is placed at the end of the clause)
  (2) It was papa made me leave. (Hemingway)
      To mi poručil táta.
This solution makes use of the Czech word-order which is primarily governed by the FSP principle. It is possibly less stylistically marked than an *it*-cleft. (Filipec 1955)

- Focalizing particles (*to, právě, zrovna…*)
  (3) I read a book about Goya, which said it was his deafness that made him into a major artist. (Lodge)
      Četl jsem knihu o Goyovi, kde se tvrdilo, že právě hluchota z něj udělala význam-ného umělce.
Another fairly frequent solution using lexical means; focalizers or rhematizers are expressions whose primary function is to signalize categories of information structure and which are typically placed in front of the emphasized unit. Rhematizers can have various specifying meanings (Panevová et al., 2014: 221–222).

- Combination (WO + focalizer)
  (4) Bobby finally grasped that it was he who was being beckoned and slowly made his way over. (Jackson)
      Bobby nakonec pochopil, že ten muž mává právě na něj, a pomalu k němu zamířil.
The rheme is indicated by word-order and further emphasized by a focalizer, which makes the information structure more evident.

- Pronoun
  (5) It wasn't I who killed you, Albus. (Rowling)
      Já jsem tě nezabil, Albusi.
In Czech, a personal pronoun in the role of subject is normally not expressed, as verbal categories are evident from the inflectional ending of the verb. When the pronoun does appear, it is emphatic or marked in some other way (Grepl et al. 2012: 412); thus being a valid equivalent for a cleft construction.

- Other
  (6) It was the telephone that was tantalizing Starling. (Harris)
      Starlingová měla celou tu dobu chuť skočit po telefonu.

This category contains examples in which the whole clause has been reformulated in some way.

- None
  (7) It was Pop who started the song, which was We Ain't Got a Barrel o' Money, and everyone took it up in shrill voices. (Bates)
      Taťka Larkin první zanotoval Peněz tak mít jako želez a všichni se k němu pronikavými hlasy přidali.

This group includes instances in which no apparent syntactic means are used to increase the communicative dynamism of the element that has been emphasized by the English cleft; they are mostly informative-presupposition *it*-clefts. These include some peripheral examples, such as:

  (8) It may be a wise child that knows its own father, but it is a laughing child that knows its own mother. (Morris)
      Chytré dítě může poznat svého otce, ale smějící se dítě jistě poznává svou matku.

This particular type has been mentioned in literature as a special case, a "proverbial" sentence with a predicational meaning instead of specificational, which is typical of *it*-clefts (Patten, 2012: 4). The construction clearly does not serve the usual function of (primarily) expressing information structure, and the translation reflects that. Or similarly:

  (9) It wasn't every man who could afford her, so she'd get the same type again and again. (Palahniuk)
      Všichni muži si ji nemohli dovolit, a tak vídala stále dokola tytéž typy.

The above example is not proverbial, but the *it*-cleft serves a similar function: it is a stylistic device that increases markedness rather than a means to highlight information structure.

5.2 Second line of research: Czech to English

Apart from the main research on English texts translated into Czech, another, secondary search was conducted, this time on Czech sentences which were translated into English as *it*-clefts. The queries were therefore identical (and performed on English text); the only difference was the direction of translation. The purpose of this research was to find out whether the equivalent pairs in the opposite direction (from Czech to English) tend to differ significantly.

For instance, if the use of *it*-clefts in Czech translations were mostly based on the source structure, it could be expected that the frequency of Czech *it*-clefts in this direction would be much lower. The types of equivalents proved to be more or less the same as in the first study, namely:

- Cleft construction
  (10) But it is she who does the instructing!
      Ale je to ona, kdo poučuje! (Kundera, *Nesmrtelnost*)

In this case, the choice of the English equivalent is quite self-evident. In fact, it would be interesting to search for the construction in the Czech section of the parallel corpus and see if there are any English translations that do *not* use cleft construction.

- Word-order
  (11) I've thought that since I was a kid, but I realized that back then it was you who denied it.

  Myslím si to od malička, spíš mi utkvělo, že jste to tenkrát popírala vy. (Kohout)
The information structure expressed by the principle of linearity in Czech is translated by an *it*-cleft.

- Focalizing particles (*to, právě, zrovna…*)
  (12) It is the second tear that makes kitsch kitsch.

  Teprve ta druhá slza dělá z kýče kýč. (Kundera, *Nesnesitelná lehkost bytí*)

- Combination (WO + focalizer)
  (13) And it was this obedience mingled with shyness that drove me wild; when I approached her, she shrank back and covered her crotch with her hands....

  A mne vzrušovala k šílenství právě ta oddanost smíšená s ostychem; když jsem k ní přistoupil, skrčila se a zakryla si rukama klín ... (Kundera, *Žert*)

- Pronoun
  (14) One of my physicians had to leave the clinic, and it was I who asked for his resignation.

  Jeden z mých lékařů musel z kliniky odejít, sám jsem ho k tomu vyzval. (Stýblová)

- None
 (15) And so it was not the flu that kept him from sleeping with Tereza on her first visit.

  Ostatně když u něho byla Tereza poprvé s chřipkou, nikdy s ní nespal. (Kundera, *Nesnesitelná lehkost bytí*)

In some cases, no apparent syntactic device in the Czech sentence seems to require the *it*-cleft; the sentence is generally reformulated, similarly to example (9). The cleft seems to express a certain markedness.

6. Results

As explained in section 4.3, the classification of *it*-clefts was somewhat complicated due to limited context provided by the concordances. What can, however, be quantified without doubt are the proportions of FSP means used in the translation equivalents. Figure 1 presents a comparison of Czech equivalents of English *it*-clefts and original Czech constructions that were translated as *it*-clefts in English. Although the results may not be strictly comparable (the size of the randomly selected sample does not significantly differ, but no special attention was paid to the year in which the texts were written, their length and other possibly relevant factors), there are some tendencies to observe.

Figure 1.
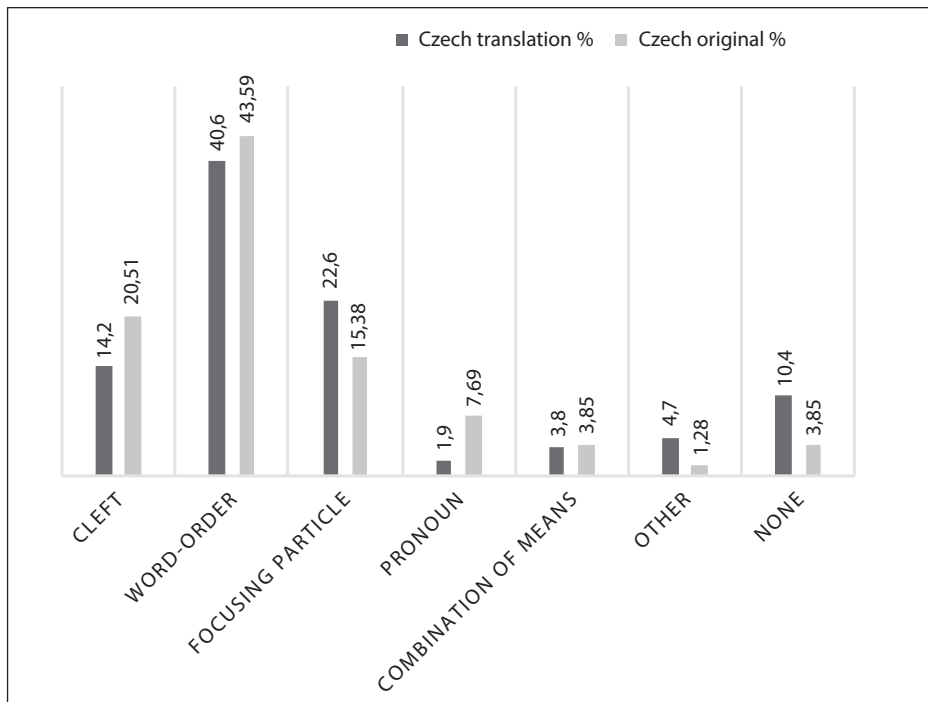
The figure reveals several evident findings: in both directions of translation, the most common Czech equivalent of an *it*-cleft is a clause in which the rhematicity of the subject is indicated by its final position (FSP principle of linearity). Focalizers and cleft constructions have approximately equal frequencies, followed by the remaining types. What is remarkable is that the differences between the two directions are not radical and the distribution of the individual categories is roughly similar. There is a lower frequency of emphatic personal pronouns in translations from English, suggesting that this way of expressing the information structure of English *it*-clefts may not be an obvious choice for Czech translators. On the other hand, focalizers/rhematizers are more frequent in this direction.

7. Conclusion

The presented preliminary study suggests that as regards fiction, translators of both languages seem to be quite reliable in choosing the appropriate FSP means characteristic of the target language, at least statistically. In other words, the translation equivalent pairs consisting of English *it*-clefts and their Czech counterparts do not display fundamental differences depending on the direction of translation. Nevertheless, this sample research in itself is not statistically very relevant, due to the lower number of examples; more equivalents will have to be analysed, and ideally with respect to other variables.

As regards further research, the results of this preliminary study show that there are still many issues that need to be addressed. The first condition of a quantitatively relevant result is that the research should be carried out on a larger scale. That, however, brings the aforementioned problem of insufficient context, which is a significant complication for the classification. For further research it will be necessary to find a suitable method to deal with this problem.

To have a better understanding of the role of cleft constructions in Czech, it may be useful to compare their occurrence in original Czech texts and those translated from English, both in terms of frequency and typology. Such a research would indicate whether there tends to be any interference from the source language, quantitative or qualitative.

REFERENCES

Dušková, L. (2010) Syntactic construction, information structure and textual role: An interface view of the cleft sentence. *Brno Studies in English* 36/1, 29–45.

Dušková, L. (2015) Textual roles of two forms of rhematic subjects: initial rhematic subjects vs. subjects rhematized by *it*-clefts. *Linguistica Pragensia* 25, 49–66.

Filipec, J. (1955) Byl to můj přítel, který… [It was my friend who…] *Naše řeč* 38/7–8, 193–198. Available at: http://nase-rec.ujc.cas.cz/archiv.php?art=4433#_ftn1 [2016-09-26].

Greenbaum, S. and R. Quirk (1990) *A Student's Grammar of the English Language*. Harlow: Longman.

Grepl, M. et al. (2012) *Příruční mluvnice češtiny*. Prague: Lidové noviny.

Chlumská, L. (2015) *Překladová čeština a její charakteristiky*. Prague: Faculty of Arts.

Panevová, J. et al. (2014) *Mluvnice současné češtiny 2: Syntax češtiny na základě anotovaného korpusu*. Prague: Karolinum.

Patten, A. (2012) *The English It-Cleft: A Constructional Account and a Diachronic Investigation*. Berlin/Boston: Mouton de Gruyter.

Prince, E. (1978) A comparison of *wh*-clefts and *it*-clefts in discourse. *Language*, 54/4, 883–906.

Sources

Český národní korpus – *InterCorp*. Praha: Ústav Českého národního korpusu FF UK. Available at: http://www.korpus.cz.

Bates, H.E. *The Darling Buds of May*
Harris, T. *The Silence of the Lambs*
Hemingway, E. *The Old Man and the Sea*
Jackson, M. *Five Boys*
Lodge, D. *Deaf Sentence*
London, J. *At the Rainbow's End*
Morris, D. *The Naked Ape*
Orwell, G. *Animal Farm*
Palahniuk, C. *Choke*
Rowling, J. K. *Harry Potter and the Deathly Hallows*

Kohout, P. *Sněžím*
Kundera, M. *Nesmrtelnost*
Kundera, M. *Nesnesitelná lehkost bytí*
Kundera, M. *Žert*
Stýblová, V. *Skalpel, prosím*

ANGLICKÉ KONSTRUKCE *IT*-CLEFTS S VYTČENÝM PODMĚTEM Z POHLEDU AČV A JEJICH ČESKÉ PŘEKLADOVÉ EKVIVALENTY

Resumé

Příspěvek představuje stručný přehled teoretických aspektů anglických vytýkacích konstrukcí s důrazem na aktuální členění větné. V praktické části se srovnávají vytýkací věty s vytčeným podmětem s jejich českými překladovými protějšky na základě vzorku získaného z paralelního vícejazyčného korpusu Inter-Corp. Vzhledem k tomu, že každý z obou jazyků používá jiné aktuálněčlenské prostředky, čeština požadovaný význam kromě vytýkací konstrukce nutně vyjadřuje i jinak, například slovosledem, zdůrazňovacími částicemi atd. Pro lepší představu o tom, nakolik může být překlad ovlivněn povrchovou strukturou originálu, byl proveden další vzorový výzkum se stejnými dotazy, pomocí nichž byly vyhledány anglické překlady obsahující vytýkací konstrukci a jim odpovídající české originály. Tyto předběžné výzkumy ukazují, že přinejmenším ve zvolené oblasti zřejmě nenastává výrazná interference zdrojového jazyka do výsledného textu.

Anna Kudrnová
Department of English Language and ELT Methodology
Faculty of Arts,Charles University
anna.kudrnova@gmail.com

WHAT COMES FIRST, WHAT COMES NEXT: INFORMATION PACKAGING IN WRITTEN AND SPOKEN LANGUAGE

VLADISLAV SMOLKA

ABSTRACT

The paper explores similarities and differences in the strategies of structuring information at sentence level in spoken and written language, respectively. In particular, it is concerned with the position of the rheme in the sentence in the two different modalities of language, and with the application and correlation of the end-focus and the end-weight principles.

The assumption is that while there is a general tendency in both written and spoken language to place the focus in or close to the final position, owing to the limitations imposed by short-term memory capacity (and possibly by other factors), for the sake of easy processibility, it may occasionally be more felicitous in spoken language to place the rhematic element in the initial position or at least close to the beginning of the sentence.

The paper aims to identify differences in the function of selected grammatical structures in written and spoken language, respectively, and to point out circumstances under which initial focus is a convenient alternative to the usual end-focus principle.

Keywords: word order, functional sentence perspective, end-focus, written language, spoken language

1. Introduction

Linearity, one of the essential properties of language, means that components of language structures can only be arranged in a linear sequence. At sentence level, this arrangement may be described as word order, or, more fittingly, as clause constituent order. The ordering of the constituents is not accidental but governed by a set of principles whose relative importance varies across languages and which serve a variety of purposes, such as indicating grammatical structure, semantic relationship between constituents, achieving an intended distribution of information over the sentence, emphasising particular elements, creating cohesion, making sentences easy to understand, etc.

Owing to the typological characteristics of the current English language, the main function of word order is to indicate the grammatical (syntactic) status of an element as

a clause constituent. As a result, positional mobility of clause constituents within sentences is severely constrained; a canonical sequence in a declarative sentence starts with the subject and continues with the verb followed by one or more postverbal constituents, such as the object (or the subject complement) and the adverbial – a sequence known in short as SVOMPT.

The respective principles governing the linear arrangement of constituents may work in agreement or contradict one another. In the latter case, the actual linear arrangement of the sentence results from their relative hierarchy, that is, how strongly they assert themselves in a given communicative situation. Viewed in this perspective, the grammatical principle of standard sequence of clause constituents often works in harmony with the tendency to put the most important information at or close to the end of the sentence, the end-focus principle (Quirk et al., 1985: 1356–1357), and with the tendency to position short, structurally light structures before longer, structurally heavier ones, the end-weight principle (Quirk et al., 1985: 1361–1362). This is especially true in the case of the subject versus the other clause constituents: the subject, the only constituent whose canonical position is preverbal, tends to be shorter and less informative than the postverbal constituents (object, subject complement). However, the subject may occasionally be long or informationally important, yet this does not necessarily mean that it can be removed from the initial position.

Additionally, the principles of end-focus and end-weight are understood as naturally correlated: it seems to be a safe assumption that structurally heavy constituents (relatively longer and syntactically more complex) tend to convey more information than short ones. Moreover, the longer constituents, expressed in full, can be expected to convey new information, while the short ones, expressed by pro-forms, are associated with old information. This makes the structurally heavy elements potentially focal/rhematic. Still, even this correlation does not apply absolutely, as can be seen from the following examples (Huddleston and Pullum, 2002: 1371).

(1) One of his daughters was running a computer store, while the other was still at university, reading law. <u>The one running the computer store</u> earned a lot of money.
(2) There's a <u>toad</u> in the large pool outside the barn.

In (1) the underlined subject of the second sentence is long but thematic (conveying context-dependent information) and remains in the canonical initial position, breaking the principle of end-weight, while conforming both to the grammatical word order and the principle of end-focus. Conversely, the rhematic subject in (2), used within the existential *there is* construction, conveys new, context-independent information and, though itself short, occupies a postverbal position, conforming partly to the principle of end-focus; in this example the linear distribution of clause constituent results from the application of the existential construction.

The relative strength of the principle of standard distribution of clause constituents in English means that the position of the rheme (focus) depends to a large extent on the grammatical structure of the sentence. Firbas (1992: 66–69) distinguishes between two essential arrangements of sentences on semantic grounds: the presentation scale and the quality scale (cf. Adam, 2013: 45–46). The presentation scale perspectives the sentence

towards the subject as the most dynamic element (the phenomenon to be presented), whereas the quality scale perspectives the sentence away from the subject (the bearer of the quality) towards the quality itself, which is more dynamic than the subject. Consequently, owing to the position of the subject in English, the vast majority of sentences representing the quality scale are characterised by a post-verbal rheme, while in those representing the presentation scale, the position of the rheme depends on the structural subtype employed: it is postverbal, though not necessarily final, in the frequent *there is* construction, final in sentences employing initial thematic adverbial followed by S-V inversion, while initial rheme is the necessary consequence of application of structures involving rhematic subjects and conforming to the standard linear distribution of clause constituents. It is to be noted that although the *there is* structure represents the most frequent subtype of presentation scale implementation, the range of predicate verbs it can employ is rather limited, while rhematic preverbal subjects readily combine with a much larger set of predicate verbs.

2. Information-packaging strategies

Despite the limited mobility of clause constituents, a change of rheme position can be achieved in English by a number of strategies referred to as information-packaging strategies (Huddleston and Pullum, 2002: 1366):
– simple word-order change (reordering), which is limited in English:
  – preposing: *This one she accepted*.
  – postposing: *I made without delay all the changes wanted*.
  – inversion: *On board were two nurses*.
– realignment (alternative pairing of syntactic functions with semantic constituents in the sentence):
  – existential: *There is a frog in the pool*.
  – extraposition: *It is clear that he's guilty*.
  – left dislocation: *That money I gave her, it must have disappeared*.
  – right dislocation: *They're still here, the people from next door*.
  – cleft: *It was you who broke it*.
  – passive: *The car was taken by Kim*.
To this list may be added one further category, provisionally called reassignment, which, although related to those listed under realignment, goes further away from the default version of the sentence by replacing the original verb with one allowing an alternative assignment of syntactic functions to semantic units. So, for example, the existential construction *There are a large number of photographs in the book* may be reformulated as follows:

(3) The book contains a large number of photographs. (BNC: HHN 295)

Compared to the *there is* construction, the syntactic functions of the nominal constituents have changed: the original adverbial (*in*) *the book* has been changed into the subject, and the notional subject *a large number of photographs* became the object, while

the semantic function of *the book*, the locative, has been preserved, regardless of its syntactic status.

With respect to the principles and operations described above, two comments must be made: while changes in constituent position may be prompted by an effort to arrive at a different linear distribution of information over the sentence, it does not mean that a movement of a constituent automatically leads to a change in its information status. In reality, the word order is the weakest of the indicators of FSP, following the context, semantics, and, in spoken communication, the prosody.

Secondly, owing to the limited mobility of clause constituents in English, the end-focus principle cannot be understood as literally referring to the absolutely final position in the sentence; rather, the end has to be interpreted in the sense of postverbal, near-final position. The truly final position is often occupied by thematic elements following a postverbal rheme. In such cases, preposing (fronting) of such thematic elements, if possible, does not affect the postverbal position of the rheme, but rather makes it more explicit.

(4) I read <u>Dr Zhivago</u> eight years ago, an English edition a friend found in Yugoslavia. (BNC: AK4 718)
(5) Eight years ago I read <u>Dr Zhivago</u>, an English edition a friend found in Yugoslavia.

The intuitive observance of the end-focus principle by native speakers is probably motivated by processibility concerns: the (near-) final position of an element is a clue for the reader/listener indicating communicative importance. Yet, considering the pragmatic differences in constructing and processing written and spoken language, it is not unreasonable to assume that the strategies of rheme placement might differ in the two modalities. Biber et al. (1999: 1067) give three interrelated principles applying to online production of spontaneous spoken language, all of them affecting the manner in which spoken language is structured: keep talking; limited planning ahead; qualification of what has been said. What, among other things, follows from these general principles is that the structures actually used in spontaneous spoken language tend to be short and potentially imperfect or incomplete, and, consequently, have to be elaborated on and modified retrospectively. According to the authors, the end-weight principle asserts itself more extremely in spoken than in written language, because subjects in spoken language tend to be very short and often consist of a single word (ibid.). Generalisations concerning the application of the end-focus principle in spoken language are much more difficult to make. Nevertheless, Biber et al. suggest that retrospective modification of the message involves tagging on as an afterthought some elements which, in a logically structured and integrated sentence, would have been placed earlier (ibid.). If this is interpreted as reference to the end-focus principle, it is possible to assume that spoken language may occasionally employ initial or near-initial focus. If this is so, the most informative element occurs relatively early in the sentence, while elements carrying less important information follow.

Generalisations about the differences between spoken and written language are made difficult by the fact that, unless the manner of presentation is used as the only criterion, the boundary between the two modalities may be difficult to draw. Neither spoken nor written language are homogeneous systems, but rather constitute a scale with a large

overlapping territory in between. The essential factor here seems to be the amount of planning: a public speech, though presented orally, definitely involves more planning than a spontaneous written memo. Consequently, the choice of structures and the linear organisation of constituents are likely to be affected by planning constraints.

Another consideration of the spoken-written distinction touches upon the signals of importance available in the respective modalities. While written language has to rely on the interplay of three non-prosodic factors (context, semantics, and linearity) (Firbas, 1992: 51), spoken language can use prosody, namely the position of the intonation nucleus, either to confirm the information structure suggested by the non-prosodic factors, or to re-evaluate it. In the latter case, prosody overrides all the other factors combined. What follows from this is that in spoken language there should be less need to use deviations from the usual grammatical linear arrangement of clause constituents as an indicator of rhematicity, since the rhematic item is sufficiently marked by intonation.

This assumption is corroborated by quantitative data (Biber et al., 1999: 909–910) suggesting that, compared with academic prose and fiction, fronting in conversation of a core constituent is three to four times less frequent, and the majority of examples represent object fronting, while predication fronting is extremely rare.

3. A corpus-based study of two information-packaging constructions

To examine the use and frequency of restructuring/realignment operations in authentic samples of written and spoken language, respectively, two constructions have been examined as they occurred in the British National Corpus: the *there is* construction, and the tough movement.

3.1 *There is* construction

The *there is* construction typically serves as a means of introducing into communication a new, context-independent nominal element functioning as the notional subject of the sentence. This subject is almost invariably rhematic, unless the sentence is repeated to create a contrast, representing a case of second instance (Firbas, 1992: 111). The most common subtype of the *there is* construction, the existential-locative construction, includes a thematic adverbial of place, which usually follows the subject, but can also be fronted, particularly when it conveys context-independent information, constituting the most dynamic variant of diatheme (Firbas, 1992: 81). The position of the adverbial does not affect the rhematic status of the subject; however, a sentence with a fronted adverbial corresponds more closely to the end-focus principle, because the notional subject occupies the absolutely final position. The purpose of the research was to compare the proportion of *there is* constructions with fronted adverbials in written and in spoken communication. The hypothesis was that there would be more instances of fronting in written language.

According to Adam (2013: 62–69), the existential construction represents the most frequent subtype of presentation scale implementation (65%), which is more than the other three types combined: rhematic subject in the preverbal position (25%); front-

ed adverbial followed by S-V inversion (8%); and locative thematic subject followed by rhematic object (2%). Adam's quantitative data are based on two types of texts, fictional narrative and biblical texts, neither of which can be considered truly spoken language. Additionally, the frequency of the respective implementation subtypes varies even in the two above-mentioned text types, with the proportions indicated representing the overall distribution in both. The prevalence of the *there is* construction may be explained by the fact that it is mostly used with the verb *be*, whose existential semantics makes it ideally suited for the purpose of presentation, while other verbs are of marginal importance. Conversely, the grammatically canonical presentation construction employing a rhematic subject in preverbal position usually must employ a verb where the meaning of existence or appearance is an extension of other semantic features, and in most cases cannot be used with the predicate *be* as an alternative to the *there is* construction, unless accompanied with the subject-verb inversion:

(6) All of a sudden, there was a picturesque castle on the hill. (Adam, 2013: 59)
(7) All of a sudden, a picturesque castle appeared on the hill.
(8) All of a sudden, on the hill was a picturesque castle.
(9) ?All of a sudden, a picturesque castle was on the hill.

Although Adam does not include (9) among the model presentation sentences, it would be marginally acceptable because the adverbial *all of a sudden* renders the structure dynamic, corresponding in meaning to the semantics of the verb *appear*. Without the initial adverbial, and therefore interpreted statically, the sentence would be infelicitous, while the version employing the verb *appear* is still perfectly acceptable:
(10) *A picturesque castle was on the hill.
(11) A picturesque castle appeared on the hill.
In the course of the present research, the *there is* construction proved to be a frequent structural pattern in BNC overall, with 245,828 hits in 3,922 different texts amounting to a total of 98,313,429 words; i.e. a frequency of 2,500.45 instances per million words. The frequency was significantly higher in spoken (4,038.96 instances per million words) than in written language (2,318.26 instances per million words).
Three hundred examples of the *there is* construction were retrieved from the written and the oral part of the BNC, respectively, and the relative proportions of three subtypes were established: with fronted adverbial, with non-fronted adverbial, and without an adverbial explicitly present. The results are given in the Table below.

Table 1. Distribution of fronted/non-fronted adverbials in spoken and written texts

there is	total		fronted Adv		non-fronted Adv		no Adv	
	absolute	per cent	absolute	per cent	absolute	per cent	absolute	per cent
written	300	100%	59	19.7%	49	16.3%	192	64%
spoken	300	100%	12	3%	75	25%	213	71%

One surprising finding is the generally high proportion of *there is* constructions without an explicit adverbial. This suggests either a high frequency of the purely existential

subtype, or, possibly, cases of ellipsis of the adverbial, which is explicitly present in the preceding context, or at least implied by it. On the other hand, the difference in the frequency of structures without an adverbial is not strikingly different in spoken and written material. What is sharply different, however, is the proportion of adverbial fronting, which is negligible in spoken language, while amounting to roughly one fifth of the examples in written language. The disproportion would be even more prominent were it not for the fact that a single short section of spoken language contained three examples of fronted adverbials used within a repeated pattern.

(12)   Then within the teams, <u>in the Policy Team there were</u> three Policy Development Officers, <u>in the Neighbourhood Development Team there were</u> two Development Officers and <u>in the Community Development Team there are</u> – one, two – there are – well there are above these six officers that you could i-, that you could identify, but there are other people that relate to them, but then it gets a bit complicated so we keep it at that. (D95 428)

Another characteristic of many examples retrieved from the spoken part of the corpus is vagueness of reference, with the adverb *there* used repeatedly both as the formal subject and a referential expression pointing to some spatial setting within the same sentence.

(13)   All I do know is that we're very conscious a great deal of money's been spent by the Council and by the Trust to try and make the place outside more inviting those tubs been put <u>there</u> er they been planted etc there's a lot of litter <u>there</u> we've put litter bins <u>there</u> <u>there</u>'s taxi rank <u>there</u> <u>there</u>'s lot's of things been put <u>there</u> I think the problem is it isn't the people who do that... (D91 758)

Moreover, some of the adverbials in the initial position are adverbials of indefinite frequency or conjuncts, rather than fronted adverbials of place, which may follow in the postverbal position in the same sentence.

(14)   <u>Sometimes</u> there's a good write up beforehand sometimes there's nothing. (D91 636)
(15)   <u>Then</u> there's a lady over here who comes or used to come on a regular basis a couple of times a month she comes now a couple of times a year. (D91 616)

Conversely, examples representing written language often include lengthy fronted adverbials, sometimes two at a time.

(16)   <u>In other political states during the twentieth century</u>, there has been strong support for the view that art should serve a social purpose. (A04 307)

The findings suggest that adverbial fronting may serve different purposes in written and spoken language. As shown in the Table above, in written language fronted adverbials slightly outnumbered non-fronted ones (19.7% to 16.3%). Although the use of some initial adverbials, especially short ones, serves the purpose of textual connection, creating

cohesion of the text, it is still reasonable to assume that, in written language, a proportion of adverbial fronting in *there is* constructions is used intentionally to boost the rhematic status of the notional subject by position (end-focus) in addition to its rhematicity already resulting from the *there is* construction. On the other hand, the conspicuous scarcity of fronted adverbials in the *there is* construction in spoken language may be due to an unconscious tendency to present the most important information early in the sentence (near-initial focus). Although the *there is* construction is traditionally interpreted as a means of shifting a rhematic subject into the postverbal position, i.e. removing it from the beginning of the sentence, strictly from the point of spoken language it can be perceived as a short signal of communicative importance of the subject: *there is* tends to be pronounced as a single unstressed syllable with a reduced vowel, or, in the case of *there are/were*, as two unstressed syllables: in both cases a unit of extremely low prosodic prominence. In some respects, it is reminiscent of the traditional English rhythmic pattern used in poetry, the iambic metre (Crystal, 1997: 74; Halliday, 1989: 49–50), as well as the well-known English tendency to start sentences with rhythmically light words (Mathesius, 1975: 159).

3.2 Tough movement

Another syntactic pattern explored in BNC from a quantitative perspective was the tough movement.

(17)   ... and my word he was difficult to follow! (KC0 7824)

Unlike the *there is* construction, where the subject is almost invariably rhematic, the tough movement is less transparent with respect to FSP: the rheme may be either in the main clause (typically carried by the postverbal adjective) or in the subordinate clause (the infinitive), while the subject of the construction is typically thematic; the construction is also known as the object-to-subject raising. In fact, the thematicity of the original object of the infinitive that has become the subject is one of the main reasons for the use of the tough movement; English sentences tend to start with informationally light, context dependent elements, and such elements naturally occupy the initial position if they are encoded as subjects. The tough movement may be interpreted as an alternative to structures involving a subject *that*-clause, either in the usual extraposition or, much more rarely, in the initial position.

(18)   ... it was difficult to follow him!
(19)   ... to follow him was difficult!

However, while the non-extraposed infinitive subject clause in (19) is unambiguous in terms of FSP, the rheme being the subject complement *difficult*, the extraposed variant in (18) is open to both interpretations: the rheme is either the adjective *difficult* or the infinitive *to follow*, depending largely on the contextual factor. The FSP interpretation of tough movement follows the same principles: the component conveying context-independent information will be likely to constitute the rheme. Analysis of tough movement

constructions in the corpus revealed a prevalence of rhematic infinitives over rhematic subject complements. In this respect, the tough movement may be considered as a structure implementing the end-focus principle, especially in comparison with the extraposed subject *that*-clause.

Accordingly, the BNC was searched for structures containing the following pattern: *be + hard/easy/difficult/impossible + infinitive*. The assumption was that these adjectives were among the most frequent ones occurring in tough movement (Biber et al., 1999: 728). The search yielded a combination of *it*-extraposition, tough movement, and possibly some other accidental structures. Consequently, the examples retrieved from both the written and the spoken component of BNC were analysed manually and the number of actual tough movement constructions was established in each. The number of both of the *it*-extraposition and the tough movement combined, and the actual number of tough movement constructions was then recalculated in relation to the whole written and spoken parts of the corpus, with the following results.

Table 2. Frequency of tough movement constructions in spoken and written texts

	Number of words	Number of *it*-extraposition + tough movement	per million words	Number of tough movement alone	per million words
written	87,903,571	5,615	63.88	2,455	27.92
spoken	10,409,858	265	25.46	116	11.14

Admittedly, owing to the method used to retrieve the examples, which included search for lexical items, the results are necessarily incomplete. Still, the differences in the relative frequency of tough movement in spoken and written language are prominent enough to suggest systematic differences. If tough movement is considered as a means of implementing the end-focus or near-end focus, corpus findings provide a clear indication that written language employs it to a much higher degree than spoken language.

4. Conclusions

Using the BNC as a source of data for the analysis proved to be a double-edged sword: on the one hand, it provided a large sample of authentic material, increasing the statistical validity of findings, on the other hand, the sheer volume of raw data turned out to be a considerable processing challenge. Clearly, general differences in information structuring in written and spoken language cannot be reduced to differences identified in the use of the two syntactic constructions explored. However, as both represent fairly frequent structural patterns, and as the differences in their distribution are sufficiently prominent, it is reasonable to assume that the findings do corroborate the initial hypothesis that spoken language uses fewer word-order modifications to conform to the end-focus principle than written language. Linear arrangement of clause constituents with the rheme at or close to the end may therefore be considered a product of planned production, which is characteristic of written language. Conversely, spoken language follows the end-focus principle less consistently, partly because of lack of planning time, partly because it uses

prosody as a signal of communicative importance of a given element, regardless of its position in the sentence.

Further research will be needed before the tendencies identified can be considered conclusive. Such research might focus on the distribution of other FSP-significant syntactic structures in the two modalities of language and should address issues like the differences in the length and complexity of clauses (clausal units) in written and spoken language, exploring the assumption that longer structures yield more potential for word-order and FSP modifications than shorter ones.

REFERENCES

Adam, M. (2013) *Presentation Sentences* (*Syntax, Semantics and FSP*). Brno: Masaryk University.
Biber, D., S. Johansson, G. Leech, S. Conrad and E. Finegan (1999) *Longman Grammar of Spoken and Written English*. Harlow: Longman.
Crystal, D. (1997) *The Cambridge Encyclopaedia of Language*, 2nd ed. Cambridge: Cambridge University Press.
Firbas, J. (1992) *Functional Sentence Perspective in Written and Spoken Communication*. Cambridge: Cambridge University Press.
Halliday, M. A. K. (1989) *Spoken and Written Language,* 2nd ed. Oxford: Oxford University Press.
Huddleston, R., G. K. Pullum et al. (2002) *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
Quirk, R., S. Greenbaum, G. Leech and J. Svartvik (1985) *A Comprehensive Grammar of the English Language*. London: Longman.
Mathesius, V. (1975) *A Functional Analysis of Present-Day English on a General Linguistic Basis.* Prague: Academia.

Sources
British National Corpus (BYU-BNC). Available at: http://corpus.byu.edu/ [2016-08-29].

CO DŘÍVE A CO POTOM: INFORMAČNÍ STRUKTURA PSANÉHO A MLUVENÉHO JAZYKA

Resumé

Základní funkcí slovosledu v současné angličtině je signalizovat syntaktické funkce jednotlivých konstituentů pomocí jejich standardního lineárního uspořádání. Tím je ovšem významně omezena variabilita slovosledu. Ostatní principy uspořádání větných členů se proto mohou uplatnit pouze v omezené míře, případně jinými prostředky než pouhou změnou slovosledu v rámci stejné syntaktické konstrukce, jak je to obvyklé v češtině.

Vedle gramatického principu slovosledu v angličtině fungují dva další obecné principy uspořádání, tzv. end-weight a end-focus, tedy tendence řadit jednotlivé konstituenty od kratších a strukturně jednodušších k delším a strukturně složitějším, a od informačně méně zatížených (tematických) k informačně nejdůležitějším (rematickým).

Cílem výzkumu založeného na analýze autentických příkladů z British National Corpus bylo zjistit, zda se princip end-focus uplatňuje stejným způsobem v psaném a mluveném jazyce. K tomuto účelu byly vyhledány příklady dvou běžných syntaktických konstrukcí: existenciální konstrukce *there is* a tzv. tough

movement. První z nich slouží k uvedení rematického podmětu v postverbální pozici, a to bez ohledu na umístění příslovečného určení, které se vedle podmětu v této konstrukci typicky vyskytuje, druhá umožňuje konstruovat tematický prvek jako podmět, tedy dosáhnout základního rozložení výpovědní dynamičnosti s rematickým prvkem na konci.

U existenciální konstrukce bylo zjištěno, že pokud se v ní explicitně vyskytuje příslovečné určení, je v psaných textech poměr jeho koncové a počáteční pozice téměř vyrovnaný, kdežto u mluveného jazyka je kanonická koncová pozice příslovečného určení téměř osmkrát častější ve srovnání s počáteční pozicí. Zdá se tedy, že v psaných textech se tendence k uplatnění principu end-focus u této konstrukce projevuje mnohem silněji než v textech mluvených.

Frekvence konstrukce tough movement se ukázala být téměř třikrát vyšší v jazyce psaném než mluveném. Je-li tato konstrukce chápána jako prostředek dosažení lineárního růstu výpovědní dynamičnosti, svědčí to rovněž o silnějším uplatnění principu end-focus v psaném jazyce. Společným jmenovatelem obou zjištění může být skutečnost, že zatímco v mluveném jazyce funguje jako důležitý indikátor intonace, která je schopna převážit nad ostatními neprozodickými indikátory FSP, v psaném projevu má při absenci intonace autor potřebu naznačit rematičnost větného členu jeho koncovým umístěním.

Vladislav Smolka
Department of English Studies
University of South Bohemia in České Budějovice
smolka@pf.jcu.cz

SYNTACTIC AND FSP ASPECTS OF FRONTING AS A STYLE MARKER

LIBUŠE DUŠKOVÁ

ABSTRACT

The paper examines contextual and emphatic fronting in academic prose, fiction narrative and fiction dialogue in order to ascertain whether the types of fronting can serve as a style marker. The differences in the distribution and their effect on style are assumed to be connected with the respective FSP structures: in emphatic fronting the fronted element is the rheme, whereas in contextual fronting it is the diatheme. Hence emphatic fronting displays a prominent deviation from the basic distribution of communicative dynamism, whereas contextual fronting achieves agreement with it. As compared with the unmarked postverbal ordering, emphatic fronting intensifies the emphatic/emotional character of the content being expressed, which is a feature of speech, while contextual fronting serves as a direct link with what precedes, hence contributes to textual cohesion, which is a characteristic of academic prose, with fiction narrative presumably occupying an intermediate position. The results of the study show more types of fronting with diversified structures and less clear-cut relations between the types of frontings and the examined text sorts.

Keywords: style marker, contextual fronting, emphatic fronting, functional sentence perspective, academic prose, fiction narrative, fiction dialogue.

1. Introduction

The subject of this paper is the fronting of clause elements in three different text sorts, examined with the aim of finding out whether different types of fronting can serve as a style marker. Attention is primarily paid to contextual and emphatic fronting, which are examined in academic prose, fiction narrative, and fiction dialogue representing conversation, on the assumption that the two types of fronting will display different distribution in these text types. Support for it is found in Biber et al. (1999: 910), cf.

"Academic prose favours a dispassionate form of expression, so does not require devices that convey special emphases. However, it does put a premium on explicitness of cohesion, which may be enhanced by predicative fronting. ...

Conversation and fiction, by contrast, strive for greater impact and stylistic effect, so we find types of fronting which chiefly convey special emphasis and contrast ...”

In this paper, the differences in the distribution of fronted elements are considered from the viewpoint of their characteristics at the level of functional sentence perspective (FSP henceforth). In particular, attention is paid to the linear arrangement of the carriers of FSP functions within the respective communicative fields, examined with regard to the basic distribution of communicative dynamism (the ordering theme – transition – rheme) and the stylistic effects involved. In contextual fronting the fronted element is context-dependent and constitutes a component of the thematic section, mostly the diatheme. Accordingly, at the clausal level of the communicative field contextual fronting achieves agreement with the basic distribution of communicative dynamism (CD).[1] In emphatic fronting the fronted element is context-independent; in the underlying regular ordering it represents the post-verbal rheme. Hence emphatic fronting, at the clausal level, displays a prominent deviation from the basic distribution of communicative dyna-mism – the basic position of the rheme in English being final or pre-final. This imparts to the rheme emphatic or emotive intensification. The connection of these types of fronting with different text sorts follows from the above quotation from Biber et al. The fronted element in contextual fronting has a linking function: it serves as a direct link between what precedes and what follows, which contributes to cohesion, a feature of academic prose and also of narrative parts of fiction. On the other hand emphatic fronting involves emotion and emphasis, which are to be sought in fiction, primarily in dialogue.

2. Material and method

Of the types of fronting described in the literature (Biber et al., 1999: 900–908; Hud-dleston and Pullum, 2002: 1366–1382; Quirk et al., 1985: 1377–1379) the following treat-ment is confined to non-clausal realization forms of the fronted elements. The elements primarily involved are integrated clause elements that regularly occupy post-verbal posi-tion, viz. objects, subject complements (the object complement did not occur) and adver-bials. However, a problem arose in connection with conjuncts, as in the case of some of their semantic roles, especially temporal, a distinction between conjuncts and adjuncts is often difficult to maintain. Moreover, even disjuncts have been shown to have a cohesive potential (cf. Pípalová, 2000). Considering that both conjuncts and disjuncts play a role in stylistic differentiation of text sorts, these two categories were included in the excerp-tion as a supplement to the detailed treatment of integrated fronted elements.

Owing to the syntactic and FSP diversity of the points under study, excerption of the research material had to be done manually, with the aid of digitalized texts. Each text sort is represented by two samples drawn from two different sources: the sources of aca-demic writing were a sociobiological text, E. O. Wilson, *On Human Nature*, and a text on psychology, B. R. Hergenhahn and Tracy B. Henley, *An Introduction to the History of Psychology*; the sources of fiction narrative and fiction dialogue (stylized conversation)

---

[1]   Where the thematic section contains more elements, initial placement of the diathematic element involves a deviation from the basic distribution of CD within this section (cf. Svoboda, 1981).

were David Lodge's *Changing Places* and P. D. James's *Devices and Desires* (see Sources). Delimitation of the text sorts had been assumed to be a straightforward, formal matter; however, that was the case only in academic prose. Here the only segments left out of account were the biographical data after the names of the discussed philosophers. As regards the two text sorts drawn from fiction, dialogue – delimited graphically by inverted commas – presented the problem of how to treat the reporting clauses. The most problematic point of the narrative part appeared in the streams of thought (free indirect and direct speech in Quirk et al., 1985: 1032–1033). Both the reporting clauses and streams of thought were left out of account, reporting clauses on the ground of forming one unit with direct speech, and streams of thought owing to their specific features that assign to them the status of a sub-sort within fiction.

The measure of the frequency of occurrence of the fronted elements was sought in the number of words needed to obtain 50 fronted elements from each sample, i.e. altogether 300 instances equally drawn from each source.

The distribution of the fronted elements (including conjuncts and disjuncts) in the three text sorts, academic writing, fiction narrative and fiction dialogue, is shown in Table 1.

Table 1. Distribution of clauses with a fronted element in academic prose, fiction narrative and fiction dialogue

	Number of clauses with a fronted element	Number of words needed	Number of words per one fronted element
Academic prose			
Wilson	50	4774	95.5
Hergenhahn and Henley	50	2414	48.3
Total	**100**	**7188**	**71.9**
Fiction narrative			
James	50	4511	90.2
Lodge	50	6493	129.8
Total	**100**	**11004**	**110.04**
Fiction dialogue			
James	50	6768	135.3
Lodge	50	4426	88.5
Total	**100**	**11194**	**111.94**

Table 1 shows a clear-cut distinction between academic writing on the one hand – nearly 72 words per one fronted element, i.e. a relatively high frequency of occurrence, and fiction narrative and fiction dialogue on the other, respectively, 110 and nearly 112 words per one fronted element, i.e. a notably lower relative frequency of occurrence. However, the two samples of each text sort display smaller or greater differences. These are partly to be ascribed to the subject matter – the content of the respective part of the text and the sub-register of the text within the particular text sort, and partly to the individual styles of the authors.

The greatest difference in this respect is found between the two samples of academic text, where the number of words needed for one fronted element is nearly twice as large in Wilson as in Hergenhahn and Henley. The two fiction samples, narrative and direct speech, quantitatively correspond in a criss-cross pattern, Lodge's narrative being comparable with James's dialogue and vice versa. This is obviously the cause of the similar distribution of fronted elements in these samples. Moreover, the narrative James sample is closely comparable with the academic Wilson sample.

3. Classification and distribution of the fronted elements in the three text sorts

3.1 Academic sample

The distribution of the syntactic functions of fronted elements in the academic sample is shown in Table 2.

Table 2. Syntactic functions of the fronted elements in the academic sample

	Wilson	Hergenhahn and Henley	**Total**
Conjuncts	9	25	34
Disjuncts	1	5	6
Subtotal	**10**	**30**	**40**
Adverbials	38	13	51
Object	1	4	5
Other clause elements	1	3	4
Subtotal	**40**	**20**	**60**
Total	50	50	100

Differences between the two sources of the academic sample shown by larger numbers of instances appear in the distribution of conjuncts and integrated adverbials. Conjuncts are notably more frequent in the psychology sample, while integrated adverbials largely predominate in the sociobiological text.

3.1.1 Wilson

As shown in Table 2, all but two fronted elements in the Wilson sample are accounted for by adverbials. The FSP functions of adverbials are generally connected with their semantic roles, temporal and locative adverbials being semantically disposed to operate within the theme. These two semantic roles account for more than a half of all fronted adverbials in this sample: there were 10 temporal + another 3 in the examples with two fronted adverbials, and 9 locative, the latter including 4 instances with inversion. Altogether, temporal and locative adverbials account for 22 out of 38 instances of fronted adverbials, i.e. 57.8%.

The disposition of these two semantic roles to constitute scene-setting elements was clearly shown in the case of temporal fronted adverbials: all of them were components of the theme where they served as links between what precedes and what follows, cf. (1) a. and b.

(1) a. [They devote lifetimes to special entities and patterns] and <u>during the early period of exploration</u> they doubt that these phenomena can be reduced to simple laws.

<div align="right">Wilson, p. 7</div>

b. [Cytology forced the development of a special kind of chemistry and the use of a battery of powerful new techniques, including electrophoresis, chromatography, density-gradient centrifugation, and x-ray crystallography.] <u>At the same time</u> cytology metamorphosed into modern cell biology.

<div align="right">Wilson, p. 9</div>

This was also partly the case of locative adverbials, cf. (2).

(2) [The societies of wasps, bees, and ants have proved so successful that they dominate and alter most of the land habitats of the Earth.] <u>In the forests of Brazil</u>, their assembled forces constitute more than 20 percent of the weight of all land animals, including nematode worms, toucans, and jaguars.

<div align="right">Wilson, p. 12–13</div>

Besides these, there was a succession of locative adverbials whose fronting was contrastive, cf. (3).

(3) "as a spectrum of possibilities, all of which properly blend into one another … <u>At one end</u>, we have *the hard, bright lights of science*; <u>here</u> we find *information*. <u>In the center</u> we have *the sensuous hues of art*; <u>here</u> we find *the aesthetic shape of the world*. <u>At the far end</u>, we have *the dark, shadowy tones of religious experience*, shading off into wave lengths beyond all perception; <u>here</u> we find *meaning*." No, <u>here</u> we find *obscurantism*!

<div align="right">Wilson, p. 10</div>

Example (3) displays, in addition to three instances of contextual fronting (*At one end, In the center, At the far end*), four occurrences of the context-dependent anaphoric proform *here*, referring to these three different locations. The locations are partly derivable from the situational context (a spectrum is a continuum with a distinguishable centre and two ends), but in connection with the immediately preceding linguistic context they introduce novel features. The proform *here*, which is prototypically context-dependent and operates within the thematic section, is here disengaged from context-dependence by the decontextualizing factors of contrast, identification of the particular location and purposeful repetition (cf. Firbas, 1995; Stehlíková, 2016). As a result, (3) displays an FSP

structure with two foci, a contrastive diatheme, constituted by *here* at the beginning, and the rheme in the post-verbal, final position.

Moreover, unlike temporal fronted adverbials, locative adverbials appeared in clauses with subject-verb inversion (3 instances) constituting a series. They are of particular interest in that they show resemblance to (3); cf. (4).

(4) [but the distribution of its intensity will have been drastically altered.] <u>At one end of this distribution will be the minority of the people</u> whose work will keep intact the technology that sustains the multitude at a high standard of living. <u>In the middle of the distribution will be found a type, largely unemployed,</u> for whom the distinction between the real and the illusory will still be meaningful … He will retain interest in the world and seek satisfaction from sensual pleasures. <u>At the other end of the spectrum will be a type largely unemployable,</u> for whom the boundary of the real and the imagined will have been largely dissolved, at least to the extent compatible with his physical survival.

<div align="right">Wilson, p. 4</div>

In (4), the fronted adverbials are partly context-dependent owing to their anaphoric postmodifier; however, they appear in a series of differing locations, just as in (3). They are thus disengaged from context dependence by the same decontextualizing factors of contrast and identification, and constitute a contrastive diatheme. As compared with (3), fronting here endows the adverbials with lesser prominence, its apparent cause being the use of inversion. Cf. the description of the function of S–V inversion after initial adverbials as a contribution to the maintenance of a consistent perspective in Biber et al. (1999) "There is a preferred distribution of this information [given – new] in the clause, corresponding to a gradual rise in information load." (pp. 896, 899) With direct word order, the fronted adverbial would be more prominent as a result of a potential pause after it, cf. *At one end of this distribution*(,) *there will be the minority of the people whose work* ... Another factor of the lesser prominence of the fronted adverbials in (4) is the complexity and weightiness of the sentence structure, as compared with the more succinct structure of (3).

The other instances with inversion (which here serves the same function as in (4)) represented contextual fronting of a context-dependent, thematic element with the linking function, cf. (5) a. and b.

(5) a. Educated people everywhere like to believe that <u>beyond material needs lie fulfillment and the realization</u> of individual potential.

<div align="right">Wilson, p. 3</div>

b. <u>Thus does ideology bow to its hidden masters the genes</u>, and the highest impulses seem upon closer examination to be metamorphosed into biological activity.

<div align="right">Wilson, p. 4</div>

The other semantic roles involve cause/source, measure/degree, means/instrument, manner/comparison, accompanying circumstances, addition/exception, accordance.

Twelve of the 16 adverbials with semantic roles other than temporal or locative (including 2 instances with inversion and two of contrastive contextual fronting) represented contextual fronting with connective function, cf. (6) a. and b.

(6) a. [The situation can be summarized by saying that biology stands today as the antidiscipline of the social sciences.] <u>By the word "antidiscipline"</u> I wish to emphasize the special adversary relation that often exists when fields of study at adjacent levels of organization first begin to interact.

Wilson, p. 7

b. Sterile castes engaged in rearing siblings are the essential feature of social organization in the insects. <u>Because of its link to haplodiploidy,</u> insect social life is almost limited to the wasps and their close relatives among the bees and ants.

Wilson, p. 12

Contrastive fronting of context-dependent adverbials was found in two instances with the semantic role of accompanying circumstances, cf. (7):

(7) It is the essential first hypothesis for any serious consideration of the human condition. <u>Without it</u> the humanities and social sciences are the limited descriptors of surface phenomena, like astronomy without physics, biology without chemistry, and mathematics without algebra. <u>With it</u>, human nature can be laid open as an object of fully empirical research, biology can be put to the service of liberal education, and our self-conception can be enormously and truthfully enriched.

Wilson, p. 1–2

Although this example displays dual focus like (3), owing to the weightiness of the predication, the degree of communicative dynamism of the verb complementation is much higher.

Proceeding to the types of fronting displayed by the non-adverbial elements, we find that both, an object and a subject complement, represent emphatic fronting, but of a different kind, cf. (8) and (9).

(8) [The situation can be summarized by saying that biology stands today as the antidiscipline of the social sciences. By the word "antidiscipline" I wish to emphasize the special adversary relation that often exists when fields of study at adjacent levels of organization first begin to interact.] <u>For chemistry</u> there is *the antidiscipline of many-body physics*; <u>for molecular biology,</u> *chemistry*; <u>for physiology,</u> *molecular biology*; and so on upward through the paired levels of increasing specification and complexity.

Wilson, p. 7

(9) [Robert Nozick begins *Anarchy, State, and Utopia* (1974) with an equally firm proposition: "Individuals have rights, and there are things no person or group may do to them (without violating their rights).] <u>So strong</u> and <u>far-reaching are</u>

<u>these rights they raise the question</u> of what, if anything, the state and its officials may do."

<div align="right">Wilson, p. 5</div>

The *for* PrepP in (8) marks a participant in the existential predication, hence it is assigned the syntactic status of object, not adverbial. It is partly context-dependent, being related to the preceding statement about biology as an antidiscipline of social sciences. However, it is disengaged from this broader context dependence by the decontextualizing factors of contrast and selection: biology x chemistry. Moreover, it is contrasted with the primary participant in the existential predication, the notional subject 'the antidiscipline many-body physics'. This contrast is further developed in the following sentences by means of parallel structures: a discipline and an antidicipline, and the relation between them, are expressed in the same way, i.e., respectively, as fronted object and existential subject, albeit elliptically (hence counted only as one example). We thus have a communicative clausal field with two foci, the first focus falling on a contrastive diatheme at the beginning and the second on the rheme at the end.

The clausal communicative field of (9) is simpler. It contains only one focus on the fronted element while the final element, the subject, is context-dependent, with no decontextualizing factor within the scope of the retrievability span. The fronted subject complement is context-independent and in the underlying ordering constitutes the rheme in the regular post-verbal position, cf. (9)':

(9)' These rights are so strong and far-reaching ...

When fronted, the rheme acquires additional emphasis. Inversion is here obligatory, fronted intensified and negated elements triggering inversion as a concomitant feature.

The findings obtained from this sample of academic writing show a great prevalence of contextual fronting, as was supposed, but exceed the initial assumptions in displaying other types fronting, contextual contrastive and emphatic, the former recurrently in series of parallel structures. Also notable is the relatively frequent occurrence of S–V inversion.

3.1.2 Hergehahn and Henley

Apart from the different representation of conjuncts and disjuncts, the psychological sample presents a fairly similar picture as the sociobiological text (3.1.1). Integrated fronted adverbials again considerably outnumber the other fronted clause elements, viz. 13 as compared with 7. Their semantic roles are varied, temporal and locative adverbials accounting for, respectively, 3 and 1 instances. Other semantic roles included accompanying circumstances, manner, respect and result, the first two recurrently, cf. (10) a., b., c., and d.

(10) a. <u>During the month-long delay</u>, Socrates was imprisoned but met regularly with his friends.

<div align="right">Hergenhahn and Henley, p. 42</div>

b. <u>In Chapter 20,</u> we will see that the extreme relativism of the Sophists has much in common with the contemporary movement called postmodernism.

<div align="right">Hergenhahn and Henley, p. 40</div>

c. <u>In his search for truth,</u> Socrates used a method sometimes called inductive definition, which started with an examination of instances of such concepts as beauty, love, justice, or truth ...

<div align="right">Hergenhahn and Henley, p. 41</div>

d. <u>In this way,</u> he sought to discover general concepts by examining specific examples.

<div align="right">Hergenhahn and Henley, p. 41</div>

With the exception of one instance, all fronted adverbials represented contextual fronting, cf. the examples listed under (10).

The only instance other than contextual fronting is the fronting of a context-independent manner adjunct, hence an example of emphatic fronting with dual focus, with the second, main focus on the verb complementation, cf. (11):

(11) Nor is it proper for him to move from one place to another. But <u>effortlessly</u> he shakes all things *by thinking with his own mind*.

<div align="right">Hergenhahn and Henley, p. 41</div>

As regards the fronted objects, they have the realization forms of prepositional *for-* and *on*-phrases, cf. (12) and (13):

(12) It is important to note that although Socrates sought the essence of various concepts, he did not believe that essences had abstract existence. <u>For him,</u> an essence was a universally acceptable definition of a concept – a definition that was both accurate and acceptable to all interested parties.

<div align="right">Hergenhahn and Henley, p. 42</div>

(13) Socrates was charged with corrupting the youth of Athens because he caused them to question all things, including many cherished traditional beliefs. <u>Perhaps on the latter charge</u> he *was guilty*.

<div align="right">Hergenhahn and Henley, p. 42</div>

Example (12) is an instance of non-contrastive contextual fronting, *for him* referring to the same referent in the preceding sentence that lacks any other potentially contrastive elements. On the other hand (13) represents contrastive contextual fronting with dual focus: the fronted object *on the latter charge* is disengaged from context dependence by the factor of selection, hence performs the FSP function of contrastive diatheme carrying the first focus while the second, main focus is on the rheme constituted by the positive polarity of the copula and its context-independent complement *was guilty* (Socrates' having been charged does not entail that he was guilty). The other two instances of object fronting have an FSP structure similar to (12).

The other three instances of fronted clause elements other than objects display a fronted modifier (cf. *he complete sees*) in parallel structure, cf. (14). The example illustrates emphatic fronting with dual focus, the first focus being on the fronted element, context-independent in the first use, and disengaged from context dependence in the following uses

by the factor of purposeful repetition and by being related to new predicates; the second, main focus is on the rhematic predicates, constituted by context-independent verbs.

(14)   <u>Complete</u> he *sees*, <u>complete</u> he *thinks*, <u>complete</u> he *hears*.

<div align="right">Hergenhahn and Henley, p. 41</div>

3.1.3 Academic samples compared

As appears from Table 2, the differences between the two sources of the academic sample mostly concern adverbials. The psychology text contains a considerably larger amount of conjuncts and disjuncts (respectively, 25 and 5) than the sociobiological text (respectively, 9 and 1). This is reflected in the representation of integrated and non-integrated fronted elements in the two texts: while in the psychology text it is the non-integrated fronted elements that predominate (30 as compared with 20), in the sociobiological text it is the integrated elements (40 as compared with 10). The differences in the distribution of both integrated and non-integrated fronted elements is connected with the different character of the two texts: the excerpted part of Wilson's sociobiological text (pp. 1–15) deals with the role and development of different species, which involves their habitats and the relevant periods, whereas the excerpted part of Hergenhahn and Henley's text is a subchapter of an introduction to the history of psychology, concerned with the relativity of truth (pp. 39–42). Moreover, the psychological text is expository instructional, whereas the book on sociobiology is more expository argumentative. Yet another, not negligible factor of the differentiation appears to be the auctorial style, Wilson's style being livelier and more varied.

Despite the much lower absolute representation of integrated fronted adverbials in the psychological text (13 as compared with 38 in the sociobiological text), they still considerably preponderate over the other integrated fronted elements (13 as compared with 7, i.e. 65% and 35%, respectively; in the Wilson text, the respective figures are 38 [95%] and 2 (5%]). The low representation of integrated fronted adverbials in the psychology text appears to be primarily due to the scarce incidence of temporal and locative adjuncts (3 and 1, respectively; as compared with 13 and 9 in the Wilson text). As mentioned above, these differences derive from the different subject matter of the two texts.

What appears to be characteristic of adverbials in the whole academic sample is the type of fronting. Prototypically, adverbials of all semantic roles (in particular temporal and locative) represent contextual fronting serving as a link between what precedes and what follows: among the 51 adverbials, this type was found in 44 instances (86.3%). Six of the other seven instances represent contrastive contextual fronting while one is a case of emphatic fronting. On the other hand, unlike the adverbials the non-adverbial integrated elements represent four different types of fronting, contextual, contextual contrastive, emphatic and emphatic fronting with two foci. Though suggestive, the figures are too small to allow generalization.

<div align="center">

3.2 Narrative sample

</div>

The distribution of the syntactic functions of fronted elements in the narrative sample is shown in Table 3.

Table 3. Fronted elements in fiction narrative

	Lodge	James	**Total**
Conjuncts	2	1	3
Disjuncts	6	3	9
Subtotal	**8**	**4**	**12**
Adverbials	38	46	84
Object	2	–	2
Other clause elements	2	–	2
Subtotal	**42**	**46**	**88**
Total	**50**	**50**	**100**

As shown in Table 3, the two sources of fiction narrative basically display a similar picture: in both, integrated fronted elements greatly preponderate over non-integrated fronted elements, and within the integrated elements, it is the adverbials that largely predominate, similarly as in the academic sample.

3.2.1 Lodge

The Lodge sample of narrative text contains 42 clauses with an integrated fronted element, among which over 90% (38 instances) are accounted for by adverbials. The most frequent semantic role is the temporal (16 occurrences, 42% of all fronted adverbials). With the exception of one instance, all temporal adjuncts represent contextual fronting with the linking function, cf. (15) a. and b.

(15) a. This new behaviour implied that they all knew perfectly well who he was, thus making any attempt at self-introduction on his part superfluous, while at the same time it offered no purchase for extending acquaintance.

Lodge, p. 39

  b. Simply keeping warm was Morris Zapp's main preoccupation in his first few days at Rummidge. On his first morning, in the tomb-like hotel room he had checked into after driving straight from London airport, he had woken to find steam coming out of his mouth.

Lodge, p. 31

The single temporal adverbial displaying a type of fronting other than the prevalent contextual type is illustrated in (16).

(16) [When he had moved his baggage into the O'Shea house, he filled the micro-refrigerator with TV dinners, locked his door, turned up all the fires and spent a couple of days thawing out.] Only then did he feel ready to investigate the Rummidge campus and introduce himself to the English Department.

Lodge, p. 31

This example represents emphatic fronting of a context-dependent element *then*, reinforced by the focalizer *only*; apparently, *then* alone is too weak an expression to be able to carry the main and only focus of the clausal communicative field. The predicative part is context-dependent through the broader preceding context. Inversion is here grammatical, being an obligatory concomitant feature of the fronting of an intensified element.

The next best represented semantic role was displayed by locative adverbials (12 instances). All fronted locative adverbials represent contextual fronting with the linking function, cf. (17) a.; a fronted locative adverbials also appears in (15) b. as a second fronted element in addition to the temporal. In two instances the fronting of locative adverbials is accompanied by subject-verb inversion, cf. (17) b.

(17) a. The faculty began to trickle back to their posts. From behind his desk he heard them passing in the corridor, greeting each other, laughing and opening and shutting their doors.

Lodge, p. 39

b. In the top right-hand one was an envelope addressed to himself.

Lodge, p. 34

As noted above, Biber et al. (1999: 898) describe inversion as a means serving the maintenance of a consistent perspective, viz. gradual rise in information load (ibid.: 896).

Of the other semantic roles (altogether 10 instances) the most frequent role was accompanying circumstances, illustrated in (18).

(18) a. In his lonely isolation, Morris turned instinctively for solace to the media.

Lodge, p. 39

b. With Wily Smith's assistance he telephoned the Chairman of the Department.

Lodge, p. 36

One instance of contextual fronting was also found among the other fronted clause elements. It is illustrated in (19). Here the fronted element is the postmodifier of the subject complement (cf. the reference of the proform *the one: the question of the two questions he was asked at the cocktail party by everyone he met*).

(19) 'How are you liking Euphoria?' Of the two questions he was asked at the cocktail party by everyone he met, this was the one he preferred.

Lodge, p. 35

The two fronted objects and the fourth fronted non-adverbial clause element, a subject complement, represent different types of fronting.

The fronted objects represent contrastive fronting of context-dependent elements disengaged from context dependence by the factor of contrast, cf. (20). The two sentences display dual focus, the second, main focus being on the predicative parts.

74

(20) [There were, of course, no commercials on the programme, but just to annoy the rival networks Boon would sometimes give an unsolicited and unpaid testimonial to some local restaurant or movie or shirt-sale that had taken his fancy.] <u>To Philip</u> it seemed obvious that beneath all the culture and the eccentricity and the human concern there beat a heart of pure show-business, but <u>to the local community</u> the programme evidently appeared irresistibly novel, daring and authentic.

<div align="right">Lodge, p. 42</div>

The fronted subject complement is emphatic fronting of the rhematic element, i.e. fronting with one focus on the fronted element, cf. (21) (the explicit expression of the subject in right dislocation constituting a separate communicative field).

(21) <u>And an extraordinary variety of listeners they were</u>, those who faithfully tuned into QXYZ at midnight ...

<div align="right">Lodge, p. 41</div>

3.2.2 James

The James sample of narrative text presents a sparsely diversified picture insofar as all integrated fronted elements (46) are adverbials out of which 37 instances (74%) are temporal. A recurrent realization form is the adverb *then*, which was found in 9 instances, all expressing the basic temporal meaning 'afterwards' and signalling successive steps in the development of a particular narrative episode, cf. (22):

(22) She stared ahead, frowned, <u>then</u> shook her head and let in the clutch. Her companion hesitated, looked at her, <u>then</u> leaned back and released the rear door.

<div align="right">James, p. 2</div>

Another recurrent adverb was *now* (6 occurrences), introducing an action happening at the particular moment. In one case it is contrasted with *already*, the two communicative fields constituting an instance of contrastive fronting with dual focus, cf. (23)

(23) It would be the end of her Friday evenings with Wayne and Shirl and the gang. <u>Already</u> they teased and pitied her because she was treated as a child. <u>Now</u> it would be total humiliation.

<div align="right">James, p. 1</div>

In a tense passage, displaying a parallel adverbial structure, the last unit contains *then* contrasting with *now*, which heightens the dramatic effect of the passage, cf. (24):

(24) She felt herself falling through time, through space, through an eternity of horror. <u>And now</u> the face was hot over hers and she could smell drink and sweat and a terror matching her own. Her arms jerked upwards, impotently flailing. <u>And now</u> her brain was bursting and the pain in her chest, growing like a great red flower, exploded in a silent, wordless scream of 'Mummy! Mummy!' <u>And</u>

<u>then</u> there was no more terror, no more pain, only the merciful, obliterating dark.

<div align="right">James, p. 4</div>

Locative adverbials had three occurrences, two of them with inversion which was also found with one temporal adverbial, cf. (25) a. and b. Adverbials with other semantic roles (five instances) are illustrated in (26).

(25) a. <u>On the uncluttered desktop was</u> a small glass vase of freesias.

<div align="right">James, p. 5</div>

    b. <u>And then came</u> salvation.

<div align="right">James, p. 2</div>

(26) She was glad when, <u>still without a word,</u> the driver drew up at the crossroads and she was able to scramble out with muttered thanks.

<div align="right">James, p. 2</div>

There were two instances of contrastive contextual fronting one of which was adduced above, cf. (23). The other instance is listed under (27) as (27) a., owing to the same realization form of the adverbial as in (27) b. and c.: all of them are de-adjectival adverbials formed by the suffix *-ly*. Their FSP functions, however, along with the respective clausal communicative field differ.

(27) a. She didn't question <u>how, so mysteriously,</u> this slim, slow-walking figure had materialized.

<div align="right">James, p. 3</div>

    b. And then, <u>miraculously,</u> her prayer was answered.

<div align="right">James, p. 3</div>

    c. After four years of silence, his new book of poetry, *A Case to Answer and Other Poems*, had been published to considerable critical acclaim which was <u>surprisingly</u> gratifying, and to even wider public interest <u>which, less surprisingly,</u> he was finding more difficult to take.

<div align="right">James, p. 4</div>

Example (27) a. is an instance of emphatic fronting of a context-independent element realized by an intensified manner adjunct, representing the same type of fronting as the manner adjunct in (11) (*effortlessly*) and the subject complement in (9) (*So strong*). In the underlying regular ordering it is a component of the rheme; when fronted it acquires additional emphasis. The communicative field of (27) a. has only one information peak (one focus), the rest of the sentence being context dependent. In (27) b., the adverbial *miraculously*, whether interpreted as a manner adjunct or a disjunct with a scope over the whole sentence, is again a context-independent element, but so is the verb and its

polarity, hence the communicative field of (27) b. has two foci, the first on the fronted adverbial and the second on the verb. Example (27) c. displays the only non-integrated fronted element, a disjunct, whose fronting may be regarded as contrastive owing to the comparative component expressing a lesser degree than in the adverbial's first occurrence, through which it is disengaged from context dependence; hence the communicative field displays a contrastive diatheme and the rheme in the verb complementation. Moreover, not only does *less surprisingly* stand in contrast to *surprisingly* in the preceding sentence, but also the predicative part *was finding more difficult to take* contrasts with *was gratifying*.

On the whole the fronted elements in the James sample appear to reflect the character of narrative text: the prevailing temporal adverbials mark successive stages in the development of series of actions depicting an episode, and thus perform an important function in the build-up of the narrative. Discernibly, however, their distribution and particular uses are also ascribable to the auctorial style. Whether the absence of fronted clause elements other than adverbials may be regarded as a feature of narrative text needs to be ascertained by a more extensive search.

3.2.3 Narrative samples compared

Overall, the two sources of fiction narrative present a similar picture not only in the distribution of integrated and non-integrated elements – in both texts integrated fronted elements vastly preponderate (accounting for 92% in James and 84% in Lodge), but also in the representation and the types of fronting of fronted adverbials. In both samples, fronted adverbials overwhelmingly outnumber other clause elements (in James to the exclusion of any other fronted integrated clause elements; in Lodge the proportion of fronted adverbials to other fronted clause elements is 38 : 4). The two samples also correspond in the types of fronting: both predominantly display the contextual type with the linking function. Some of the specific features of the two sources are to be ascribed to the auctorial style.

3.3 Fiction dialogue

The distribution of the syntactic functions of fronted elements in fiction dialogue is shown in Table 4.

Table 4. Fiction dialogue

	Lodge	James	**Total**
Conjuncts	28	14	42
Disjuncts	5	19	24
Interjections	5	1	6
Subtotal	**38**	**34**	**72**
Integrated elements			
adverbials	6	15	21
other clause elements	6	1	7
Subtotal	**12**	**16**	**28**
Total	**50**	**50**	**100**

3.3.1 Lodge

As shown in Table 4, the Lodge sample of fiction dialogue displays only 12 integrated clause elements, all other non-subject openings of clauses being accounted for by non-integrated elements. Although conjuncts may be regarded as an intermediate category between integrated and non-integrated elements – especially in the case of the semantic role of time, this does not apply to the most frequent conjunct in this sample, viz. *well*, which has 23 occurrences, a notable feature of the utterances of the British protagonist of the novel and other British characters. The connective function of *well* can hardly be described as primary among the many diverse functions of this polyfunctional device largely serving as a discourse marker rather than a conjunct. Compare the two uses of *well* in (28). In (28) a. *well* serves as a hesitant response gambit with an attitudinal undertone. On the other hand, in (28) b. (continued turn of the same speaker) it performs a connective function.

(28) a. 'Mrs Swallow, may I ask you a personal question about your husband?'
       She looked at him in alarm. <u>'Well, I don't know.</u> It depends…'

       Lodge, p. 50

     b. 'The schools are pretty good,' he said. '<u>Well, one or two</u> –'

       Lodge, p. 47

The connective function of the other conjuncts was more pronounced, cf. (30) a. and b., but these were few: *so* (2 occurrences), *also*, (inferential) *then* and *after all*.

(29) a. ['Your wife with you?'
       'No.' She responded with a gesture which implied clearly enough that his assumption was therefore demonstrably unwarranted.] 'I would have liked to have brought her,' he said. 'But my visit was arranged at rather short notice. <u>Also we have children</u>, and there were problems about schooling and so on.'

       Lodge, p. 45

     b. [Also we have children, and there were problems about schooling and so on. And there was the house…' … 'Do you have children yourself?' he concluded desperately. 'Two. Twins. Boy and girl. Aged nine.'] 'Ah, <u>then</u> you understand the problems.'

       Lodge, p. 45

As for the disjuncts, also numerous in this sample (24 instances), nearly all were of the epistemic type expressing a degree of factivity (*sure, as a matter of fact, really*), with the exception of *on second thoughts*, which presupposes an antecedent.

A specific feature of this sample was the occurrence of interjections, both primary and secondary:

(30) a. 'You don't mean my copy of *Playboy,* by any chance? But that's ridiculous, *Playboy* isn't pornography, for heaven's sake! <u>Why, clergymen read it</u>. Clergymen write for it!'

       Lodge, p. 64

b. 'Exactly!' cried Mrs Zapp. 'I do, I do!' She looked at her watch suddenly, and said, '<u>My God, I must go</u>,'

<div align="right">Lodge, p. 47</div>

The integrated fronted elements displayed the following patterns of fronting. One locative (*here*) and one temporal adverbial (*now*) had deictic reference, cf. (31) a. and b. Two other temporal adverbials and the remaining one (accompanying circumstances) were instances of contextual fronting, cf., respectively, (31) c. and d.

(31) a. <u>here's</u> th' old josser himself.

<div align="right">Lodge, p. 40</div>

b. <u>Ah, now</u>, that's better, that's better ...

<div align="right">Lodge, p. 41</div>

c. [Boon was still talking away in the background, about his plans for a TV arts programme:] 'Something entirely different... art in action... train a camera on a sculptor at work for a month or two, <u>then run the film through</u> at about fifty thousand frames per second, see the sculpture taking shape... put an object in front of two painters, ...

<div align="right">Lodge, p. 44</div>

d. '<u>In that case,</u> you take my bed, and I'll sleep in here.'

<div align="right">Lodge, p. 62</div>

It is to be noted that (31) a., b. and c. are monologues, (31) a. and b. a spectator's running commentary on a TV show, (31) c. a description of a future arrangement.

Two other instances of contextual fronting with the linking function, albeit as anacolutha, were also found among the other fronted integrated elements. In (32) a. the initial context-dependent element, which takes up what was said before, introduces the theme of the following utterance, in (32) b. the conversational topic is suggested by an object in the current situation.

(32) a. '<u>And the one I went to,</u>' said the Confederate Soldier, 'we didn't know who was the group leader,

<div align="right">Lodge, p. 58</div>

b. 'Tell me, <u>that badge you're wearing</u> – what is Kroop?'

<div align="right">Lodge, p. 37</div>

While with the exception of (31) c. and d. none of these types have been encountered either in academic writing or fiction narrative, the four remaining instances of fronted integrated elements appear to be non-specific to a text-sort. Presumably significantly, these instances are again constituted by fronted objects and fronted subject complements. Example (33) illustrates contextual contrastive fronting with two foci, the first on the

fronted object with the FSP function of contrastive diatheme, the second, main focus on the predicative part.

(33) ['What bargain?' said the woman sharply. 'Just a figure of speech.] I mean, <u>for me</u>, it's a great opportunity, a paid holiday if you like. But <u>for her </u>it's just life as usual, only lonelier.

<div align="right">Lodge, p. 45</div>

One of the remaining two fronted elements is again an object, the other is a subject complement; both constitute emphatic fronting of the rheme which in the underlying regular ordering occurs in post-verbal position. The clausal communicative fields have only one focus on the emphatically fronted initial element, the rest being context-dependent.

(34) a. '<u>Terrific cheek </u>he had, that chap Boon. I wonder what became of him.'

<div align="right">Lodge, p. 53</div>

b. [somebody just told me there's an English guy at this party who asked Hogan to introduce him to Karl Kroop. I'd love to have seen the old man's face.' 'Ask him,' said Ringbaum, nodding towards Philip. Philip blushed and laughed uneasily. 'Oh my God, you aren't the English guy by any chance?' 'You goofed again, Sy, dear,' said the woman.]
'I'm terribly sorry,' said the man. '<u>Sy Gootblatt</u> is the name.

<div align="right">Lodge, p. 43</div>

Apart from the type of fronting, (34) a. shows another specific feature of conversation, right dislocation of the subject (cf. also (21) in 3.2.1).

3.3.2 James

The James sample of fiction dialogue, as shown in Table 4, also contains relatively few integrated clause elements, viz. 16 instances, all of them adverbials except one. As regards the non-integrated elements, conjuncts and disjuncts have a fairly comparable distribution (14 and 19, respectively), interjections being marginal (1 instance). The conjuncts included *well* (5 occurrences, i.e. nearly a third of the conjuncts in this sample: 5 out of 14); equally represented was *then*, the other conjuncts being *anyway* (with 3 occurrences) and *so*. Like Lodge's group of disjuncts, the disjuncts in the James sample were mostly epistemic content disjuncts (*perhaps*, *of course*, *presumably*, all recurrent), with the exception of two occurrences of the style disjunct *honestly*.

The integrated fronted elements in this sample include 15 adverbials and one subject. Most of the adverbials are instances of contextual fronting serving as diathemes in the thematic section of the clause. By being fronted they serve as a device contributing to the basic distribution of CD. This is especially the case of temporal adjuncts, which account nearly for a half of all fronted adverbials in this sample. Compare (35) a. and b.

(35) a. And <u>now</u> he's got the latest Comare report to fuel his spurious concern.

<div align="right">James, p. 38</div>

b. 'Two weeks. Two weeks too bloody many. <u>Then</u> I ran away and found a squat.'

<div align="right">James, p. 25</div>

Other semantic roles, which were very diverse (manner, cause, accompanying circumstances, place and others) are illustrated in (36) a. and b.

(36) a. '<u>In the interests of the station</u> I think you should drop it.'

<div align="right">James, p. 36</div>

   b. <u>All over the world</u> ordinary people are marching, demonstrating, making their voices heard,

<div align="right">James, p. 28</div>

The contribution of fronted adverbials to the ordering of clause elements according to a gradual rise in communicative dynamism is reinforced where they co-occur with subject-verb inversion, as in (37).

(37) <u>And out there somewhere is a mass murderer</u> who enjoys strangling women and stuffing their mouths with their hair.

<div align="right">James, p. 43</div>

Example (38) shows another locative adverbial followed by subject–verb inversion.

(38) <u>and out pops an Identikit of sonny</u> complete with prints, collar-size and taste in pop music.

<div align="right">James, p. 45</div>

Unlike (37), which is an instance of contextual fronting of a context-dependent thematic element, here the adverbial is context-independent (in the underlying regular ordering it is the rheme in its usual post-verbal position) and the fronting is emphatic. By being fronted the rhematic adverbial acquires a higher degree of CD and becomes more emphatic. Again unlike (37), which is a presentation sentence offering optional insertion of existential *there*, here the inversion occurs in a quality scale with a context-independent carrier of quality and final specification, cf. the underlying structure without inversion '*and an Identikit of sonny complete with prints, collar-size and taste in pop music pops out.*' The communicative field contains two foci, a minor one on the initial adverbial[2] and the main on the final subject.

The last example of a fronted element is the only instance of fronted subject. In English, subject fronting appears to be a contradiction since the subject is the canonical initial clause element, hence movement to the left is ruled out. Consequently, a special construction is required; in (39), fronting of the subject is achieved by left dislocation. The dislocated subject has a weighty structure due to its apposition construed with a comment clause.

---

[2] The possibility to front the adverbial element of the verb phrase is adduced as one of the criteria for distinguishing a free combination of intransitive verbs + adverbs from intransitive phrasal verbs (cf. Quirk et al. 1985: 1153 (16.3).

(39) <u>This proposed job</u> – rumoured job, I should say – I suppose we're entitled to ask whether <u>it</u> has been formally offered to you yet? Or aren't we?

<div align="right">James, p. 33</div>

Example (39) is an instance of contextual fronting. The fronted subject introduces a known discourse topic that operates both as the theme in the sense of 'what is going to be talked about' (i.e. the textual theme) and the theme in FSP terms, here the diatheme, with the theme proper realized by its pronominal postcedent in the regular position.

Comparing this example with (34) and (21), which also display dislocated subjects, two points are to be noted: first, fronting involves only movement to the left and leaves an unfilled gap in the regular ordering; dislocation involves movement of a clause element either to the left or to the right (as shown in (21) and (34)), with concomitant employment of a pro-form in the regular position (cf. postponed and anticipated identification in Quirk et al. 1985: 1310) Secondly, and more importantly, unlike (39), the fronted and the dislocated elements in (21) and (34) are not identical, the fronted elements being, respectively, the subject complement and the object, while the dislocated elements are the subjects. Consequently, there is no relation between the two processes – the sentences undergo both fronting and right dislocation independently of each another. In (39) the fronted and the dislocated element are identical, viz. the subject, and left dislocation here serves as a device of fronting. What is common to all three instances is the stylistic effect produced by dislocation: this structure is a feature of informal speech, which is here imitated by the fiction dialogues.

3.3.3 Dialogue samples compared

The fiction dialogue is the only of the three examined text sorts in which the non-integrated elements outnumber the integrated ones. In this respect, the two sources of fiction dialogue appear to be basically similar, cf. 38 non-integrated and 12 integrated fronted elements in the Lodge sample, with 34 and 16, respectively, in the James sample. However, within these groups there are considerable differences which at least in the more numerous group of non-integrated elements may be considered significant indicators of the character and content of the two novels, as well as of the auctorial style.

The two major groups of non-integrated elements, conjuncts and disjuncts, have a patently different distribution. While in the Lodge sample it is conjuncts that greatly predominate (28 as compared with 5 disjuncts), the James sample displays a noticeable preponderance of disjuncts (19 as compared with 14). Notably, however, the cause of the relatively high representation of conjuncts in the Lodge sample is the recurrence of *well* (23 out of 28), which here serves as a characteristic of British speech and attitudes. A similar ground may be sought in the James sample for the much higher representation of disjuncts (19 as compared with 5 in Lodge) insofar as a half of the epistemic content disjuncts were expressions lowering the truth value of the statement being made (cf. six occurrences of *perhaps*, two of *presumably*).

Unequal distribution was also found with the third non-integrated element registered in both samples, interjections, which is evidently connected with the different character of the two novels, Lodge's humoristic subject matter providing more situations for their occurrence than James's grim story (5 and 1 respectively).

The representation of the integrated fronted elements is too small to offer ground for ascertaining source-specific features, even though the equal number of adverbials and other clause elements (6 instances each) in the Lodge sample suggests more diverse language, again resulting from the character of the story.

4. Discussion of the results

4. 1 Distribution of fronted elements in the three text sorts

The distribution of fronted elements in all three text sorts is shown in Table 5.

As appears from the figures, the fronted elements in the three text sorts show a clear-cut distinction between academic writing and narrative text on the one hand, and fiction dialogue on the other, with respect to the distribution of the integrated and non-integrated clause elements. While in both academic writing and narrative text it is the integrated clause elements that predominate (60 instances of integrated elements as compared with 40 instances of non-integrated in academic writing; in the narrative part, the predominance of integrated elements is even more prominent: 88 instances as compared with 12), in fiction dialogue the number of integrated elements forms a minor part: 28 instances, the remaining 72 initial elements being accounted for by conjuncts, disjuncts and interjections. These figures call for further research into fiction dialogue, the number of integrated fronted elements being too small to allow drawing plausible conclusions.

Differentiation is moreover found between the two sources of each sample, which is largely to be ascribed to their different content and sub-register, and partly also to the auctorial style. On the other hand, there are similarities between the academic and the narrative sample as regards the representation of different fronted integrated clause elements. In both samples, they include a large majority of adverbials which account for 85% (51 out of 60) in the academic sample and for 95.4% (84 out of 88) in the narrative sample. This partly reflects the optional character of adverbials, and their different syntactic status and diversified semantics which enable them, in contrast to the constitutive clause elements, to occur side by side in the same sentence (cf. (11), (15) b., (26), (27) b).

However, there are notable differences in the distribution of semantic roles. Most adjuncts in the narrative text are temporal (63.1%), while in the academic sample they account for only 27.4%. The largest group of adverbials in the academic sample has semantic roles other than temporal and locative: adjuncts of cause, means, manner, accompanying circumstances and others together comprise 27 instances of the total of 51 adverbials, i.e. 53%). The different distribution of adverbial semantic roles in the two samples reflects the respective character of the text. The greater variety of semantic roles in the academic sample derives from its expository and argumentative nature, while the prevailing temporal adverbials in the narrative mark successive stages in the development of series of actions depicting the current episodes. Discernibly, however, the distribution and particular uses of adverbials are also due to the auctorial style.

Fiction dialogue appears to differ from the other two samples in all the examined points, apparent similarities being either source-specific, rather than text-sort specific, or superficial. Thus the predominance of adjuncts over other fronted integrated clause

Table 5. Distribution of fronted elements

	Wilson sociobiology	Hergenhahn and Henley psychology	Acad. sample	Narrative Lodge	Narrative James	Narrat. sample	Fiction dialogue Lodge	Fiction dialogue James	Fiction dialogue sample
Conjuncts	9	25	**34**	2	1	**3**	28	14	**42**
Disjuncts	1	5	**6**	6	3	**9**	5	19	**24**
Interjections	–	–	–	–	–	–	5	1	**6**
Non-Integrated elements	**10**	**30**	**40**	**8**	**4**	**12**	**38**	**34**	**72**
Temporal adjuncts	13	2	14	16	37	53	3	6	9
Locative Adjuncts	9	1	10	12	3	15	2	3	5
Other Adjuncts	16	10	27	10	6	16	1	6	7
Subtotal	**38**	**13**	**51**	**38**	**46**	**84**	**6**	**15**	**21**
Other clause elements	2	7	**9**	4	–	**4**	6	1	**7**
Integrated elements	**40**	**20**	**60**	**42**	**46**	**88**	**12**	**16**	**28**
Total	**50**	**50**	**100**	**50**	**50**	**100**	**50**	**50**	**100**
Instances with S–V inversion	6	–	**6**	3	3	**6**	2	2	**4**

elements, comparable with that in the narrative and academic samples, was found only in one fiction-dialogue source (15 and 1, respectively, in James), whereas the other source shows an equal distribution (6 and 6, respectively, in Lodge). Similarly, the relative correspondence of the representation of conjuncts in fiction dialogue and the academic sample is due only to one source in each (psychology and Lodge). Moreover, there are essential qualitative differences between them: while in the psychology sample 18 of the 25 conjuncts were different lexical items, the only recurrent conjunctive adverbials being *therefore* with six occurrences, and *furthermore*, *in other words* and sets of listing conjuncts with two occurrences each, the 28 conjuncts in Lodge are with five exceptions accounted for by *well*.

Table 5 also includes, in addition to the fronted clause elements, instances of subject-verb inversion.[3] While the distribution of the different clause elements in the three samples displays significant differences, inversion apparently tends to equal or near-equal distribution: cf. the same number of instances – six – in academic writing and narrative text, and four instances in direct speech. What these figures show conclusively is a comparatively infrequent occurrence of inversion, 16 instances in 300 clauses. The ascertained distribution of instances with subject–verb inversion is partly comparable with the corpus findings given in Biber et al. (1999: 926); here inversion is described as "more frequent in written registers than in conversation, with the highest frequency in fiction. ... At the other extreme, inversion is least common in conversation."

4.2 Types of fronting in the three text sorts

The distribution of the types of fronting of integrated clause elements in all three samples is shown in Table 6.

The table shows a larger number and diversity of the types of fronting than the two proposed at the beginning as the distinguishing features of the text sorts under study, viz. contextual fronting and emphatic fronting of a rhematic element. In addition to these two types, the examined samples displayed three other types, fronting of a deictic element, contextual contrastive fronting and emphatic fronting with two foci.

Contextual fronting, defined as fronting of a context-dependent element that constitutes a component of the thematic section, and hence contributes to the basic distribution of the principal FSP functions, theme, transition and rheme, appears to be the most common type of fronting in all three text sorts. The greatest prevalence of this type is shown in the narrative sample (nearly 91%, 80 instances out of 88), academic writing ranking second with 73.3% instances of contextual fronting (44 out of 60) and fiction dialogue showing the relatively lowest percentage of contextual fronting, viz. 67.85% (19 instances out of 28). These results fail to confirm the initial assumption of contextual fronting as a characteristic of academic writing. Yet at least a partial confirmation of the expected stylistic distinctions may be seen in the highest representation in fiction dialogue of the fronting of other types, which here account for 32.15% (9 instances out of 28), as compared with 26.7% (16 out of 60) in academic writing and 9% (8 instances

---

3   One of the realization forms of the presentation sentence (cf. Adam, 2013; Chamonikolasová and Adam, 2005; Dušková, 2015).

Table 6. Types of fronting in academic prose, fiction narrative and fiction dialogue

	Fronting of deictic element	Contextual fronting	Contextual contrastive fronting	Emphatic fronting of rheme	Emphatic fronting with 2 foci	**Total**
Wilson	–	29	10	1	–	40
Hergehahn and Henley	–	15	1	–	4	20
Academic Sample	–	**44**	**11**	**1**	**4**	**60**
Lodge narrative	–	38	2	2	–	42
James narrative	–	42	2	2	–	46
Narrative sample	–	**80**	**4**	**4**	–	**88**
Lodge Fiction dialogue	3	5	2	2	–	12
James Fiction dialogue	1	14	–	–	1	16
Fiction dialogue	**4**	**19**	**2**	**2**	**1**	**28**

out of 88) in fiction narrative. The fact that these figures show more similarity between the academic sample and fiction dialogue than between fiction narrative and fiction dialogue is due to the different content and sub-registers of the samples and partly also to the auctorial style.

As regards the syntactic functions of the contextually fronted elements, nearly all were adverbials. Of the other fronted clause elements, only two had non-adverbial functions, viz. a prepositional object in (12) and the postmodifier of a subject complement in (19).

In all three samples some of the instances of contextual fronting displayed subject-verb inversion (altogether 16 instances, see Table 5). In Quirk et al. (1985: 1377), this type, illustrated by *To this list may be added ten further items of importance*, is described as "very common both in speech and in conventional writing, often serving the function of so arranging clause order that end-focus falls on the most important part of the message as well as providing direct linkage with what has preceded." According to Biber et al. (1999: 926), quoted here on p. 85, inversion is least common in conversation and most frequent in fiction. The figures in Table 5 appear to support the more general statement of Quirk et al., but being small, do not offer ground for a founded comparison. On the other hand all the examples of contextual fronting with inversion corroborate the contributive role of inversion in textual cohesion.

The second initially proposed type, emphatic fronting of the rheme, defined as fronting of a context-independent element that constitutes the postverbal rheme in the regular ordering, was expected to favour fiction dialogue and fiction. It is represented by 7 instances most of which come from the narrative sample. The syntactic functions of the emphatically fronted elements are mostly adverbials, viz. 4 instances, (16), (27) a., b., (38); the other non-adverbial syntactic functions include 2 subject complements, (9) and (21), and one object (34) a. The prevalence of adverbials over other clause elements here again reflects the generally much higher frequency of occurrence of this clause element, due to its prevalent valency-independence, differing degree of integration, and semantic specificity and diversity, as compared with the constitutive elements, pointed out on p. 83 in Section 4.1.

Besides these two initially proposed types of fronting, the material under study displayed three other types: fronting of a deictic element, contextual contrastive fronting and emphatic fronting with two foci.

Fronting of a deictic element is a situational counterpart of contextual fronting: the speaker takes up, as a starting point, an element present in the situation of utterance. As follows from the nature of deixis, its province is informal speech, which is reflected in the material under study in the restriction of this type of fronting to the sample of fiction dialogue. This is presumably at least partly the reason for its being the least frequent type: there are four instances, three adverbials, (31) a., b., (35) a., and one object, (32) b.

Contextual contrastive fronting ranks second in the frequency of occurrence, next to contextual fronting but greatly separated from it in absolute figures: 17 instances (11, 4 and 2, respectively, in academic prose, fiction narrative and fiction dialogue) as against 143 instances of contextual fronting.

In contextual contrastive fronting the first contrasted element is context-dependent, but disengaged from context-dependence by one of the decontextualizing factors (cf. Firbas, 1995: contrast, selection, identification, summarizing effect and purposeful repetition; Stehlíková, 2016), here by contrast. By being contrastive, the fronted element acquires a heightened degree of communicative dynamism, hence constitutes a contrastive diatheme within the thematic section. Due to the nature of contrast, at least two items are intrinsically involved. Moreover, it is not only the two items that are contrasted, but also what is said about them. Hence contextual contrastive fronting typically occurs in parallel structures in which each unit carries two foci. In addition to the main focus on the context-independent predicative part (the intonation centre/nuclear tone is as a rule carried by the rheme), there is a second, minor focus on the contrastive diatheme. The FSP structure has two information peaks, realized by the contrastive diatheme and the rheme in the predicate (cf. divided focus in Quirk et al., 1985: 1378; Hajičová et al., 1998: 151). Compare examples (3), (7), (8), (13), (20), (23), (33).

The last type of fronting, emphatic fronting with two foci, differs from contextual contrastive fronting in that the first contrasted element is context-independent whereas in contextual contrastive fronting it acquires its context-independence owing to a decontextualizing factor. Since in emphatic fronting with two foci the contrastive element is introduced into discourse for the first time, it carries a higher degree of CD than the contrastive diatheme in contextual contrastive fronting, hence both contrastive items can be assigned the same degree of prosodic prominence. Clear examples of this type of fronting are (11), the first clause of (14) and (27) b.; example (38) is an instance of potentiality.

5. Conclusion

Considering the aim of the study in the light of the obtained results, it appears that both the proposed structures, and the relations between the structures and the three examined text sorts are more complicated and diversified than was assumed. First, the types of fronting found in the excerpted samples are more diverse. Still, the types found in addition to the proposed ones, deictic fronting, contextual contrastive fronting and emphatic fronting with two foci are based on the latter and may be regarded as their subtypes.

As regards the stylistic aspect of the different types of fronting, the initial assumption of contextual fronting as a distinctive feature of academic prose and fiction narrative was shown to be erroneous, contextual fronting being the most frequent type even in fiction dialogue. Moreover, the best represented sample in this respect was not academic prose, as was supposed, but fiction narrative. On the other hand, stylistically relevant distinctive features appeared in the distribution of the fronted integrated and non-integrated elements. In fiction dialogue fronted integrated elements were much less frequent than in the narrative parts and in academic prose. The small number of fronted integrated elements in fiction dialogue (28 instances out of 100) calls for a study based on a much more extensive material. A finding of stylistic, as well as of structural and textual interest appeared in the relatively frequent occurrence of contextual contrastive fronting in series of parallel structures, which contributed to the second rank of this type in the frequency of occurrence, even though greatly distanced in absolute figures from the contextual fronting.

While the results concerning the fronting of integrated clause elements in the three text sorts brought few stylistically notable points, features of stylistic significance appeared in the distribution and realization forms of non-integrated clause and sentence elements: conjuncts, disjuncts and interjections. Interjections – to which can be added dislocation and deictic reference – were exclusive features of fiction dialogue, while disjuncts and conjuncts showed significant differences in the distribution and occurrence of particular semantic roles and realization forms. An important stylistic factor was moreover demonstrated in the differences between the samples of the same text sort. The differences appeared to be due not only to the different subject matter and different sub-registers of the same text sort – sociobiological expository argumentative vs. psychological expository instructional, a humoristic vs. a dark novel, but also to the auctorial style, both in fiction, where it is expected as a matter of course, but also in academic prose, which is generally greatly standardized.

REFERENCES

Adam, M. (2013) *Presentation Sentences (Syntax, Semantics and FSP). Spisy Pedagogické fakulty Masarykovy univerzity* 162. Brno: Masaryk University.

Chamonikolasová, J. and M. Adam (2005) The presentation scale in the theory of functional sentence perspective. In: Čermák, J., A. Klégr, M. Malá, P. Šaldová (eds.), *Patterns. A Festschrift for Libuše Dušková,* 59–69. Prague, Faculty of Arts, Charles University, and Czech Modern Language Association (Kruh moderních filologů).

Biber, D., S. Johansson, G. Leech, S. Conrad and E. Finegan (1999) *Longman Grammar of Spoken and Written English.* Harlow: Longman.

Dušková, L. (2015) Syntactic forms of the presentation scale and their differentiation. In: *From Syntax to Text: the Janus Face of Functional Sentence Perspective*, 202–208. Reprinted from *Linguistica Pragensia* 8, 1998/1, 36–43.

Firbas, J. (1992) *Functional Sentence Perspective in Written and Spoken Communication.* Cambridge: University Press.

Firbas, J. (1994) Substantiating Daneš's view of givenness as a graded phenomenon. In: Čmejrková, S. and F. Štícha (eds.), *The Syntax of Sentence and Text*, 119–129. Amsterdam/Philadelphia: Benjamins.

Firbas, J. (1995) Retrievability span in functional sentence perspective. *Brno Studies in English* 21, 17–45.

Hajičová, E., B. Partee and P. Sgall (1998) *Topic-Focus Articulation, Tripartite Structures and Semantic Content*. Dordrecht: Kluwer.

Huddleston, R. and G. K. Pullum (2002) *The Cambridge Grammar of the English Language*. Cambridge: University Press.

Pípalová, R. (2000) Intersentential connectives – a textual analysis. *Acta Universitatis Carolinae, Philologica,* 5/1997, *Prague Studies in English,* XXII, 91–115.

Quirk, R., S. Greenbaum, G. Leech, and J. Svartvik (1985) *A Comprehensive Grammar of the English Language*. London: Longman.

Stehlíková, L. (2016) *Contextual Disengagement in FSP*. Doctoral dissertation. Brno: Masaryk University, Faculty of Arts.

Svoboda, A. (1981) *Diatheme*. Brno: Univerzita J. E. Purkyně, *Spisy Filozofické fakulty* 229.

Sources

Academic sample

B. R. Hergenhahn and T. B. Henley, *An Introduction to the History of Psychology*. 1st ed. 1986. Wadsworth CENgage Learning, 7th ed. 2014, United Kingdom, United States.

E. O. Wilson, *On Human Nature*. With a New Preface. 1st ed. 1978. Cambridge, Massachusetts, 2004, London, England.

Narrative sample

D. Lodge, *Changing Places: A Tale of Two Campuses.*1st ed. 1975. Penguin Books, Harmondsworth, 1978.

P. D. James, *Devices and Desires*. Faber and Faber, London, 1989.

Fiction dialogue

D. Lodge, *Changing Places: A Tale of Two Campuses.*1st ed. 1975. Penguin Books, Harmondsworth, 1978.

P. D. James, *Devices and Desires*. Faber and Faber, London, 1989.

PREPOZICE POSTVERBÁLNÍCH VĚTNÝCH ČLENŮ JAKO UKAZATEL FUNKČNÍHO STYLU

Resumé

Článek se zabývá otázkou prepozice postverbálních větných členů s cílem zjistit jejich distribuci a typy v odborném stylu, narativní části beletrie a v dialogu beletrie, reprezentujícím konverzaci. Zkoumají se dva typy prepozice, kontextuální a emfatická. Předpokládané rozdíly v jejich distribuci vycházejí z různého řazení aktuálněčlenských funkcí, které se v kontextuální prepozici shoduje se základním rozložením výpovědní dynamičnosti, zatímco v emfatické prepozici stojí v počáteční pozici réma. V prvém případě prepozice přispívá k textové kohezi, což je rys odborného stylu, v druhém případě intenzifikuje emfázi/emotivitu vyjadřovaného obsahu, což je rys konverzace. Beletristický narativ je podle předpokladu disponován k výskytu obou typů. Výsledky výzkumu ukázaly více prepozičních typů s rozmanitější strukturou a složitější vztahy mezi prepozičními strukturami a zkoumanými druhy textu.

Libuše Dušková
Department of English Language and ELT Methodology
Faculty of Arts, Charles University
libuse.duskova@ff.cuni.cz

ENCODING THE GLOBAL THEME IN RESEARCH ARTICLES: SYNTACTIC AND FSP PARAMETERS OF ACADEMIC TITLES AND KEYWORD SETS

RENATA PÍPALOVÁ

ABSTRACT

Research articles rank among the most prominent academic genres familiarizing their readers in a succinct way with the latest results of academic research. Due to the immense rate of publication these days, it has become increasingly difficult to gain reader attention. Among other things, such attention may be achieved by devoting particular care to the selection of a suitable title and perhaps reinforced with the help of the keyword set. Since titles and keyword sets are freely available and visible even in paid online journals, they are in open competition and serve a multitude of functions (e.g., enticing readers or raising their expectations). Crucially, they encode the Global Theme of Research articles, though in two distinct ways. This paper is based on a specialized corpus, composed of titles and keyword sets gathered from linguistic Research articles recently published by six renowned international journals. By examining their ideational, interpersonal, and textual functions, the author of this paper has sought to identify some of the prominent patterns and tendencies in their formulation, paying particular attention to their syntactic and FSP parameters.

Keywords: specialized corpus, research article, simple and hanging/compound titles, keyword sets, FSP, syntax

1. Introduction

Titles and keyword sets (KW sets) represent specific academic subgenres, marked by their brevity as well as their fixed initial or post-initial positions, and are clearly set off from the Research article (RA) itself. Both the subgenres seem to reflect the endeavour of the authors to delimit their research topic, to establish links between present and previous academic discourse and to advertise what they, in turn, add to the common pool of scientific knowledge, thereby claiming the general validity of their research for the entire community of practice. Admittedly, unlike the titles, the KWs do not represent an obligatory subgenre of RAs. However, their use appears to be increasing, at least in some fields, presumably thus reflecting the ever growing user-friendliness of academic

culture, especially of the writer-responsible (rather than reader-responsible) type, and, more generally, the impact of electronic media.

The two subgenres have a rather specific standing, since both are relatively independent (stand-alone, see Bhatia, 2004), but at the same time form an indispensable part of an RA (embedded, Bhatia, 2004). Both subgenres crucially encode the Global Theme of the RA, although in two distinct ways. They also cohere with one another and particularly with the RA itself. Indeed, they are expected to lure the target readers to the RA, but simultaneously deter others, whether expert or lay, lacking the necessary specialized knowledge. Among other factors at play here, they indicate from the outset that familiarity with the requisite terms and concepts is presupposed. Furthermore, both the KWs and the titles facilitate the automatic retrieval of the respective RAs, although the KWs appear to do so more directly.

Since the titles are freely available and visible even in paid online papers, they lend themselves to open competition and serve a multitude of functions. To borrow Halliday's framework (1985), they serve simultaneously all, the ideational, the interpersonal and the textual functions. To particularize further, the ideational function shows, among other things, that titles should identify the Global Theme of the RA, and possibly also suggest its epistemological tradition and disciplinary background. Likewise, they might encode the data, the methods used, the anticipated results, and the like. The interpersonal function stems from their capacity to entice readers, to raise their expectations and curiosity, to motivate or demotivate them instantly, thus potentially saving a lot of processing time (owing to their language economy). Through the title the writer also cooperates with the recipient by hierarchizing the RA content (foregrounding some of its elements) and by activating the relevant content frame/schema in order to facilitate and enhance the coherent reception of the RA. Furthermore, the textual role of titles shows their ability to signpost (label) the RA itself (endophoric link), reinforce the house style of the journal in a denominationally decorative way, establish intratextual and intertextual links with other titles from the same subgenres by triggering diverse cohesive chains which interrelate the titles with both the KWs and the actual body of the texts. In addition, they may establish interdiscursive links with similar academic subgenres, as well as with other academic genres dealing with the same or related topics (exophoric links).

Most of the above tasks can also be ascribed to the KW sets. However, while the titles seem to perform them fully, the functional load of KW sets appears to be somewhat more limited. For instance, unlike the titles, the KW sets do not fulfil the denomination role nor rank among those parts of RAs most likely to be cited in other academic studies. Moreover, they cannot anticipate the results. This restriction may explain why they do not represent an obligatory academic subgenre. Although the KWs also enter into interdiscursive relationships, their primary function appears to be intratextual and largely pragmatic.

2. Research preliminaries

2.1. Research available

Studied primarily in titlelogy (Baicchi, 2004), titles have been described as the "earliest point of contact between writer and reader" (Haggan, 2004), "combining informativeness with economy" (ibid.), as well as "consequential texts" (Gesuato, 2008) that determine whether the reader will continue to read. Their shape has been found to vary with respect to their fields.

The last three decades have seen a growing interest in academic titles, the research being conducted particularly, though not exclusively, within the scope of English for special purposes. Attention has been given to a wide range of parameters for academic titles. To name but a few, researchers have explored their phoricity (Baicchi, 2004); disciplinary peculiarities and cross-disciplinary variations (including differences between the soft and hard sciences, e.g., Anthony, 2001; Mendéz, Alcaraz, 2015; Cianflone, 2010; Nagano, 2015; Salanger-Meyer, Ariza, 2013; Sisó, 2009; Soler, 2007); genre-specific characteristics and cross-generic peculiarities (e.g., titles of monographs, dissertations, RAs, e.g., Gesuato, 2008; Jalilifar, 2010; Soler, 2007); cross-language and cross-cultural distinctions (e.g., Blaheta, 2006; Busch-Lauer, 2000; Soler, 2011); their external or graphic features (e.g., Anthony, 2001); particular or unusual features of titles (e.g., direct speech, abbreviations; Pułaczewska, 2009; Mendéz, Alcaraz, 2015). Though most of the existing studies are synchronic, some diachronistic research is also available (e.g., Mendéz, Alcaraz, 2015).

As far as relevance for the present study is concerned, a number of authors have focused on the syntactic parameters of titles (e.g., Haggan, 2004; Soler, 2007; Wang, Bai, 2007; Cianflone, 2010; Jalilifar, 2010; Moattarian, Alibabaee, 2015), although the criteria upon which the writers establish their taxonomies have frequently been conflated and many title types have been identified by some of their most conspicuous features. Authors concur in stating that there are both single- and double-unit titles, the latter being sometimes referred to as compound (e.g., Haggan, 2004), hanging (e.g., Anthony, 2001), and colonic titles (e.g., Hartley, 2004). In her study, Haggan (2004) further features full-sentence, compound and nominal titles, the last class with or without modification. Cianflone (2010) includes nominal, full-sentence, compound and question titles. Jalilifar (2010) also proposes a classification based on syntactic encoding and discriminates between NP, VP, PP, full-sentence and compound titles.
As far as the author knows, no specialized research focusing on the FSP parameters of titles has been carried out to date. However, valuable insight can be gained, chiefly from FSP studies by some prominent Czech scholars, including Svoboda (1968, 1989), Firbas (1992), and Dušková (2015).

In contrast, it seems that KWs have as yet been explored mainly in terms of their capacity to facilitate the automatic retrieval of an RA.

2.2 Research aims

Given the immense functional load of these subgenres, which contrasts sharply with their brevity, it appears worthwhile to examine some of the leading tendencies in their formulation. This paper seeks to uncover some of their conspicuous syntactic and FSP characteristics and to compare relevant aspects of titles and KW sets as two distinct approaches to the Global Theme of the RA. It is based on a specialized corpus of linguistic RAs and tries to balance quantitative and qualitative findings. Although generally more attention is given to RA titles, the findings are set against the trends detected in the KW sets.

This exploration of syntactic and FSP parameters means the paper has several foci. Firstly, the aim is to bring to light the leading formulation tendencies in recent linguistic RA titles. Secondly, the same exercise is repeated in the case of the corresponding KW sets. Finally, the paper is designed to correlate and compare the findings for the two subgenres, viewed as two distinct embodiments of the Global Theme of the RAs.

2.3 Corpus

This paper is based on data gathered from six renowned, international, peer-reviewed linguistic journals, namely *Lingua* (hereinafter: 1), *Language & Communication* (2), *Journal of Pragmatics* (3), *Language & Speech* (4), *International Journal of American Linguistics* (5) and *Language* (6). The corpus was composed of research articles (RAs) exclusively, reviews and editorials being disregarded. To qualify for inclusion, the RA had to exhibit both a title and a set of keywords as well as an abstract. An attempt was made to select only authors displaying English-sounding names or those affiliated with a university established in an L1 English-speaking country (within the inner circle of the Kachruvian model).

In quantitative terms, from each of the above journals, 50 RA titles and the same number of KW sets were drawn. As a result, the specialized corpus comprised 300 recent RA titles and 300 KW sets, i.e., 600 specimens in all. Since the journals differed in the total of issues per year as well as in the number of native authors publishing their RAs in them, I explored the most recent papers (2015) first and, when necessary, looked at back issues from 2014 and earlier to assemble the same number of RAs from each journal.

With regard to length, the title sub-corpus naturally proved to be larger, accounting for 3,029 words altogether, while the KW sub-corpus was found to be more constrained, comprising 2,409 words, with the entire corpus amounting to 5,438 words. It is noteworthy that all the data are available for free online.

2.4 Graphic aspect

2.4.1 Titles
In the data, the titles seldom exceeded the graphic space of two lines, whatever the size of fonts adopted, in extreme cases spanning four lines. The mean length of the titles examined turned out to be 10.09 words. Naturally enough, the individual journals varied in this respect, with their mean lengths ranging between 8.26 words (5) and 11.62 words (2).

This seems to suggest that the contributors recognized, attuned to and co-constructed a kind of journal house style. The titles were constructed as single- or two-unit structures, with one exception verging on a three-constituent title: *(Almost) everyone here spoke Ban Khor Sign Language – Until they started using TSL: Language shift and endangerment of a Thai village sign language* (2). It should be pointed out that the boundaries of title units were signalled graphically (chiefly by punctuation marks, i.e., mostly by colons, or by some less common means, such as stops or dashes, with the punctuation being combined with, and potentially reinforced by, the immediately succeeding capitalization, unless the journal in question capitalized all the title content words anyway).

In the academic subgenre of RA titles, multi-word constructions represent an established norm. In the corpus, a single content word never constituted an RA title. Nevertheless, a single content word did on rare occasions form one part of a hanging/compound title, e.g., *Requests: Knowledge and entitlement in writing tutoring* (2); *The Combining of Discourse Markers – A Beginning* (3). The shortest titles in the corpus were detected among the single-unit ones and comprised only two graphic words, e.g., *Communicating content* (2); *Intercultural impoliteness* (3); but they proved to be rather marginal. Conversely, the longest titles turned out to be two-unit structures, composed in extreme cases of 25 graphic words: *Working the overall structural organization of a call: How customers use their position as leverage for gaining service representatives' assistance in dealing with service problems* (2); *Functional Load and the Lexicon: Evidence that Syntactic Category and Frequency Relationships in Minimal Lemma Pairs Predict the Loss of Phoneme Contrasts in Language Change* (4).

2.4.2 KWs

In contrast, the space provided for KW sets varied between one and two lines, with a single exception filling three graphic lines (2). Since the KW sub-corpus of 2,409 words comprised 300 KW sets and encompassed altogether 1,545 KWs (i.e., KW items), the mean number turned out to be 5.15 KWs per KW set, embodying on average 8.03 words, with a typical KW (unit) involving 1.56 words. With regard to the differences between the individual journals, the mean lengths ranged between 1.52 words (5) and 1.83 words (4) per KW item. It would appear that contributors either had to follow the respective style sheets, or else deliberately decided to adjust their KWs to established house-style conventions. As for the extreme corpus findings, the shortest KW set detected exhibited only three words (2), whereas the longest was made up of twenty words (4).

In the data, KW sets featuring five KWs were the commonest of all (35.33% of the KW sub-corpus), followed by six-item KW sets (28.33%) and four-item ones (23.67%). In fact, these three groups composed a clear majority of instances (87.33% of all), the remainder being rather rare. It should be noted that the extreme sets comprised three KWS (*Semantics; Pragmatics; Proposition;* (3) and ten KWs (*person shift, pronouns, narrative peak, narratives, evaluation, imperatives, tense shift, historical present, discourse structure, Silverstein hierarchy* (6), respectively.

3. Research results – titles

3.1 Title typology

As follows from Figure 1 below, there are two types of titles in terms of the number of their constituents – simple (i.e., single-unit) ones, e.g., *The Syntax of Sanskrit Compounds* (6) and hanging/compound ones (i.e., composed of two or even more graphically marked units), e.g., *Ensuring Language Acquisition for Deaf Children: What Linguists Can Do* (6). In the data, simple titles (52.33%) prevailed over hanging/compound ones (47.67%).

Syntactically, both the titles and their constituent parts may be divided into clause-type and phrase-type structures. In addition, some ambiguities between the clause-type and phrase-type construction may also be detected, e.g., *The nature of Old Spanish verb second reconsidered* (1). Within the clause-type titles/title parts, we can distinguish finite (see below), non-finite, e.g., *Analyzing Ethnic Orientation in the Quantitative Socio-linguistic Paradigm* (2), and verbless (i.e., elliptical) cases, e.g., *Time for a change?* (1). Furthermore, the finite ones may take either the shape of a main/independent clause, e.g., *Does Deliberate Metaphor Theory Have a Future?* (3), or that of a subordinate clause employed independently, e.g., *How language communities intersect* (2), an instance of a nominal relative clause. Conversely, the phrase-type titles are constructed either as, though rarer, prepositional phrases (PPs), e.g., *On reference work and issues related to the management of knowledge:* (…) (3), or as more prevalent noun phrases (NPs), e.g., *The Role of Facial Motor Action in Visual Speech Perception* (4). Hence the syntactic realization of titles is rather varied and most seem marked, at least judging by typical independent structures in the written language. In fact, they can be arranged on a cline, since their variegated character is further diversified by punctuation, which may be conclusive (e.g., full stop, question mark), inconclusive (e.g., colon, dash) or missing altogether.

Figure 1. Title typology

A.	SIMPLE VS. HANGING/COMPOUND (also COLONIC) titles
B.	CLAUSE-type vs. PHRASE-type (+ AMBIGUITIES)
	i. CLAUSE-type: finite (main clause; subordinate clause), non-finite, verbless (elliptical);
	ii. PHRASE-type: noun phrase, prepositional phrase
C.	PUNCTUATION: conclusive, inconclusive, absent

In terms of the syntactic relationships involved, the titles exhibited predication, typically coupled with complementation, e.g., *Does space structure spatial language?* (6); *Understanding change through stability* (1); coordination, e.g., *Patwin Phonemics, Phonetics and Phonotactics* (5); modification, e.g., *Mechanism of Disyllabic Tonal Reduction in Taiwan Mandarin* (4) and apposition (manifested not only within title structures, as in *Polarity particle responses as a window onto the interpretation of questions and assertions* (6), but could also be ascribed to the relationship between the two parts of hanging/

compound titles). Exceptionally, a single, unmodified head noun was used to form part of a hanging/compound title, e.g., *The combining of Discourse Markers – A beginning* (3).

3.2. Titles – quantitative results

The corpus findings for title realizations are provided in Figure 2 and further discussed below.

Figure 2. Corpus findings for Simple and Hanging/Compound Titles and their realizations.

SIMPLE TITLES (52.33%)	TOTAL	REALIZATIONS		HANGING/ COMPOUND TITLES (47.67%)	TOTAL	CONSTITUENT REALIZATIONS	
PHRASES	85.99%	NP	85.35%	PHRASE + PHRASE	65.03%	NP + NP	60.84%
		PP	0.64%			NP + PP	4.19%
						PP + PP	0%
AMBIGUITIES	1.27%	(PHRASE/CLAUSE)		PHRASE + CLAUSE	28.68%	NP + CL	27.98%
						PP + CL	0.70%
CLAUSES	12.74%	NON-FIN	7.01%	CLAUSE + CLAUSE	6.29%	AMB + NON-FIN	0.70%
		FIN	5.73%			V-LESS + NON-FIN	0.70%
						FIN + NON-FIN	0.70%
						FIN + FIN	4.19%

As regards the simple titles, the majority (85.99%) were constructed as phrases, chiefly as NPs (85.35%) and marginally as PPs (0.64%). Three cases were potentially ambiguous between clause and phrase construction (1.27%). Clause-type titles accounted for 12.74%, with non-finite clauses (7.01%) slightly prevailing over finite ones (5.73%).

Looking at the feasible combinations of structures in the hanging/compound titles, a convincing majority was constituted by two phrases (65.03%), over a third by a clause and a phrase (28.68%) and a small proportion (6.29%) by two clauses (including ambiguities). However, when each of the two constituents was counted separately, phrases in hanging/compound titles amounted to 77.97% of all constituents, whereas clauses composed the remaining 22.03% of units. Hence, compared with simple titles, hanging/compound ones are marked by an increase in verbal features and thus dynamicity.

As for two-phrase hanging/compound titles (i.e., 65.03% altogether), they typically embraced two NPs (60.84%) and less commonly there was a combination of a PP and an NP (4.19%). In the data, two PPs were never juxtaposed to form a hanging/compound title. Furthermore, the corpus included 28.68% of instances where hanging/compound titles correlated a phrase (primarily an NP) and a clause-type structure, whether main or subordinate, and whether they featured a finite verb, a non-finite verb, or were indeed verbless. In such cases, the phrase tended to occur on the right, being preceded by a more dynamic clause construction. The reverse was attested to in only

one fifth of specimens, the most frequent group exhibiting an NP followed by a non-finite clause. Last of all, there were several hanging/compound titles (6.29%) employing a clause in each part. Three of these combined a main clause structure with a non-finite one, three coupled a subordinate clause with a non-finite one, and a single hanging/compound title matched a main clause with a subordinate counterpart. In the data, no single title correlated two main clauses. Admittedly, there was also a verbless clause matching a non-finite one, and a non-finite clause correlated with a structure potentially ambiguous between a clause/phrase. All in all, phrases (chiefly NPs) clearly prevailed, realizing both constituents of hanging/compound titles in 65.03% of cases, and one of the parts in 28.68% of specimens.

3.3 Titles – qualitative results

3.3.1 Clause-type titles/Title constituents

3.3.1.1 Main clause titles

Simple titles realized by main clauses were mostly marked by direct word order and the absence of terminating punctuation, e.g., *Hip-hop rhymes reiterate phonological typology* (1). The finite verb forms were all in the indicative mood. Hence such titles mostly corresponded to single-clause, declarative sentences, though devoid of the final period. However, the corpus also contained one instance of a *yes/no* interrogative with a regular question mark, e.g., *Does deliberate metaphor theory have a future?* (3).

With regard to main-clause constituents of hanging/compound titles, the situation turned out to be similar. Most of the finite verbs were in the indicative mood and were employed almost equally either in declarative clauses without the final period (eight specimens), e.g., *Onsets contribute to syllable weight: …* (6) or in interrogative clauses terminated by a question mark, mostly *yes-no* category (five instances), e.g., *Do you speak Arabic? …* (2), and less frequently *wh*-ones (two cases), e.g., *When are public apologies 'successsful'?...* (3). Two titles were encoded as imperative clauses with imperative verb forms, e.g., *Don't Listen With Your Mouth Full:…* (4).

In relation to the FSP aspect of titles (title constituents) realized as main clauses, they all proved to be in line with the rising CD (see Firbas, 1992: esp. 7–20, 66–67), observing end-focus and end-weight principles (e.g., *When are public apologies* 'successsful'?... (3). Wherever the verb was transitive, the Rheme coincided with the particular verb complementation, e.g., *Hip-hop rhymes reiterate phonological typology* (1); *Do you speak Arabic?* (2); *When in doubt, read the instructions* (1). (Note: The Rhemes are underlined).

To sum up the findings on the titles encoded as main clauses: although the data included declarative, interrogative and imperative clauses, the declarative turned out to be the most frequent, accounting for 56%, followed by the interrogative (32%) and imperative (12%). This presumably points to communicative priorities. The declarative category was used to anticipate some of the findings from the RAs, the interrogative foregrounded the interpersonal function, engaging the recipient in interaction, while the rare occurrence of the imperative could serve to exemplify some of the phenomena analysed in the body of the RA itself. On the whole, main clause constructions, constituting altogether 9% of the data, realized simple titles less frequently (5.73% of simple titles) than hanging/compound ones (6.29% of their constituents). Note: For the sake of accuracy, from now on

figures will only be provided for hanging/compound title constituents, for the realizations of the two title constituents frequently differed.

3.3.1.2 Subordinate clause titles (with finite verbs)

In the data, a subordinate clause never encoded a simple title, but was occasionally employed to form a constituent of a hanging/compound title. This was the case in 1.75% of the hanging/compound title constituents, and 1.13% of the entire corpus. The findings were all realized as nominal relative clauses employed to denominate, whether introduced by *How* or by *What*, e.g., *How language communities intersect* (2); *What linguists can do* (6); *How concepts and conventions structure the lexicon* (1).

Regarding their FSP aspect, all the instances proved to be in line with the rising CD, with the verb, or, where transitive, verb complementation, representing the Rheme, e.g., *How language communities <u>intersect</u>* (2); *What linguists <u>can do</u>* (6); *How concepts and conventions structure <u>the lexicon</u>* (1).

It should be noted, however, that there were two other specimens of finite verb clauses, this time not employed independently, but used to postmodify their nominal heads. In line with Greenbaum and Quirk (1990: 363–382), Huddleston and Pullum (2005: 183–191), and others, they are treated here as modifications, e.g., *The Woman who was a Fox:* (5); *Evidence that Syntactic Category and Frequency Relationships in Minimal Lemma Pairs Predict the Loss of Phoneme Contrasts in Language Change* (4). Interestingly, both were detected in hanging/compound titles.

All in all, titles featuring a finite verb (primary predication), constructed as main or subordinate clauses, proved to be generally rather rare, constituting 5.73% of simple titles, 13.29% of the first parts of hanging/compound titles, and 2.80% of second parts.

3.3.1.3 Non-finite clause titles

Statistically, the non-finite verb forms were more frequent than the finite ones, realizing 14.67% of the corpus titles. More specifically, they accounted for 7.01% of simple titles and 23.08% of hanging/compound titles (11.54% of constituents). In fact, they proved to be far more popular in the first parts of hanging/compound titles (13.99%) than in the second parts (9.09%). Naturally, non-finite titles were not concluded by terminating punctuation marks. The simple titles and the second parts of the hanging/compound ones were unambiguously marked by the absence of punctuation, whereas all the first parts of the hanging/compound titles realized in this way displayed a colon at the end. Moreover, most verbs were transitive, calling for complementation, typically by a direct object, e.g., *Exploring the source of differences and similarities in L1 attrition and heritage speakers competence:* (1); *Understanding change through stability* (1); *Inferring difficulty:* (4), and less commonly by a prepositional object, e.g., *Talking to a stranger* (2). Occasionally, a transitive verb was employed intransitively, e.g., *Relinquishing in musical masterclasses:* (3), or else an intransitive verb was preferred, e.g., *Drifting without an anchor* (4). Though potentially ambiguous, the last three instances were counted as non-finite structures chiefly owing to their dynamic lexico-semantic features.

Among other characteristics, the use of non-finite verb forms seems to testify to the researcher's modesty, which follows from the agency. Indeed, such structures at the least included the researcher among the implied agents, e.g., *Investigating the influence of atti-*

tudes on language accommodation (2); *Defining Nominal Compounding as a Productive Word-Formation Process in Chuxnabán Mixe* (5), but at times the agent was conceived more broadly, and was by no means confined solely to the producer, e.g., *Ensuring language acquisition for deaf children: …* (6); *Understanding change through stability:* (1). In fact, it appeared to be deliberately indeterminate, strategically vague, with its denotation varying from case to case, and potentially embracing the readers, the informants, the community of practice, the community as a whole, and the like.

In this context, it should be noted that non-finite verb structures may occasionally produce ambiguities, at least for non-specialists, e.g., *Voicing cooccurrence restrictions in Afrikaans* (6), here treated as premodification realized by a participle, as corroborated among others by the KWs: (*Afrikaans, grammatical change, obstruent voicing, cooccurrence restrictions, lexicon*). Interestingly, two hanging/compound titles in the corpus even saw coordination of non-finite verb forms: *Sharing and negotiating stance within the turn constructional unit* (3); *Managing and evaluating problematic interactions in a multilingual medical environment* (2). Another remarkable title involved near-parallelism: *Crossing into the past and crossing out the present* (2), verging on a pun and resembling the type of title practice that is common in journalism.

As for the FSP dimension of these non-finite clauses, one of the most conspicuous features appears to be their Theme-less nature. It follows, then, that the structures were typically "perspectived"/oriented towards Rhematic post-verbal complementation, and marked by rising CD. The following should serve as an illustration: *Working the overall structural organization of a call: …* (2); *Narrating participation and power relations in a social inclusion program* (2); *Examining the Acquisition of Phonological Word Forms with Computational Experiments* (4); *Avoiding emotivism* (2). Hence, such title constructions were encoded as Theme-less, presumably to foreground the Rheme, and to meet the need for language economy.

Naturally, non-finite verb forms involve a whole range of phenomena finely graded with respect to their share of dynamic features. In addition to the above instances, mostly complemented by direct or prepositional objects, there were gerunds apparently manifesting more nominal features, since they were premodified, e.g., *Morphosyntactic Marking on Intransitive Verbs* (5); *Advice-giving in newspaper weather commentaries* (3); *Lexical cloning in English:* (3); *Vowel patterning of Mormons in Southern Alberta, Canada* (2); *LF-copying without LF* (1). Occasionally the same pattern occurred even without modification: *Asking or Telling – Real-time Processing of Prosodically Distinguished Questions and Statements* (4). Due to their increased share of nominality, for the purpose of quantification, these cases were treated here as syntactic nouns (NPs), even though they should be clearly distinguished from regular (deverbal) nouns, e.g., *The Marking of Nonsingular Verbal Objects in Natchez* (5); *The processing of book polysemies* (1). It may be superfluous to note that non-finite verb forms can also be employed in attribute functions, e.g., *Factors affecting African American English usage and accommodation in adolescent peer dyads* (2), subsumed under NPs below.

3.3.1.4 Verbless titles

Another marginal type displayed potential finite-verb ellipsis, with the ellipsis justified by the terminating punctuation e.g., *Intact grammar in HFA?* (1). Strictly speaking,

these titles were constructed as NPs, but were always followed by a terminating punc-
tuation mark (a question mark exclusively in the data examined), which turned them
into verbless clauses devoid of a finite verb. In the corpus, such structures proved to
be generally peripheral, constituting 1.13% of all the data. In fact, no such cases were
detected in simple titles, but there were some findings in hanging/compound titles, all in
the first parts (1.75% of all the hanging/compound title constituents). Some of these con-
structions corresponded to *yes*/*no* questions, e.g., *Time for a change?* (1), while others to
alternative questions (Huddleston and Pullum, 2005:163) e.g., *Synchronic or Diachronic
Derivation?* (5).

In view of their FSP parameters, like the non-finite verb forms, these encodings seem
to be marked by deliberate, strategic vagueness. They are Theme-less, the whole elliptical
structure epitomizing the Rheme. Since, linguistically, these structures were in fact real-
ized as NPs, the Rheme proper, if there was one, fell mostly on the modifier, e.g., *Time for
a change?* (1); *Direct Off-record Requests?* (3); *Synchronic or Diachronic Derivation?* (5).
Nevertheless, some marginal instances were potentially vague, e.g., *A Typological Novel-
ty?* (5). Presumably, when featured in a journal (issue) focusing on typology, the Rheme
proper would fall on *novelty*, the analysis also being supported by the semantic factor
(Firbas, 1992: 41–65). Conversely, if the journal dealt with new discoveries in diverse
research fields, the Rheme proper would coincide with the modifier (for more on the
treatment of NP modification in FSP, see below).

3.3.1.5 Ambiguities

Among the potentially elliptical structures, there were also three borderline, periph-
eral cases (with the finally positioned past participle following the syntactic noun head),
in fact equivocal between clause- and phrase-like interpretations i.e., *Borders traversed,
boundaries erected?* (2); *Cherokee noun incorporation revisited* (5); *Precede-and-command
revisited* (6). The preferred clause-type treatment would involve finite verb ellipsis, while
the phrase-type view would consider these specimens as NPs featuring non-finite post-
modifications. Although this pattern was marginal, composing only 1% of all the data,
it was detected in both, simple (1.27%) and hanging/compound titles (0.35% of the con-
stituents). What opened space for possible elliptical interpretations was especially the
dynamic features of the non-finite verb form and/or the word order, once reinforced also
by a question mark. By and large, though infrequent, such title encodings only added
to the rich scale of conceivable title realizations, epitomizing their strategic vagueness.

In terms of their FSP, it seems that all these instances were couched in line with the
rising CD with the Rheme falling on the final *-ed* form. Such an interpretation seems to
be also corroborated by the prosodic and semantic factor (i.e., the lexical meaning of the
non-finite verb, in fact recurrent: *Cherokee noun incorporation revisited* (5); *Precede-and-
command revisited* (6) and in one case it seems justified by, inter alia, the two less dynam-
ic near-synonyms employed in parallel initial Thematic functions, which only confirms
the context-independence of the latter participle, e.g., *Borders traversed, boundaries erect-
ed?* (2). If we adopt the clause view, in this instance the first element appears to take the
subject function, and the latter that of a verb-less predicate.

Both the verbless (elliptical) and the potentially ambiguous title constructions virtual-
ly embody the transition between the clearly hierarchized FSP communicative fields, that

of the clause and that of the phrase, with the latter, as has been shown by Svoboda (1968), representing but a constituent of the former.

3.3.2 Phrase-type titles/Title constituents

3.3.2.1 Prepositional phrases (PPs)

Occasionally, titles or their constituents were encoded as prepositional phases (PPs). These structures realized mostly the first parts of hanging/compound titles (6 cases), once a simple title, and once the second part of a hanging/compound title. Altogether, this pattern turned out to be infrequent, accounting for only 2.67% of titles (i.e., 0.64% of simple and 2.45% of hanging/compound title constituents). The most usual preposition proved to be *on* (four specimens), e.g., *On reference work and issues related to the management of knowledge:* (3), followed by *beyond* (two cases), e.g., *Beyond truth conditions:* (6), with other prepositions featuring only once: *About bound and scary books* (1); *In your dreams* (3). Interestingly, all the PP realizations of titles were attested to in only three journals (1, 3, 6). Potentially, many PPs in titles might be interpreted as special cases of pragmatic ellipsis of the respective nominal heads, leaving the communicatively more significant post-modifications stranded, with such implied heads as *study, research article, analysis*, etc. Nevertheless, in the present study, these cases were kept separate and not counted as elliptical.

With regard to the FSP aspect of such titles, the researcher can draw chiefly on Svoboda (1968), who included PPs within the broadly conceived communicative field of NPs. Thus, it seems that in such titles, Rhemes (Proper) fall on the modifiers, if there are any at all, and in their absence, they coincide with the nominal heads of the PPs, e.g., *On ref-erence work and issues related to the management of knowledge* (3); *On the linguistic effects of articulatory ease, with a focus on sign language* (6); *In your dreams* (3); *Beyond sarcasm* (3). Hence, such structures proved to be mostly in line with the rising CD.

3.3.2.2 Nominal titles (NPs)

However, the decisive proportion of titles, both within the phrase-type category and within the entire corpus of linguistic RAs, were constructed as noun phrases (NPs). In fact, this central category accounted for 85.35% of the simple titles, 65.73% of the first parts of hanging/compound titles and 85.31% of their second parts.

The overwhelming majority of such NP titles (95.72%) displayed modification, the remaining 4.28% being devoid of modifiers. Interestingly, although the corpus comprised a few non-modified, bare nominal heads, they never realized a simple title, presumably because they were perceived as extremely short and under-informative. However, they were employed as components of hanging/compound titles. The corpus involved four such isolated syntactic nouns (1.14% of nominal titles/constituents), of which three constituted the first parts of hanging/compound titles and one a second part, e.g., *The Combining of Discourse Markers – A Beginning* (3); *Chitimacha: A Mesoamerican Language in the Lower Mississippi Valley* (5).

Coordinated bare syntactic nouns, devoid of modification, e.g., *Form and Function* (5); *Asking or Telling* (4), turned out to be more numerous (accounting for 3.85% of hanging/compound title constituents and for 2.48% of the entire corpus). They were distributed rather evenly between the first and second parts of hanging/compound titles. This seems to point to some conventional length limits imposed even on the constituents

of (hanging/compound) titles which apparently govern their stylistic acceptability, not to mention communicative efficiency.

Modified coordinated nominal heads were found to be more abundant as title realizations, since they made up over 17.42% of the nominal titles/constituents. They were distributed almost equally between single-unit (30 instances) and hanging/compound titles (31 cases). In terms of their attributes, some appeared as premodifiers exclusively, e.g., *Reformulative appositions and clausal ellipsis* (1), or predominantly (considering the number of content words employed in premodification vs. postmodification), e.g., *Resumptive pronouns, structural complexity, and the elusive distinction between grammar and performance* (1), more were distributed equally between pre- and postmodification, e.g., *Turn order and turn distribution in a multi-party storytelling* (3), and still others appeared in postmodification only, e.g., *The pleasures and possibilities of roadside shop talk in Tamil Nadu, India* (2) or primarily, e.g., *Language shift and endangerment of a Thai village sign language* (2). When one disregards the coordinated heads with an equal share of pre- and postmodification (9 specimens in all), then those premodified and primarily premodified instances (21 in all) are clearly outnumbered by the postmodified and largely postmodified cases (31 in all), the ratio being approximately 2:3. In this respect, the findings for the simple and hanging/compound titles proved to be similar, even though in the former the rate was somewhat less contrastive in favour of premodification.

The most decisive proportion of nominal titles were those exhibiting a single nominal head with premodification, postmodification, or a combination of the two (i.e., mixed type). Such a pattern was evident in 104 simple titles, 66 first parts of hanging/compound titles, and 104 of second parts. Some displayed solely premodification, e.g., *Intercultural impoliteness* (3) or primarily premodification, judging by the number of content words in the respective functions e.g., *Collaboratively organized stancetaking in Japanese:* (3). Others featured pre- and postmodification equally, e.g., *A pragmatic approach to anonymity* (3). Still others preferred postmodification, e.g., *A case study of Chinese participant perspectives* (3) or displayed it exclusively, e.g., *Effects of Age, Sex, and Syllable Number on Voice Onset Time* (4).

With respect to simple titles, premodification (5 instances) was outnumbered by postmodification (39 cases), but the mixed category, combining pre- and postmodification, had a crucial share (60 specimens in all). Within this mixed category, three subgroups were distinguished, based on the relative proportion of content words in their premodification and postmodification parts respectively. The nominal titles with predominant premodification and those with an equal share of pre- and postmodification (8 cases each) were clearly outnumbered by those with postmodification exclusively or primarily (44 instances). Hence, we can conclude that within simple nominal titles, postmodification dominated convincingly.

As regards the first parts of hanging/compound titles, the NPs were generally less frequent (66 cases altogether). Those solely displaying premodification and those featuring only postmodification were represented equally (17 items each), with the mixed type almost double the rate (32 cases). To gain a deeper insight, the mixed group was explored further, which revealed a decisive rate of postmodification. Indeed, there were three first parts where premodification prevailed, eleven first parts where pre- and postmodification were used equally, and 18 instances where postmodification had the crucial share.

Hence, even in the first part of hanging/compound titles, postmodification outnumbered premodification, though less so than in simple titles.

An examination of the second parts of hanging/compound titles reveals a remarkable growth in nominality, with as many as 104 nominal title parts. Among them, premodification (9 instances) was by far outnumbered by postmodification (54 cases), with the mixed type falling in between (41 examples). On closer scrutiny, however, within the mixed type group, title parts featuring mostly premodification (3 specimens) or those where pre- and postmodifications were balanced (8 items) were strongly surpassed by those where postmodification prevailed (30 second parts). Hence, postmodification turned out to be the leading pattern in the second parts of hanging/compound titles.

If we combine the findings for simple titles and both parts of hanging/compound titles, we find that of the 270 instances of NPs with single nominal heads and modification, there were 15.29% displaying premodification, 41.76% exhibiting postmodification but mixed type instances with pre- and postmodification made up the largest share of 42.95%. Further analysis of the mixed group confirmed the decisive proportion of postmodification. Indeed, nominal titles with premodification or prevailing premodification put together covered 18.82% of instances, those with an equal share of pre- and postmodification accounted for 11.18% of all, and nearly three fourths (70% of the total) showed a decisive share of postmodification or postmodification exclusively.

To summarize all the findings for the 350 NP/title parts: there were 4.28% of titles not featuring any modification at all, but the overwhelming majority (95.72%) of the nominal heads were modified. Within the latter group, there were 49 instances of premodification, 127 cases of postmodification, but the most remarkable proportion was the mixed type – 159 specimens in all. It follows from the data that heavily modified nominal titles represent the norm rather than the exception. In order to identify the tendencies, further investigation was undertaken within the mixed category. The results show that there were 18.86% of titles/title parts displaying premodification or a prevalent share of premodification. There were 10.29% of titles/title parts exhibiting an equal share of the two types of modification. However, a clear majority of titles/title parts, 66.57% of all, were marked by postmodification or primarily by postmodification.

With respect to the FSP aspect of NPs, this has been given particular attention in several studies, including e.g., Svoboda (1968: 49–101, 1989: esp. 101–104), Firbas (1992: esp. 14–20, 88–97), and Dušková (2015:135–159, 335–349, 362–377). Svoboda (1968), whose framework here is preferred to that in his later treatment, posits a hierarchy of distributional fields, with an NP constituting a distributional field of its own, embedded in the communicative field of the clause as its constituent. The author shows that unless a modifier is context-dependent, the head of the NP is Thematic, whereas the modification is more dynamic, and hence Rhematic. It should be noted that such Communicative Dynamism holds irrespective of whether the NP features pre- or postmodification. Recently, these findings have been confirmed and further developed by Dušková (2015: 349) who observes "a tendency for the noun phrase to be construed with postmodification at the first occurrence and premodification when reiterated."

Hence, the titles being contextually independent (unbound), the dominant FSP pattern proved to be the Rhematic modifier, irrespective of whether it was construed as a premodifier or as a postmodifier, coupled with the Thematic head noun, e.g., *Mecha-*

nism (Theme) *of* <u>*Disyllabic Tonal Reduction in Taiwan Mandarin*</u> (Rheme) *(4); Implications of* <u>*an Exemplar-Theoretic Model of Phoneme Genesis:*</u> *A* <u>*Velar-Palatalization Case Study*</u> (4). Interestingly, many Thematic head nouns were often reiterated across the titles (e.g., *account, analysis, approach, case, change, control, effect, evidence, function, identity, implication, problem, relation(ship), role, sketch, study, type, typology, use, variation, view*), both within and across individual journals, which further corroborated this tendency and disclosed an established pattern. Noteworthy, too, was the fact that such head nouns usually denoted general academic concepts and were not featured in KWs.

In view of the FSP theories cited above, the results of the present research, which found postmodification in nominal titles as the central pattern, can hardly be surprising. Indeed, being the very first academic subgenre in an RA, the title is not contextually dependent/bound. Therefore, its NPs should duly give preference to postmodification. It may be worth noting that postmodification also appears to offer comparably more space for particularization on the first mention, since the number of stylistically plausible premodifying elements seems to be rather limited.

3.3.2.3 Factors for premodification

Given the fact that titles are not contextually dependent structures, what seems surprising is that some authors employed premodification in NP titles from the outset. The investigation uncovered a number of possible reasons for such premodifications. In what follows, some of the factors (morphological, lexical, syntactic, FSP, stylistic, pragmatic, processing and graphic) will be briefly discussed individually, even though they should not be seen as tightly compartmentalized and mutually exclusive. Rather, they often seem to work in concert with one another.

Premodifying elements in titles were frequently realized by adjectives whose prehead position is grammaticalized in English, although exceptionally they may be found postposed e.g., *Narrative illocutionary acts* <u>*direct and indirect*</u> (3), due to coordination and other reasons, and thus attain the Rheme Proper function. Some adjectives formally coincide with non-finite verb forms, e.g., *The* <u>*clichéd*</u> *juxtapositions and* <u>*pleasing*</u> *patterns of political advertising* (2), and their postposition would change the meaning:* *patterns pleasing political advertising.* Interestingly, the above example illustrates the general scarcity of evaluative attributes in academic titles, here employed in parallel structures, presumably with an attention-drawing, foregrounding effect.

Some titles featured well-established, pre-fabricated terminological collocations and bundles, treated as an integral part of general or disciplinary background knowledge or as an indispensable scientific toolkit. They were constructed as NPs, with their head nouns premodified. In such instances, the fixed order of the constituents and the established character of the terminological collocation/bundle was confirmed when the title was compared with the respective KW set, since the same NP was usually detected even there, e.g., <u>*Relative clause*</u> *attachment in German, English, Spanish and French*; KWS: <u>*Relative clause*</u> *attachment; Cross-linguistic comparison; English; French; German; Spanish;* (1); (vs. ** An Attachment of relative clauses*). The existence of premodification in such bundles is indicative of previous research on the topic and hence testifies to the embeddedness of the bundle in the terminological network to such an extent that familiarity with such a complex term may be taken for granted and presupposed.

Sometimes, chiefly in cases with heavy modification of the head noun, part of the modifiers was preposed and part was postposed, presumably to mark the FSP juncture and set off the Rheme from the Rheme Proper, e.g., *A Stress "Deafness"* (Rheme) *Effect* (Theme) *in European Portuguese* (Rheme Proper) (4).This arrangement might also have been motivated by the tendency to keep the semantically rather homogeneous groups of modifiers together and to dissociate them by means of the medially positioned nominal head, e.g., *A Consonant/Vowel Asymmetry in Word-form Processing* (4), rather than **Asymmetry between Consonants and Vowels in Word-form Processing* (4).

Syntactic factors for the use of premodification in NP titles may also include the attempt to reduce the piling up of postmodifiers, e.g., *Vowel Patterning of Mormons in Southern Alberta, Canada* (2), rather than **Patterning of Vowels of Mormons in Southern Alberta, Canada.* This circumstance may go hand in hand with the need to prevent excessive use of function words and stylistic long-windedness. Ultimately, such considerations as ease of processing and pragmatic clarity may also be at play, precluding false interpretations, garden paths, and the like: e.g., *Biomechanically Conditioned Variation at the Origin of Diachronic Intervocalic Voicing* (4) rather than **Variation Conditioned Biomechanically at the Origin of Diachronic Intervocalic Voicing.*

In some specimens, premodification in titles would have been seen as less conspicuous, whereas postmodification appears more explicit, communicatively richer and easier to process, even though correlated with premodification in the KW sets. Naturally enough, such postmodifications proved to be longer and prosodically heavier, e.g., *Duration of American English Vowels by Native and Non-native Speakers: Acoustic Analyses and Perceptual Effects*; vs. KWs: *Vowel duration, duration pattern, non-native speakers, vowel intelligibility* (4); *Communication between hospital doctors: Underaccommodation and interpretability*; vs. KWs: *Underaccommodation; Interpretability; Health Communication; Inter-professional Communication* (1).

In fact, most premodifiers fell within a relatively restricted range of semantic fields, including those denoting particular languages, dialects, disciplines, theories, and so on, which were apparently perceived as an indispensable part of the necessary background knowledge, and by no means meant to pose a processing challenge on first mention, e.g., *Arabic interdialectal encounters: …* (2); *A Social psychological perspective* (2); *Morphosyntactic complexity: …* (6); *… : A relevance-theoretic reappraisal* (2). Given the fact that all the journals dealt with linguistic issues, it seems interesting that some of the fixed, established terminological clusters displayed even semantically rather redundant premodifiers, such as linguistic, language, etc., e.g., *Language variation and ethnic identity* (2); *Linguistic repertoire and ethnic identity in New York City* (2); rather than **Variation of language…*Repertoire of language.* Significantly, such premodifiers were at times missing altogether, presumably due to compression and language economy, e.g., *The acquisition of prosody in American Sign Lang*uage, as against the respective KW set: *language acquisition, sign language, prosody, phonology, American Sign Language* (6).

Moreover, English being an analytic language, preposing a noun before another noun converts the former into a syntactic adjective. Hence, wherever the nominal term seemed well-established, it was simply preposed before another noun and thus converted, and no derivation process had to take place (e.g., *Reference Constraints and Information Structure Management in Kokama Purpose Clauses* (5); rather than ** referential constraints*; ** infor-*

mational structural management…, although the two structures are virtually equivalent in meaning. Sometimes the comparison between the KW sets and titles did nevertheless reveal derivation, e.g., *Reformulation* (as a KW) as against the title: *Reformulative apposition and clausal ellipsis* (1), typically traced in denominal adjectives. In this case the two structures would not be semantically interchangeable. Obviously, extensive use of derivational suffixes or function words (in postmodifications) could be undesirable, since it would counteract the endeavour to keep the titles short.

Hence it seems that for premodifying functions in titles, the authors were prone to select chiefly notions which were considered well-established in the community of practice sharing the specialized knowledge, taken for granted at least to some extent, or even partly redundant, while for the postmodifying functions they opted for those that were communicatively more significant. This, however, did not exclude occasional use of various stylistic devices (such as parallelism) through which even the pre-position becomes more attention-drawing.

4. Research results – keywords

4.1 Keyword typology

The above findings on RA titles will now be set against the background of KW tendencies. In fact, similar to academic titles, the KWs also fell into two syntactic groups: single (graphic) words (hereinafter simple KWs), e.g., *Homophones* (4), as well as multi-word combinations, chiefly NPs (hereinafter complex KWs), e.g., *speech perception* (4). The simple KWs (45.18% of all) were outnumbered by the complex ones (54.82% in total). Nonetheless, two journals defied this tendency, as one exhibited an equal share between simple and complex KWs (1) and another (5) in fact reversed the ratio. Hence, even KW sets appear to be affected by house style patterns. The simple KWs were mostly syntactic nouns (or were converted into syntactic nouns, which was indicated co-textually, e.g., by their mere positioning among other syntactic nouns, or by other means, such as graphic, e.g., 'Now'(1); 'Then' (1), with very few exceptions.

4.2 Keywords – quantitative and qualitative analysis

Comparing the syntactic relationships within the complex KWs with the titles, one can notice some conspicuous patterns, namely the absence of predication (and complementation), a marginal share of both coordination (e.g., *onset and coda consonant* (4) and apposition (e.g., *False Cognates (Camouflaged Forms)* (2), with the two constituting solely 1.06% of the complex KWs; and the clearly dominant role of modification (e.g., *historical linguistics* (5), (6), accounting for 98.70% of the complex KWs. It should be noted that the corpus also included two odd instances (0.24%, e.g., *outside-in* (3).

Further investigation showed that within complex KWs with modification (98.70%), premodification (*phonological encoding* (4) prevailed strikingly over postmodification (*ideologies of difference* (2), since the percentage of instances of the former was 96.11%, compared to only 2.59% of the latter, including cases where postmodified KWs were also

premodified (*syntactic realization of illocutionary forces* (1). Interestingly, three specimens of the former premodification group even exhibited coordination of premodifiers (*onset and coda consonant* (4). The research revealed that the number of premodifiers was mostly limited to one, since 83.29% of all instances displayed only a single premodifying item. The incidence of two premodifiers (*Discourse temporal connective* (1) was much lower (14.49%), although there were three additional cases with coordination of pre-modifiers (0.37%). The rate of KWs with three premodifiers (*New York City English* (2) proved to be noticeably low (1.48%), and the proportion of heads with four premodifiers (*Vaupés River Basin linguistic area* (5) negligible (0.37%). Hence the number of premodifiers in KWs seems to be inversely related to their frequency. Tellingly, if we disregard the simple KWs, the average length of the complex ones corresponds to 2.02 words.

5. Discussion

5.1. Comparison of titles and kws

Titles were composed mostly of one or two units, and displayed all syntactic relationships, viz. predication, complementation, coordination, modification and apposition. They entered intertextual (e.g., very likely to be cited) as well as intratextual relationships. In contrast, a single KW set embraced a number of KW units, in the data between two and ten, typically five. The scale of syntactic relationships proved to be restricted, with predication and verb complementation not featured at all. A prototypical KW corresponded to a modified NP. KW sets seem to be motivated rather pragmatically, and, unlike the titles, they are not likely to be cited. Furthermore, the KWs are syntactically more uniform and stereotypical, whereas titles tend to be more diversified, in line with the writer's communicative intent, creativity, style, strategy, and the like, allowing for the integration of core disciplinary concepts in one or two title structures. Hence, in the titles the mutual relationships between the concepts are marked explicitly. On the other hand, KW sets are characterized by atomization, the mutual relationships between the individual KWs in sets being only implied, inferable with the help of background disciplinary knowledge, since their mere juxtaposition is expected to activate the particular cognitive frame, with the formulation and sequencing of KWs failing to suggest explicitly any hierarchy or emphasis.

Notwithstanding the differences noted above between titles and KWs, in the data, nominality proved to be one of the shared features, with the modifiers carrying the Rhematic functions. It was shown that an overwhelming majority of title NPs featured postmodification or largely postmodification. This convincingly corroborates the tendency observed by Dušková (2015: 347) for NPs to be construed first with postmodification and, when reiterated, with premodification. Indeed, the NPs are first mentioned in the titles and when re-expressed in KWs, they have already become context-dependent. However, other factors may also be at play. For example, through their integration into one or two syntactic constructions in titles, some heads of various well-established terminological bundles, listed mostly as disparate KWs, were frequently deprived of their head (NP) status and turned into modifiers of title heads instead, e.g., KW: *Chinese attitude-bearing*

wh-questions; Second language; Computational complexity; L1 transfer vs. Title: *The effect of computational complexity on L1 transfer: Evidence from L2 Chinese attitude-bearing wh-questions.* (1) In this context the role of the recurrent nominal heads, as a rule missing in KW sets, seems to be crucial (e.g., *effect*, *evidence*, see above). However, the analysis conducted here suggests a variety of other reasons for the occasional use of premodifying functions in NP titles on their first mention.

5.2. Global theme encoded in titles and kws

In Pípalová (2008 a, b: esp. 99–111) the content aspect of Textual Themes, including the Global Theme, is conceived of as composed of three cognitive layers, resembling a pyramid. The narrowest layer embodies its most conspicuous, foregrounded element(s), the central layer embraces a number of interrelated, regularly co-occurring elements organized in a cognitive structure or content frame. The lowest and broadest layer, simultaneously the most diffuse of all, involves all the backgrounded elements, established in the particular speech event.

With regard to the way the Global Theme was found encoded in the title data, several approaches were noticed. Some authors decided to opt for the formulation of the Theme in the narrowest sense, identifying only its most conspicuous, foregrounded element, e.g., *Gitsan Modals* (5). Such practice was in line with deliberate strategic vagueness, but might simultaneously leave the recipient all at sea in many relevant respects. Most researchers therefore favoured the Global Theme in its broader meaning, which, apart from identifying the Theme in the narrow sense, also contextualized it within the particular disciplinary or research cognitive frames: e.g., *A Corpus-based Study of Fillers among Native Basque Speakers and the Role of Zera* (4); *A sociolinguistic view of null subjects and VOT in Toronto heritage languages* (1); *Consonants are More Important than Vowels in the Bouba-kiki Effect* (4). As a result, the titles proved to be longer, and being packed with more information and in line with the densification of discourse, they not only provided the reader with the Theme in the narrow sense (*fillers*), but also suggested the disciplinary background (*sociolinguistic view*), identified the informants (*Native Basque Speakers*), the approach (*Corpus-based*), the data (*Toronto heritage languages*), anticipated some of the results (*Consonants are More Important than Vowels in the Bouba-kiki Effect* (4), and the like. Hence, such patterns seem to encode the top and central layers of the Global theme framework. Elements of the lowest layer, which encompasses the established, backgrounded components of the communicative event, such as the producer, the recipient, the genre, research or discourse, were encoded for instance in non-finite titles (suggesting the implied producer/recipient, and the like, see above), e.g., *Examining the Acquisition of Phonological Word Forms with Computational Experiments* (4); in interrogative structures invoking the recipient, e.g., *Does deliberate metaphor have a future?* (3) and in titles constructed as PPs, implicitly pointing to the research, genre or discourse, e.g., *On reference work and issues related to the management of knowledge:* (3).

In contrast, KW sets seldom encoded elements of the lowest Theme layer, but did feature components of both the top and central layers (KW: *Sociophonetics* (central layer of the Global theme, disciplinary background); *Canadian English* (Central layer, data);

Mormon dialect (central layer of the Global theme, particular data and informants*); Vowels* (top layer of Global theme)*; Vowel raising* (top layer of Global theme)*,* as against the Title: *Vowel patterning of Mormon in Southern Alberta, Canada* (2). Compared with the Title encodings, the major difference seems to be the failure of KWs to hierarchize such elements and layers. Instead, they are merely juxtaposed for the sake of search engines and/or for a possible check of conceptual familiarity on the part of the potential processor.

6. Conclusion

This paper explored syntactic and FSP aspects of RA titles and KW sets in six renowned linguistic journals. In contrast to the relatively homogeneous and syntactically rather stereotypical KWs, the titles were shown to exhibit a wide range of syntactic relationships and structural realizations. The ultimate choice made appears to reflect the particular author's communicative intent, their originality, strategy or style, the house style of the journal and the patterns established within the particular community of practice. This is in sharp contradistinction to the KWs which rather than reflecting an individual author's creativity, accentuate the conventional pool of established disciplinary knowledge.

Notwithstanding the above differences between titles and KWs, in the data, nominality proved to be one of the shared features, with the modifiers (irrespective of whether premodifiers or postmodifiers) taking on the Rhematic functions. While in complex KWs, premodification convincingly prevailed (96.11%), attesting the prevalent Rh-Th arrangement, in the titles it constituted less than a fifth (18.86% of instances). This finding would appear to corroborate the tendency observed by Dušková (2015: 349) for the NPs, "to be construed with postmodification at the first occurrence and premodification when reiterated". Indeed, it was shown that the NPs are first mentioned in the titles, and when re-expressed in KWs, they have already become context dependent. However, the analysis conducted here also suggests a wide variety of reasons for the occasional use of premodifying functions in titles on their first mention. One source for such premodification can be traced to the prefabricated terminological bundles epitomizing part of the background disciplinary knowledge.

Even though the specialized corpus confirmed the dominant use of nominal titles, it simultaneously pointed to a whole scale of less frequent and even peripheral patterns. The wide range of realizations, serving specific communicative ends, and their respective frequency rates seem to reflect the difference in weight given by authors to at least two conflicting needs, viz. to reinforce the established, conventional patterns and the ambition to stand out from the crowd and to draw attention to one's own research.

Whatever the realization selected, most titles, in contrast with most KWs, were found to be in line with the rising CD, placing their Rhemes finally. Such an arrangement is well-suited to all three title functions, viz. the Ideational function (e.g., identifying the Global theme in the narrow sense and possibly also giving relatively ample space to contextualize it, to hierarchize its elements and show emphasis), the Interpersonal function (e.g., the endeavour to be reader-friendly, inviting rather than deterring the readership, and facilitating smooth, on-line processing) and the Textual function (e.g., establishing

intratextual, intertextual and interdiscursive relationships, with the Rhemes likely to trigger prominent, focal or text exhaustive, cohesive chains). The research demonstrated that very few titles encoded only the top layer of the Global theme content, while a decisive proportion were long enough to accommodate the central layer, and at times elements of the lowest also. In this way, the authors seem to strike the right balance between the need for strategic vagueness and that for discourse densification. Admittedly, the results of the present research cannot but be preliminary, being established on a corpus of limited size. More extensive research would therefore be necessary to verify the tentative results.

Symbols and Abbreviations

AMB: Ambiguities
CD: Communicative Dynamism
CL: Clause
FIN: Finite (verb)
FSP: Functional Sentence Perspective
KW: Keyword
NON-FIN: Non-Finite (verb)
NP: Noun Phrase
PP: Prepositional Phrase
RA: Research Article
V-LESS: Verbless
VP: Verb Phrase

REFERENCES

Anthony, L. (2001) Characteristic features of research article titles in computer science. *IEEE Transactions of Professional Communication* 44/3, 187–194.

Baicchi, A. (2004) The cataphoric indexicality of titles. In: Aijmer, K. (ed.), *Discourse Patterns in Spoken and Written Corpora,* 17–38. Philadelphia: John Benjamins Publishing Company.

Bhatia, V. K. (1993) *Analysing Genre: Language Use in Professional Settings.* London: Longman.

Bhatia, V. K. (2004) *Worlds of Written Discourse. A Genre-Based View.* New York: Continuum.

Blaheta, R. (2006) Anglický gerundiální knižní titul v překladu do češtiny [English book titles realized by gerunds in Czech translation]. *Časopis pro moderní filologii* 88/2, 65–81.

Busch-Lauer, I. (2000) Titles in English and German research papers in medicine and linguistics. In: Trosborg, A. (ed.), *Analysing Professional Genres,* 77–97. Amsterdam: John Benjamins.

Cheng, S. W., C.-W. Kuo and C.-H. Kuo (2012) Research article titles in applied linguistics. *Journal of Academic Language & Learning* 6, 1–14.

Cianflone, E. (2010) Scientific titles in veterinary medicine research papers. *English for Specific Purposes World* 30/9, 1–8.

Dušková, L. (2015) *From Syntax to Text: The Janus Face of Functional Sentence Perspective.* Prague: Charles University in Prague, Karolinum Press.

Firbas, J. (1992) *Functional Sentence Perspective in Written and Spoken Communication.* Cambridge: Cambridge University Press.

Gesuato, S. (2008) Encoding of information in titles: Academic practices across four genres in linguistics. In: Taylor, C. (ed.), *Ecolingua. The Role of E-corpora in Translation and Language Learning,* 127–157. EUT-Edizioni Universita di Trieste.

Greenbaum, S. and R. Quirk, (1990) *A Student's Grammar of the English Language*. Harlow: Longman.

Haggan, M. (2004) Research paper titles in literature, linguistics and science: Dimensions of attraction. *Journal of Pragmatics* 35/2, 293–317.

Halliday, M. A. K. and R. Hasan (1985) *Language, Context, and Text: Aspects of Language in a Social-semiotic Perspective.* Oxford: Oxford University Press.

Hartley, J. (2004) Planning that title: Practices and preferences for titles with colons in academic articles. *Library & Information in Science Research* 29/4, 553–568.

Huddleston, R. and G. K. Pullum (2005) *A Student's Introduction to English Grammar.* Cambridge: Cambridge University Press.

Hyland, K. (2000) *Disciplinary Discourses: Social Interactions in Academic Writing.* Harlow: Pearson Education Limited.

Jalilifar, A. (2010). Writing titles in applied linguistics: A comparative study of theses and research articles. *Taiwan International ESP Journal* 2/1, 29–54.

Méndez, D. I. and M. A. Alcaraz (2015) Astrophysics titles in scientific American magazine (1990–2014): Linguistic and discourse practices. *International Journal of Applied Linguistics and English Literature.* Available at: http://www.journals.aiac.org.au/index.php/IJALEL/article/view/1776 [2016-07-15].

Moattarian, A. and A. Alibabaee (2015) Syntactic structures in research article titles from three different disciplines: Applied linguistics, civil engineering, and dentistry. *Journal of Teaching Language Skills (JTLS)* 7 (1), 78/4, 27–50.

Nagano, R. L. (2015) Research article titles and disciplinary conventions: A corpus study of eight disciplines teaching language skills. *Journal of Academic Writing* 5/1, 133–144.

Paisa, C. E., J. P. da Silveira Nogueira Lima and B. S. R. Paiva (2012) Articles with short titles describing the results are cited more often. *Clinics* 67/5, 509–513. Available at: http://doi.org/10.6061/clinics/2012(05)17 [2016-04-19].

Pípalová, R. (2008a) *Thematic Organization of Paragraphs and Higher Text Units*. Prague: Faculty of Education, Charles University.

Pípalová, R. (2008b) On narrow and broad P-Theme paragraphs. *Linguistica Pragensia* XVIII/1, 37–51.

Puławewska, H. (2009) 'I bet they are going to read it': Reported direct speech in titles of research papers in linguistic pragmatics. *Lodz Papers in Pragmatics* 5.2, 271–291.

Salanger-Meyer, F. and M. Á. A. Ariza (2013) Titles are 'serious stuff': A historical study of academic titles. *JAHR* 7/4, 257–271.

Sgall, P., E. Hajičová and J. Panevová (1986) *The Meaning of the Sentence in its Semantic and Pragmatic Aspects*. Prague: Academia.

Sisó, M. J. (2009) Titles or Headlines? Anticipating conclusions in biomedical research article titles as a persuasive journalistic strategy to attract busy readers. *A Journal of English and American Studies* 39, 29–54.

Soler, V. (2007) Writing titles in science: An exploratory study. *English for Specific Purposes* 26, 90–102.

Soler, V. (2011) Comparative and contrastive observations on scientific titles in written English and Spanish. *English for Specific Purposes* 30, 124–137.

Stainton, R. J. (2007) The pragmatics of non-sentences. In: Horn, L. R. and G. Ward (eds.), *The Handbook of Pragmatics*, 3rd ed., 266–287. Oxford: Blackwell Publishing.

Svoboda, A. (1968) Functional perspective of the noun phrase. *Brno Studies in English* 7, 49–101.

Svoboda, A. (1989) *Kapitoly z funkční syntaxe* [Chapters from Functional Syntax]. Prague: Státní pedagogické nakladatelství.

Swales, J. (1990) *Genre Analysis. English in Academic and Research Settings*. Cambridge: Cambridge University Press.

Wang, Y. and Y. Bai (2007) A corpus-based syntactic study of medical research article titles. *System* 35, 388–399.

FORMULACE/UCHOPENÍ GLOBÁLNÍHO TÉMATU V ODBORNÝCH ČLÁNCÍCH: SYNTAKTICKÉ A AKTUÁLNĚČLENSKÉ PARAMETRY ODBORNÝCH TITULŮ A KLÍČOVÝCH SLOV

Resumé

Odborné tituly a soubory klíčových slov jsou ve studii koncipovány jakožto dvě různá uchopení či formulace globálního tématu odborného textu. Studie se věnuje zejména titulům anglicky psaných odborných článků a zkoumá jejich syntaktické a aktuálněčlenské tendence. Opírá se přitom o specializovaný korpus složený z nejnovějších odborných článků otištěných šesti renomovanými mezinárodními lingvistickými časopisy. Odborné tituly i klíčová slova se probírají z hlediska Hallidayovské funkce ideační, interpersonální a textové/textotvorné. Studie dále přináší zejména typologii syntaktických struktur, které se ve zkoumaném korpusu objevily, poznatky o jejich poměrném zastoupení a analýzu jejich aktuálněčlenských parametrů. Určitá pozornost se též věnuje porovnání odborných titulů a klíčových slov, a to jak z hlediska výrazných syntaktických, tak i aktuálněčlenských tendencí. Studie poukazuje zejména na převahu nominálních struktur v obou dílčích odborných útvarech, na podstatně pestřejší škálu syntaktických konstrukcí uplatněných v titulech oproti poměrně jednotvárné realizaci klíčových slov, na převahu postmodifikace u titulů na straně jedné a na typickou premodifikaci u klíčových slov na straně druhé, jakož i na obvykle stoupající škálu komunikačního dynamismu u odborných titulů v protikladu k opačné tendenci zjištěné u klíčových slov. Studie přináší též poznatky o poměrně značné délce anglických odborných titulů, která skýtá dostatečný prostor nejen pro nastolení samotného obsahového jádra globálního tématu, ale též pro jeho explicitní zasazení do příslušného odborného kontextu.

Renata Pípalová
Department of English Language and Literature
Faculty of Education, Charles University
renata.pipalova@pedf.cuni.cz

"… AND OUR STUDY MIGHT THEREFORE HAVE BEEN SLIGHTLY UNDERPOWERED": A CROSS-LINGUISTIC ANALYSIS OF HEDGING IN ENGLISH AND CZECH MEDICAL RESEARCH ARTICLES

JANA KOZUBÍKOVÁ ŠANDOVÁ

ABSTRACT

Attenuating devices are frequently employed in both spoken and written language to weaken the illocutionary force of utterances as well as their directness. In this way, they are associated with expressing linguistic politeness. In scientific discourse, hedges are largely used to express negative politeness since the authors aim to protect themselves against disapproving or critical remarks of text recipients. Linguistically, hedging devices are realised in various ways, e.g. as modal adverbs (*possibly, perhaps, probably,…*), modal adjectives (*possible, (un)likely, probable,...*), modal nouns (*assumption, possibility, suggestion,…*), modal verbs expressing possibility (*might, could, would,…*), epistemic verbs (*assume, seem, appear, suggest,…*), approximators such as *approximately* and *roughly*, etc. The aim of this paper is to analyse hedging devices in written academic discourse, in particular in English and Czech medical research papers published in medical journals with an impact factor, and compare them with respect to their types, occurrence and communicative functions they perform. In other words, to find out whether the pragmatics of hedging is or is not culture-specific. Hedging devices may be classified from various points of view. In this paper a modified version of Hyland's taxonomy (1998) was adopted.

Keywords: Hedging, medical discourse, research articles, cultural specificity, pragmatic function

1. Introduction

In the traditional approach to science, the view prevailed that the language of science should be as precise as possible, objective and matter-of-fact. This opinion has been gradually changing and nowadays, communicating scientific findings is more interactional. Scientists do not present their findings as set and invariable but they are more dialogic and involve readers in the cognitive process. However, not only the reader but also, and primarily, the whole personality of the author is included in the cognitive process, together with his emotion and, as Daneš points out, with a certain degree of irrationality and subjectivity (2000: 81). In this context, Daneš cites the German linguist Harald

Weinrich, who claims that a new finding may gain scientific currency only when it is spread "through a certain controlled process" in the scientific community and in this way exposed to criticism (Weinrich, 1995: 159). From this it follows that science is essentially a communicative process.

It is important to emphasise that scientific knowledge should not be reduced to the way findings are formulated. We must distinguish between processes leading to gaining scientific knowledge from the scientific discourse and text production as such when the researcher attempts to formulate and organise his/her thoughts for the recipient. At the same time, his/her imagination, close relationship to the subject matter of the research, excitement and aesthetic experience s/he undergoes are reflected, in a way, in scientific articles, as Daneš claims (2000: 82). Thus, a certain degree of subjectivity in these texts cannot be avoided.

As early as the 1940s Mathesius argued that every utterance carries its own factual meaning and that utterances also reflect speaker attitude to reality and his relationship to the recipient. All these factors – factual meaning, situational context, speaker attitude to reality, and his relationship to the recipient – form the overall semantic structure of the utterance (Mathesius, 1982: 93).

Hedging as a communicative strategy is frequently employed by speakers and writers to mitigate the illocutionary force of their statements. It is an important strategy also in academic writing because hedges contribute "to an appropriate rhetorical and interactive tenor, conveying both epistemic and affective meaning" (Hyland, 1998b: 349–350). Hedging is often associated with expressing linguistic politeness. In scientific discourse it is largely used to express negative politeness since the authors aim to protect themselves against disapproving or critical remarks of text recipients.

This study aims to analyse and compare hedging devices occurring in English and Czech medical research papers, with the focus on their types, occurrence and the communicative functions they perform. Hedges may be classified from various points of view. In this paper a taxonomy introduced by Hyland (1996, 1998a) was adopted since it stresses the so-called polypragmatic approach to the analysis of hedging devices. However, it was necessary to modify this taxonomy to a certain extent, as explained in Section 6 in greater detail.

In recent years some contrastive studies in different languages have appeared, focusing on the concept of hedging in academic discourse. It is an important area of language study since the use and functions of attenuating devices seem to be culture- and language-specific. To find out whether the pragmatics of hedging is or is not culture-specific is another aim of this study into English and Czech medical discourse.

2. Cross-linguistic studies dealing with hedging

Clyne's investigation into scientific texts written by German scholars has shown that they employ more hedging devices both in scientific articles written in German and in English than English native speakers (Clyne, 1991). Another study comparing English and German academic discourse has revealed that the main function of hedges in English articles is to soften the presented arguments, whereas in German it is predominantly "assertion and authority" (Kreutz and Harres, 1997).

English and Bulgarian academic texts were the subject of investigation of Vassileva (1997). She compared research articles written by Bulgarian scientific writers, articles written by English scientific writers, and finally, articles by Bulgarian writers in English, from the point of view of hedging distribution and form. Scientific texts written by English authors exhibited the highest occurrence of hedging expressions while the lowest number of them appeared in the papers written in English by Bulgarian authors.

Other contrastive studies on hedging have been carried out by Olmo, who compared English and Spanish medical discourse. His studies revealed differences between the two languages in the distribution of hedges. In general, they occur more frequently in English (Olmo, 2004, 2005). Martín-Martín (2008) investigated lexico-syntactic expressions with hedging function in English and Spanish as well. However, he focused on research articles from the field of psychology. He came to a similar conclusion as Olmo (2004), namely that hedging is slightly more frequent in English.

Atai and Sadr (2008) researched academic papers published in the field of applied linguistics written by English and Persian native speakers with the aim to examine the impact of language on the employment of hedging strategies in these texts. The research revealed that English native speakers employed a wider spectrum of linguistic means expressing tentativeness and indeterminacy.

Figueiredo-Silva (2001) compared academic texts written in English and Portuguese and found out that scientific articles written in English are more hedged than Portuguese texts, which appear to be more direct. A similar study (Yang, 2003) contrasting English and Chinese revealed that Chinese scientific writers use a small amount of plausibility shields and, on the contrary, a higher number of approximators, which results in their sounding more direct.

As we can see, the outcomes of the above-mentioned studies prove certain differences between various languages as regards hedging.

3. Hedging in Czech academic discourse

If we now turn to Czech academic discourse in regard to the employment of hedging expressions, we find that no systematic research has been conducted up to now apart from a few studies made by Daneš (2000), Čmejrková and Daneš (1997), Čmejrková et al. (1999), and a cross-cultural study into Czech and German academic discourse carried out by Dontcheva-Navrátilová (2013).

Generally, Czech academic discourse is typical of "modalization" and "authorial modesty", which are achieved most frequently by various modal expressions and first person plural forms (Čmejrková et al., 1999: 28–30). As Čmejrková and Daneš point out, in the case of modality and the use of hedging devices, there are distinct differences between individual authors and also between text genres. They speak of a continuum "between the pole of the straightforward and economical expository style and that of the narrative ("redundant") style. Czech expository texts (in the humanities) occupy positions on the scale nearer to the narrative pole […]" (Čmejrková and Daneš, 1997: 46). It is also important to emphasise that any text reflects the idiosyncrasies of the writers.

Further, Čmejrková and Daneš have found out that compared to English academic writers, Czech authors "formulate their pronouncements in a far less assertive, direct, and matter-of-fact tone" (1997: 44), which has also been confirmed by this study. The reasons why Czech scientific authors use mitigating devices may be modesty, adopting a defensive position, and showing distance from their claims (Daneš 2000).

4. Taxonomy of hedges introduced by Hyland

A relatively influential classification was proposed by Hyland (1996, 1998a). It is a sociopragmatic model designed for the examination of hedging devices in scientific texts (see Figure 1). Hyland works on the assumption that these expressions have various semantic readings depending, firstly, on the context in which they appear and, secondly, on the speaker or writer who has used them. Literally he claims that "linguistic features […] cannot be seen in isolation from particular socio-institutional activities and broader cultural understandings" (1998a: 157). Therefore, one and the same expression may carry more pragmatic functions. In this connection it is also important to stress that one cannot say that a given expression will always function as a hedge in all possible contexts. "Because indeterminacy appears to be an inherent feature of the epistemic use of language, an adequate account of hedging in scientific discourse must look beyond a mono-meaning model" (ibid.). Hence, attenuating devices require a " 'more-or-less' rather than an 'all-or-nothing' account" (ibid.).

Hyland claims that the writer of a scientific text anticipates the reader's opposition to his/her claims. This opposition is divided into two types, content-oriented and reader-oriented.

Content-oriented hedges soften "the relationship between propositional content and a non-linguistic mental representation of reality; they hedge the correspondence between what the writer says about the world and what the world is thought to be like" (Hyland, 1998a: 162). Reader-oriented hedges are focused on participants of the communication process. They "address the various dimensions of the social relationship between writer and reader […]" (ibid. 177). Even if Hyland makes this distinction and categorises both groups in greater detail, he highlights the fact that one of the major features of hedges is indeterminacy; so attenuating devices ascribed to one category may very often involve the meaning of another.

Hyland defines two categories of content-oriented hedges, namely, accuracy-oriented, which "involve the writer's desire to express propositions with greater precision in areas often subject to revision" (1996: 440), and writer-oriented, which "enable writers to refer to speculative possibilities while at the same time guard against possible criticism" (1996: 443). Recognizing different motivations and forms, Hyland distinguishes two subgroups of accuracy-oriented hedges, attribute and reliability hedges. Attribute hedges "enable writers to restructure categories, define entities, and conceptualize processes more exactly to distinguish how far results approximate to an idealized state […] (Hyland, 1996: 440). Reliability hedges indicate "the writer's uncertain knowledge and indicate the confidence he or she is willing to invest in the validity of a claim […]" (Hyland, 1998a: 166).

For a more detailed description of Hyland's taxonomy, see Hyland 1996 or 1998a.

Figure 1. Hyland's taxonomy of hedges (Hyland, 1998a)
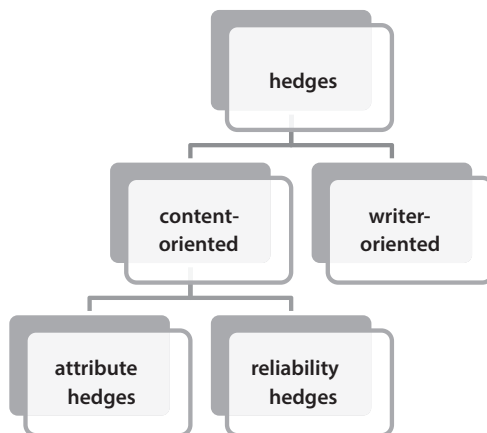
5. Some notes on Hyland's taxonomy

However, Hyland's classification is not unproblematic. This is caused, besides other things, by their very nature because hedges represent a very diverse and heterogeneous phenomenon as far as their surface form is concerned. Their pragmatic functions may overlap hence sometimes it is difficult to categorise them, and also, different approaches to hedging have been adopted. Moreover, a subjective element in classifying these expressions plays a role. Hyland is aware of this subjectivity in categorising hedges, therefore, he works with so-called "core examples" representing each category.

Problems arise when instances of hedges of different categories presented by Hyland as core examples are not clear. This may be illustrated by several examples taken from Hyland's work (1998a):

A) … **it appears** possible that the mechanism causing the light-activated fluorescence quenching **may** be triggered by either photosystem. (1998a: 167)

B) From this discussion, then, **it would appear that** some of the changes in the amino acid concentrations… (ibid. 173)

C) **It seems that** the stomata do not use the Calvin cycle… (ibid. 173)

Hyland regards the expressions in the first sentence as reliability hedges, but in the other two as writer-oriented. He explains that the primary motivation for hedging in the first case (example A above) is "a desire to clarify the state of knowledge, a hedge against complete accuracy, rather than a wish to seek protection against overstatement" (Hyland, 1998a: 167), whereas in examples B and C, the hedging implies that "the writer does not wish to be thought fully and personally committed to a belief in the proposed state of affairs" (ibid. 173). However, these realisations of hedges and the contexts in which they

occur seem to be almost identical. Although Hyland speaks of "higher-level claims" and "lower-level claims" when distinguishing these two categories, neither is this of any help since in practice these claims are rather difficult to determine.

Distinguishing between content- and reader-oriented hedges is in some cases also difficult. Hyland states that the explicit presence of the author in the text signals reader-oriented hedging, while the absence of it is regarded as content-oriented hedging. However, these distinctions are not so clear, as example D illustrates.

D) This insertion, which **we suspect** is the membrane anchor, could associate peripherally with the membrane or **might** span half the bilayer... (1998a: 167)

All highlighted expressions were judged as content-oriented reliability hedges, even though the first case is an explicit author reference signalled by the personal pronoun *we* and should therefore be classified as an instance of reader-oriented hedging.

The next difficulty is connected with authorial agentivity, as correctly pointed out by Varttala (2001: 88). It seems that any occurrence of a personal or possessive pronoun referring to the author (*I, we, my*, and *our*) is automatically regarded as an instance of reader-oriented hedging by Hyland, e.g. in *our findings, my data,* etc. These instances do not necessarily have to be cases of hedging but just ways to identify the authors of a given article in contrast with other scientific writers. These questionable cases were not included nor were they analysed in this study. When the above-mentioned pronouns collocated with clear instances of epistemic language means, only then were these expressions included.

In spite of these difficulties and occasional problems with its application, Hyland's taxonomy of hedging devices in medical discourse is a useful approach for analysing these diverse and multifunctional language means. As Varttala correctly states, Hyland's classification "is at its most valuable in summarising the major functions that hedges may have in the context of RAs. [...] Which of these functions an individual hedge can be seen to fulfil is a more complex matter" (2001: 90).

6. Materials and methods

This study is based on a comparative analysis of British and Czech research articles taken from prestigious medical periodicals with an impact factor released in 2014 and 2015. The British journals from which the articles were drawn are *The British Medical Journal, BMJ Open,* and *The Lancet.* The Czech medical journals used for this research were *Česká a Slovenská neurologie a neurochirugie* and *Epidemiologie, mikrobiologie, imunologie.* The British corpus totals 60,619 words, the Czech corpus contains 60,638 words. Abstracts, tables, graphs, notes, and references were excluded both from the word count and from the analysis itself. After both parallel corpora of medical articles were created, they were tagged manually for all hedging expressions present in the texts. Then these expressions were counted, categorised and mutually compared.

The theoretical framework for classification of hedging devices employed in this paper was a modified version of Hyland's taxonomy (1998a). In spite of the above-mentioned

Figure 2. Taxonomy of hedges used in this study
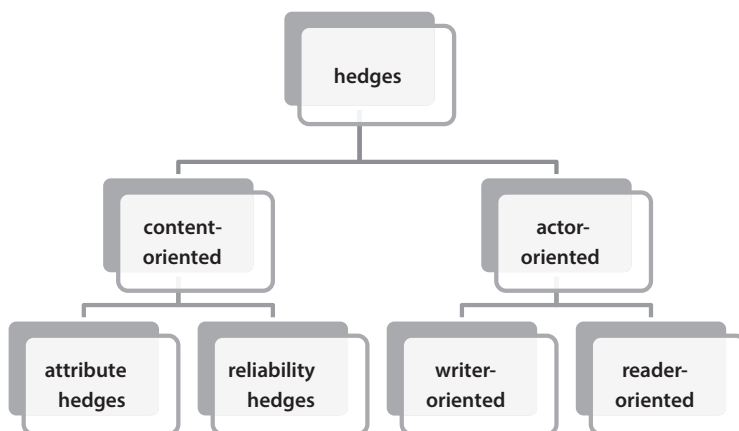
weaknesses, he suggested a well-applicable classification. However, some of the categories of hedging expressions had to be modified, as may be seen below.

I suggested two basic categories of hedging expressions according to their relation to the main components of the situation of discourse, which are the proposition and participants. Content-oriented hedges focus on the proposition itself and relate to the extent to which writers wish to modify its content as to its directness and (im)precision. Actor-oriented hedges aim at participants of the communicative situation, in this case at the author of a scientific text and at its recipient. These two subcategories of actor-oriented hedges also contribute to better understanding of hedging as a means of interaction.

As we can see, Hyland subsumes writer-oriented hedges under the category of content-oriented hedges. However, writer-oriented hedges do not mitigate the content of the proposition as such but rather reduce the presence of the writer in the text so they have a slightly different function and orientate more towards the actor. This is the reason for including them in a different category than Hyland originally suggested.

The results of both quantitative and qualitative analyses are discussed in the following sections.

7. Results and discussion

The total number of hedges and their types occurring in the two corpora of medical papers is summarised in Table 1.

As shown by the figures, the occurrence of hedges in the corpus of English medical articles is higher than in the Czech data. In every thousand words there appear almost 15% of hedging expressions in the English corpus while about 10% of hedges may be found in the Czech corpus. This finding supports the claims mentioned in the theoretical sections of this study which argue that the expression in Czech academic discourse is more direct than in English.

Table 1. Occurrence of hedges in English and Czech research articles (raw counts/frequency per 1,000 words).

Types of hedges		English corpus	Czech corpus
content-oriented	attribute hedges	111 / 1.83	87 / 1.43
	reliability hedges	532 / 8.78	372 / 6.13
actor-oriented	writer-oriented hedges	95 / 1.57	18 / 0.30
	reader-oriented hedges	61 / 1.01	43 / 0.71
other		102 / 1.68	72 / 1.19
TOTAL		901 / 14.86	592 / 9.76

When comparing the two basic categories of hedges, content- and actor-oriented, we can observe that in both corpora there appear more content-oriented hedges. When using them, the authors focus more on explanation or presentation of facts and mitigate the relationship between the content and a depiction of reality. The motivation for their use may be the author's focus on an accurate description of facts and research findings but also on self-protection as well as the prevention of opposition of a particular discourse community. Of content-oriented hedges, reliability hedges are those that are more frequent in both corpora. The element of the author's self-protection is apparent in Example 1 below. The author uses the epistemic modal verb *might* to weaken the force of his/her statement in order to prevent potential criticism and to present the proposition as an assumption rather than as a claim:

(1) The ALSPAC pregnancies occurred over 20 years ago, and, as then, there **might have been** changes in clinical practice. (EA1)

In Example 2, the highlighted reliability hedges show the writer's caution when suggesting possible improvements of future research in his/her field. Hedging enables the author to leave some space for potential discussion with other researchers.

(2) In conclusion, measurements of blood pressure recorded during the second half of pregnancy, […], **can improve** the identification of women who are at risk of developing pre-eclampsia later in pregnancy and **could be used to differentiate** women who require more intensive monitoring from those who are likely to have a normal pregnancy. (EA1)

Regarding the surface forms of reliability hedges, in the majority of cases they are expressed by the modal auxiliaries *can, could, may, might*, and *would*, then by probability adjectives and adverbs such as *possible, probable, likely, possibly, probably, perhaps, potentially, apparently*, by tentative cognition nouns (*assumption, estimate*), nouns of tentative likelihood (*probability, possibility, likelihood*), and verbs expressing tentative cognition (*estimate, assume*).

(3) The risks of channeling bias are **probably** small considering similar pregnant populations […]. (EA4)

(4) There is therefore **the possibility** that these measurements are not a reasonable representation […]. (EA9)

In (3) and (4) above, the reliability hedges used express the author's opinion tentatively, leaving room for other possibilities.

As already said, content-oriented hedges, and particularly reliability hedges, are the most frequent hedging expressions also in the Czech corpus. They have the same functions as those in the English data, i.e. focusing on an accurate description of findings, expressing writer's tentativeness, opening room for other possible interpretations of the research findings and leaving space for scientific discussion (Examples 5 and 6 below). Also, the surface forms of Czech hedges are very similar to those appearing in English, although their variety is not so wide. Czech reliability hedges are typically expressed by the modal verb *moci* [*be able*], the probability adjectives *možný* [*possible*], *pravděpodobný* [*probable*], the probability adverbs *pravděpodobně* [*probably*]*, lze* [*is-possible*], and the nouns of tentative likelihood *možnost* [*possibility*], *pravděpodobnost* [*probability*].

(5) Pozitivní vliv **mohla mít** [could have] i samotná hospitalizace pacientů a s ní spojený režim na oddělení a podpůrná psychoterapie […]. (CA6)

(6) K limitujícím faktorům určitě náleží poměrně krátké trvání studie – a to tři týdny, což **může být** [can be] příliš krátké na plný rozvoj terapeutického účinku antidepresiv, ale i rTMS. (CA6)

In Examples 7 and 8, there occurs a compound reliability hedge *lze pravděpodobně* [*is-possible probably*], which occurs quite frequently in the Czech corpus. It expresses a higher degree of tentativeness and again, opens space for further discussion on the topic.

(7) Rozdíl mezi muži a ženami **lze pravděpodobně** přisoudit výchovnému stylu, který v našich kulturních podmínkách preferuje inhibici emocí u mužů. (CA3)

(8) O něco horší výsledek než v našem případě […] **lze pravděpodobně** přikládat tomu, že do Barešovy studie byli zařazeni výhradně nemocní na léčbu rezistentní. (CA6)

Compound reliability hedges appear in the English corpus too:

(9) Secondly, people with hypertension and peripheral arterial disease **may be more likely** to be screened for cardiovascular disease than people without those disorders. (EA2)

The other subcategory of content-oriented hedges, namely attribute hedges, represents the second most frequent category of hedging expressions in both corpora. Attribute hedges weaken the illocutionary force of the arguments. They are used when the writers approximate their research results to an expected or a usual state of knowledge

and attempt to find precision in expression, as illustrated below (Examples 10–13). For this reason, the authors employ approximative adverbs (*approximately, almost, nearly*) and adjectives and adverbs of indefinite degree (*modest, slight, quite, somewhat, slightly*). When using any of these means, the degree of strength of the given expression is modified. In Czech very similar language means are used as attribute hedges, e.g. *přibližně* [*approximately*]*, zhruba* [*roughly*]*, asi* [*about*] (approximative adverbs), and *relativně* [*relatively*]*, poměrně* [*relatively*] (adverbs of indefinite degree). Adjectives of indefinite degree were not found in the Czech corpus.

(10) **Approximately** 1% of the population have intellectual disability, defined as a significant deficit in cognitive and adaptive function with onset during the developmental period. (EA 11)

(11) In women who reported that they were in fair or poor health, being unhappy was associated with a **slightly** lower mortality than being happy most of the time […]. (EA12)

(12) Mortalita těchto pacientů je **přibližně** 50%. (CA7)

(13) Tumory thalamu jsou **relativně** vzácné léze a představují **asi** 5 % intrakraniálních nádorů. (CA5)

The category of actor-oriented hedges consists of two subtypes directing either at the writer or reader. As shown in Table 1, they are more recurrent in the English corpus, which does not necessarily mean that the Czech medical articles do not take discourse participants into account. The reason for rarer occurrence of actor-oriented hedges in the Czech corpus may be a greater orientation towards the explication and effort of the authors to explain things as thoroughly and matter-of-factly as possible. As regards the particular subgroups of actor-oriented hedges, writer-oriented, which diminish the presence of the writer in the text, are more frequent in the English articles. The writers weaken the strength of the propositional content and make their claims more indirect. This may be a face-saving strategy and also prevention of opposition from the scientific community (Examples 14–17 below).

(14) **It has been suggested** that related subjective measures of wellbeing […] could independently affect mortality. (EA12)

(15) […] and heavy alcohol consumption – a risk factor for all types of stroke – might be a contributing factor because employees working long hours **seem to be** slightly more prone to risky drinking than are those who work standard hours. (EA10)

(16) Tento časový interval **se zdá být** [seems to be] z pohledu hodnocení dynamiky TCD PbtO2 dostatečný […]. (CA4)

(17)   Ve většině případů nelze příčinu SKT prokazatelně identifikovat a onemocnění je **považováno** [is considered] za idiopatické. (CA1)

As seen from these illustrative examples, writer-oriented hedges may be realised by nonfactive reporting verbs (*suggest*), tentative cognition verbs (*consider, think, estimate; považovat* [*consider*], *předpokládat* [*suppose*]), and by tentative linking verbs (*seem, appear; zdát se* [*seem*]).

The incidence of reader-oriented hedges is similar in both corpora, with the frequency of 1.01 per 1,000 words in the English corpus and the frequency of 0.71 in the Czech corpus. Within the Czech corpus, reader-oriented hedges are more frequent than writer-oriented. When using reader-oriented hedges, the authors show respect for the audience, aim at involving the readers in the argumentation process and present their views so that the readers feel that they may form their own judgements (Examples 18–21).

The surface forms of reader-oriented hedges are quite varied. Scientific writers aim at avoiding conflict and do not want to threaten the readers' negative face so they choose non-imposing phrases or expressions. Therefore, they use means of reader involvement and means of attenuating their claims. To fulfil these functions, both Czech and English writers use singular or plural first person pronouns, expressions of personal belief, author's self-reference, personal attribution, etc. As already mentioned above, Hyland's taxonomy is questionable at some points, for instance, in determining reader-oriented hedges. In this study, only a clear author reference collocating with an epistemic expression was treated as an illustration of a reader-oriented hedge, as in Examples 18–21.

(18)   Furthermore, we do not have information on whether patients stop taking anti-thrombotics when treated with NSAIDs; however, given post-myocardial infarction treatment guidelines, **we think** that this is unlikely. (EA7)

(19)   **Our findings suggest** that tailoring of information delivery to the communities being served might be useful. (EA8)

(20)   **Autoři** [této studie] **se domnívají** [the authors suppose], že problémem byla porušená žilní drenáž v. thalamostriata sin., která byla v těsném kontaktu s cévnatým okrajem tumoru. (CA5)

(21)   Prezentované normy pro všechny zkoušky VF mohou, **podle našeho názoru** [in our opinion], významně přispět k hodnocení kognitivní výkonnosti v klinické praxi. (CA22)

Hedging devices are very difficult to categorise because they constitute a very diverse group of language means, and one and the same attenuating expression may fulfil several different functions depending on context. It is then rather problematic to suggest a clear-cut taxonomy. When analysing both parallel corpora of medical research articles, hedging expressions occurred which did not fit any of the above-defined categories but are evidently instances of hedging. These are usually quantifying expressions signalling indeterminacy and vagueness, such as *several, some, at least, most (of), majority / některý*

[*some*], *řada (z)* [*most (of)*], *většina* [*majority*], and the attributive adjectives *slight* or *recent,* as we may note in the examples below. These expressions are used when there are no exact numbers or data available or the author does not consider it necessary to quote the precise figures. In this study they fall within the category "other".

(22) This analysis also has ***several*** limitations. (EA2)

(23) Whatever the cause, the interaction is potentially clinically important because folic acid supplementation might be more likely in **some** patient groups taking lamotrigine […]. (EA6).

(24) **Několik málo** [several] předchozích studií referuje záchyt FiS u cca 3–4 % mladých pacientů do 50 let v době přijetí pro akutní iCMP. (CA2)

(25) **Řada z nich** [most of them] se uplatnila i v podrobnější stratifikaci anaplastických gliomů. (CA23)

8. Conclusion

As we have seen, the phenomenon of hedging in medical research papers is quite prevalent although it has been frequently claimed that scientific language should be precise and matter-of-fact. A cross-linguistic perspective was taken in this study to compare a corpus of English and Czech medical articles published in peer-reviewed medical journals with an impact factor.

The research revealed that of these two languages, hedging occurs more frequently in English. This result contributes to the discussion on universality or culture-specificity of language means used in academic discourse and supports the view that the use of hedging expressions is culturally determined. Czech medical discourse is more straightforward and direct compared to English medical discourse. This is connected with the fact that Czech authors focus more on the content they convey and present their findings and thoughts matter-of-factly in non-modalised statements. However, this does not mean that Czech scientific writers do not take the reader into account. There are many instances of modalised utterances in the Czech corpus which clearly turn to the reader and present the claims as opinions open to discussion rather than as definitive facts.

Focusing now on the distribution of the specific types of hedging devices, reliability hedges were the most frequent type in both corpora. These hedges indicate that a statement is tentative and not definitive and opens space for dialogue. Without them, assertions would be rather categorical and face-threatening. Reliability hedges are followed by attribute hedges as regards their frequency in both corpora. This category of hedges suggests that the research results are approximate and the authors try to find precision in expression and to evaluate the accuracy of their arguments. As regards the two types of actor-oriented hedges, writer-oriented and reader-oriented, both are more frequent in the English data. This confirms the finding made above that Czech scientific authors concentrate more on the content of their texts and conveying facts. Writer-oriented

hedges reduce the voice of the author in the text, thereby diminishing writer involvement with the textual claims. Reader-oriented hedges contribute positively to the relationship between the writer and the reader. The writers show that the role of the reader is active in the ratification of their assertions and involve him/her in the argumentation process.

What is also important to take into account when examining hedging is the structure of research articles. Scientific papers published in the most prestigious journals usually have the IMRAD format. It would be interesting to analyse the incidence of hedging expressions across the different sections of research papers because they are not distributed evenly within an article. Unfortunately, focusing on this was beyond the scope of the present study but it will be the subject of another analysis of hedging.

To conclude, hedges play a significant role in academic discourse since they enable authors to present their arguments with appropriate accuracy and modesty rather than regard the conclusions as invariant, hence their unproven claims are explained with caution. Hedging is a positive and necessary phenomenon because it makes author's assertions more accessible for discussion and develops a writer-reader relationship. This contributes to the fact that hedging should be understood as a means of interaction. Although hedges are polypragmatic, which means that one and the same attenuating device may fulfil different functions in different contexts, it must be noted that they are "a resource, not a problem" (Skelton, 1985: 41) and that this issue should be addressed, for instance, in teaching academic writing.

REFERENCES

Atai, M. and L. Sadr (2008) Cross cultural genre study on hedging devices in discussion section of Applied Linguistics research articles. *Journal of Teaching English Language and Literature Society of Iran* 2 (7), 1–22.

Clyne, M. (1991), The sociocultural dimension: The dilemma of the German speaking scholar. In: Schroder, H. (ed.), *Subject-oriented Texts: Languages for Special Purposes and Text Theory*, 49–68. Berlin: Mouton de Gruyter.

Čmejrková, S., F. Daneš and J. Světlá (1999) *Jak napsat odborný text*. Prague: Leda.

Čmejrková, S. and F. Daneš (1997) Academic writing and cultural identity: The case of Czech academic writing. In: Duszak, A. (ed.) *Culture and Styles of Academic Discourse*, 41–62. Berlin: Mouton de Gruyter.

Daneš, F. (2000) Jakou řečí mluví věda. Modalizace vědeckého diskurzu. *Slovo a slovesnost* 61 (2), 81–92.

Dontcheva-Navrátilová, O. (2013) Lexical bundles indicating authorial presence: A cross-cultural analysis of novice Czech and German writers' academic discourse. *Discourse and Interaction* 6 (1), 7–24.

Figueiredo-Silva, M. I. R. de (2001) Teaching academic reading: Some initial findings from a session on hedging. *Proceedings of the Postgraduate Conference 2001,* Department of Theoretical and Applied Linguistics, The University of Edinburgh.

Hyland, K. (1996) Writing without conviction? Hedging in science research articles. *Applied Linguistics* 17 (4), 433–454.

Hyland, K. (1998a) *Hedging in Scientific Research Articles*. Amsterdam, Philadelphia: John Benjamins.

Hyland, K. (1998b) Boosting, hedging and the negotiation of academic knowledge. *Text* 18 (3), 349–382.

Kreutz, H. and A. Harres (1997) Some observations on the distribution and function of hedging in German and English academic writing. In: Duszak, A. (ed.), *Culture and Styles in Academic Discourse*, 181–202. Berlin: Mouton de Gruyter.

Martín-Martín, P. (2008) The mitigation of scientific claims in research papers: A comparative study. *International Journal of English Studies* 8 (2), 133–152.

Mathesius, V. (1982) Řeč a sloh. In: Vachek, J. (ed.), *Jazyk, kultura a slovesnost*, 92–146. Praha: Odeon.

Oliver del Olmo, S. (2004) *Análisis Contrastivo Español/Inglés de la Atenuación Retórica en el Discurso Médico: el Artículo de Investigación y el Caso Clínico*. Unpublished doctoral dissertation. University of Pompeu Fabra, Spain.

Oliver del Olmo, S. (2005) The role of passive voice in hedging medical discourse: A corpus-based study on English and Spanish research articles. *LFE: Revista de lenguas para fines específicos* 11–12, 205–218.

Skelton, J. (1985) The care and maintenance of hedges. *ELT Journal* 42 (1), 37–43.

Varttala, T. (2001) *Hedging in Scientific Oriented Discourse: Exploring Variation According to Discipline and Intended Audience*. Unpublished doctoral dissertation, University of Tampereen Yliopisto, Finland.

Vassileva, I. (1997) Hedging in English and Bulgarian academic writing. In: Duszak, A. (ed.) *Culture and Styles in Academic Discourse*, 141–157. Berlin: Mouton de Gruyter.

Weinrich, H. (1995) Wissenschaftssprache, Sprachkultur und die Einheit der Wissenschaft. In: Kretzenbacher, H. L. and H. Weinrich (eds.), *Linguistik der Wissenschaftssprache*, 155–172. Berlin/New York: W. de Gruyter.

Yang, Y. (2003) *A Contrastive Study of Hedges in English and Chinese Academic Discourse*. Unpublished MA thesis. Jilin University, Changchun, China.

Sources

English corpus

Cleynen, I. et al. (2015) Inherited determinants of Crohn's disease and ulcerative colitis phenotypes: a genetic association study. *The Lancet*. Available at: http://www.thelancet.com/pdfs/journals/lancet/PIIS0140-6736(15)00465-1.pdf. (EA14)

Emdin, C. A. et al. (2015) Usual blood pressure, peripheral arterial disease, and vascular risk: cohort study of 4.2 million adults. *BMJ*. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4586462/. (EA2)

Furu, K. et al. (2015) Selective serotonin reuptake inhibitors and venlafaxine in early pregnancy and risk of birth defects: population based cohort study and sibling design. *BMJ*. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4410618/. (EA4)

Geddes, J. R. et al. (2015) Comparative evaluation of quetiapine plus lamotrigine combination versus quetiapine monotherapy (and folic acid versus placebo) in bipolar depression (CEQUEL): a 2 × 2 factorial randomised trial. *The Lancet*. Available at: http://www.thelancet.com/pdfs/journals/lanpsy/PIIS2215-0366(15)00450-2.pdf. (EA6)

Jagger, C. et al. (2015) A comparison of health expectancies over two decades in England: results of the Cognitive Function and Ageing Study I and II. *The Lancet*. Available at: http://www.thelancet.com/pdfs/journals/lancet/PIIS0140-6736(15)00947-2.pdf. (EA3)

Kivimäki, M. et al. (2015) Long working hours and risk of coronary heart disease and stroke: a systematic review and meta-analysis of published and unpublished data for 603 838 individuals. *The Lancet*. Available at: http://www.thelancet.com/pdfs/journals/lancet/PIIS0140-6736(15)60295-1.pdf. (EA10)

Knipe, D. W., A. Fraser, D. A. Lawlor and L. D. Howe (2014) Is interpregnancy interval associated with cardiovascular risk factors in later life? A cohort study. *BMJ Open*. Available at: http://bmjopen.bmj.com/content/4/3/e004173.full.pdf+html. (EA9)

Liu, B. et al. (2015) Does happiness itself directly affect mortality? The prospective UK Million Women Study. *The Lancet*. Available at: http://www.thelancet.com/pdfs/journals/lancet/PIIS0140-6736(15)01087-9.pdf. (EA12)

Macdonald-Wallis, C. et al. (2015) Antenatal blood pressure for prediction of pre-eclampsia, preterm birth, and small for gestational age babies: development and validation in two general population cohorts. *BMJ*. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4647185/#. (EA1)

Schjerning Olsen, A. M. et al. (2015) Impact of proton pump inhibitor treatment on gastrointestinal bleeding associated with non-steroidal anti-inflammatory drug use among post-myocardial infarction patients taking antithrombotics: nationwide study. *BMJ*. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4609736/. (EA7)

Shah, A. et al. (2015) High-sensitivity cardiac troponin I at presentation in patients with suspected acute coronary syndrome: a cohort study. *The Lancet.* Available at: http://www.thelancet.com/pdfs/journals/lancet/PIIS0140-6736(15)00391-8.pdf. (EA15)

Sheehan, R. et al. (2015) Mental illness, challenging behavior, and psychotropic drug prescribing in people with intellectual disability: UK population based cohort study. *BMJ.* Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4556752/. (EA11)

Stephan, B. et al. (2015) Usefulness of data from magnetic resonance imaging to improve prediction of dementia: population based cohort study. *BMJ.* Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4476487/. (EA13)

Thorlund, J. B. et al. (2015) Arthroscopic surgery for degenerative knee: systematic review and meta-analysis of benefits and harms. *BMJ.* Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4469973/. (EA5)

Wardle, J. et al. (2015) Effects of evidence-based strategies to reduce the socioeconomic gradient of uptake in the English NHS Bowel Cancer Screening Programme (ASCEND): four cluster-randomised controlled trials. *The Lancet.* Available at: http://www.thelancet.com/pdfs/journals/lancet/PIIS0140-6736(15)01154-X.pdf. (EA8)

Czech corpus

Beránková, D. et al. (2015) Addenbrookský kognitivní test – orientační normy pro českou populaci. *Cesk Slov Neurol N* 78/111(3), 300–305. (CA17)

Bezdíček, O. et al. (2014) Validita Montrealského kognitivního testu pro detekci mírné kognitivní poruchy u Parkinsonovy nemoci. *Cesk Slov Neurol N* 77/110(1), 47–53. (CA11)

Ďuriš, K. et al. (2014) Vztah mezi parametry transkraniální dopplerometrie a tkáňovou oxymetrií u pacientů s těžkým subarachnoidálním krvácením. *Cesk Slov Neurol N* 77/110(2), 196–201. (CA4)

Fibír, A. et al. (2015) Mobilita medianu před dekompresí karpálního tunelu a po ní. *Česká a Slovenská neurologie a neurochirugie* 78/111(6), 675–679. (CA1)

Flanderková, E. et al. (2014) Posuzování gramatičnosti v Brocově afázii – příklad dvou pacientů. *Cesk Slov Neurol N* 77/110(2), 202–209. (CA12)

Hrabálek, L. et al. (2015) Algoritmus operací mnohočetného myelomu a solitárního plazmocytomu páteře. *Cesk Slov Neurol N* 78/111(1), 64–71. (CA20)

Kalita, O. et al. (2015) Současný stav léčby anaplastických gliomů v České republice. *Cesk Slov Neurol N* 78/111(3), 306–316. (CA23)

Káňová, M. et al. (2015) Delirium u kriticky nemocných – prospektivní studie. *Cesk Slov Neurol N* 78/111(6), 662–667. (CA21)

Košťálová, M. et al. (2015) Dotazník funkcionální komunikace (DFK) – validace originálního českého testu. *Cesk Slov Neurol N* 78/111(2), 188–195. (CA18)

Krůtová, M. et al. (2014) Diagnostika Clostridium difficile infekcí – porovnávací studie dvou imunoenzymatických metod s konfirmací pomocí PCR a kultivace s následnou ribotypizací kmene. *Epidemiologie, mikrobiologie, imunologie* 63(2), 99–102. (CA10)

Kříž, J. and V. Hyšperská (2014) Vývoj neurologického a funkčního obrazu po poranění míchy. *Cesk Slov Neurol N* 77/110(2), 186–195. (CA9)

Laczó, J. et al. (2014) Rozdíly v postižení prostorové orientace u demencí neurodegenerativní etiologie. *Cesk Slov Neurol N* 77/110(4), 449–455. (CA15)

Mrlian, A. et al. (2014) Možnosti kontinuálního monitoringu průtoku krve mozkem v detekci vazospazmů u pacientů po těžkém subarachnoidálním krvácení. *Cesk Slov Neurol N* 77/110(3), 326–333. (CA7)

Nikolai, T. et al. (2015) Testy verbální fluence, česká normativní studie pro osoby vyššího věku. *Cesk Slov Neurol N* 78/111(3), 292–299. (CA22)

Rektorová, I. et al. (2014) Hodnocení písma u pacientů s Parkinsonovou nemocí. *Cesk Slov Neurol N* 77/110(4), 456–462. (CA8)

Smrčka, M. et al. (2015) Chirurgický přístup k tumorům thalamu. *Cesk Slov Neurol N* 78/111(2), 172–180. (CA5)

Světlák, M. et al. (2015) Měření úrovně emočního uvědomění – pilotní studie ověřující základní psychometrické vlastnosti české verze Škály úrovně emočního uvědomění LEAS. *Cesk Slov Neurol N* 78/111(6), 680–688. (CA3)

Šaňák, D. et al. (2015) Poruchy srdečního rytmu u mladých pacientů s kryptogenní ischemickou cévní mozkovou příhodou. *Cesk Slov Neurol N* 78/111(6), 669–674. (CA2)

Šonka, K. et al. (2014) Přínos opakování nepotvrzujícího testu mnohočetné latence usnutí (MSLT) pro stanovení diagnózy narkolepsie. *Cesk Slov Neurol N* 77/110(4), 444–448. (CA16)

Štěpánková, H. et al. (2015) Mini-Mental State Examination – česká normativní studie. *Cesk Slov Neurol N* 78/111(1), 57–63. (CA13)

Timkovič, J. et al. (2015) Možnosti ovlivnění diplopie při paralytickém strabismu konzervativní léčbou. *Cesk Slov Neurol N* 78/111(2), 181–187. (CA19)

Ustohal, L. et al. (2014) Repetitivní transkraniální magnetická stimulace v léčbě depresivní poruchy – randomizovaná, jednoduše slepá, antidepresivy kontrolovaná studie. *Cesk Slov Neurol N* 77/110(5), 602–607. (CA6)

Žemličková, H. et al. (2014) Rezistence k erytromycinu, ciprofloxacinu a tetracyklinu u humánních izolátů *CAMPYLOBACTER* spp. V ČESKÉ republice, vyšetřená standardní metodou EUCAST. *Epidemiologie, mikrobiologie, imunologie* 63(3), 184–190. (CA14)

"…AND OUR STUDY MIGHT THEREFORE HAVE BEEN SLIGHTLY UNDERPOWERED": KOMPARATIVNÍ STUDIE HEDGINGU V ANGLICKÝCH A ČESKÝCH LÉKAŘSKÝCH ODBORNÝCH ČLÁNCÍCH

Resumé

Jazykové prostředky zeslabující ilokuční sílu a přímost výpovědi se často používají jak v mluveném, tak v psaném jazyce a bývají spojovány s vyjadřováním zdvořilosti. Ve vědeckém diskurzu se tyto prostředky většinou používají k vyjádření negativní zdvořilosti, protože jedním z cílů autorů vědeckých článků je ochrana před nesouhlasnými nebo kritickými projevy ze strany recipientů. Z lingvistického hlediska mohou být zeslabující jazykové prostředky (hedges) vyjádřeny různými způsoby, např. modálními adverbii (*possibly, perhaps, probably; možná, pravděpodobně*…), modálními adjektivy (*possible, (un)likely, probable; možný, (ne)pravděpodobný*…), modálními substantivy (*assumption, possibility, suggestion; možnost, domněnka*…), modálními slovesy vyjadřujícími možnost (*might, could, would; moci*…), slovesy epistémickými (*assume, seem, appear, suggest; zdát se, domnívat se*…), výrazy vyjadřujícími přibližnost jako např. *approximately* a *roughly* (*přibližně, zhruba*), apod. Cílem této komparativní studie je analýza jazykových prostředků zeslabujících ilokuční sílu výpovědi v psaném akademickém diskurzu, konkrétně v anglických a českých lékařských článcích, které byly publikovány v odborných lékařských časopisech s impakt faktorem, porovnat je z hlediska jejich typů, výskytu a komunikativních funkcí. Jinými slovy, cílem je zjistit, zda je použití těchto výrazů v lékařském diskurzu kulturně specifické nebo není. Zeslabující výrazy mohou být klasifikovány z různých hledisek, v této studii byla využita modifikovaná Hylandova taxonomie (1998).

Jana Kozubíková Šandová
Department of English, Faculty of Philosophy
University of South Bohemia in České Budějovice
sandova@ff.jcu.cz

VERB ERRORS IN ADVANCED SPOKEN ENGLISH

TOMÁŠ GRÁF

ABSTRACT

As an experienced teacher of advanced learners of English I am deeply aware of recurrent problems which these learners experience as regards grammatical accuracy. In this paper, I focus on researching inaccuracies in the use of verbal categories. I draw the data from a spoken learner corpus LINDSEI_CZ and analyze the performance of 50 advanced (C1–C2) learners of English whose mother tongue is Czech. The main method used is Computer-aided Error Analysis within the larger framework of Learner Corpus Research. The results reveal that the key area of difficulty is the use of tenses and tense agreements, and especially the use of the present perfect. Other error-prone aspects are also described. The study also identifies a number of triggers which may lie at the root of the problems. The identification of these triggers reveals deficiencies in the teaching of grammar, mainly too much focus on decontextualized practice, use of potentially confusing rules, and the lack of attempt to deal with broader notions such as continuity and perfectiveness. Whilst the study is useful for the teachers of advanced learners, its pedagogical implications stretch to lower levels of proficiency as well.

Keywords: advanced spoken English, learner corpora, error analysis, tenses, verbs, present perfect

1. Introduction

The Common European Framework of Reference (Council of Europe, 2001, pp. 24–27) defines the advanced levels (i.e. C1 and C2) of language proficiency in its Global Scale and self-assessment grids using such terms as coherence, spontaneity, fluency, precision, complexity, flexibility, efficiency, and effortlessness. It further describes advanced accuracy as being characterized by a high degree of grammatical accuracy, consistent grammatical control and rare occurrence of errors. It is the experience of many a language teacher that whilst the speech of truly advanced learners displays all of the above mentioned characteristics and is certainly not riddled with errors, occasional instances of inaccuracies still occur, and sometimes it might even surprise the hearer how seemingly basic these

appear to be. As has been shown, for example by Granger (1999), Götz (2015), and Gráf (2015a), such errors often involve the use of tenses. It is the aim of this study to explore the nature of grammatical errors in the use of verbs in spontaneous spoken production of Czech advanced learners of English. At the time of writing no recent analysis of such material is available.

1.1 Error analysis and its offshoots

It might seem somewhat anachronistic to revert our attention to errors decades after the gradual decline of Error Analysis and at a time when the increased use of English as a lingua franca and the communicative language teaching methodologies call for increased tolerance to less than perfect L2 performance. The recent advances in the field of Learner Corpus Research (see especially Granger et al., 2015) and the development of Contrastive Interlanguage Analysis (Granger, 1996) and Computer-aided Error Analysis (Dagneaux et al., 1998) have made available not only new sources of data but also new analytical and interpretative techniques. Together these can yield deeper insight into the nature of L2 production and acquisition and can inform the fields of language acquisition and pedagogy. Last but not least new research in this area within the framework of English as a foreign language (EFL) may show to what extent the original findings are applicable to the English spoken by a generation of young learners (Meriläinen, 2010) who have had entirely different opportunities throughout the learning process than the preceding generations, e.g. early starting age at schools, opportunities to practise and travel, easy access to ELT and authentic materials, significant exposure stemming from the use of English as international language on the internet and elsewhere, and this list could easily continue. Aided by the status of English and its omnipresence, large numbers of learners nowadays frequently attain much higher levels of foreign language competence than is customary in other foreign languages than English. This gives researchers the opportunity to study advanced language proficiency and answer questions regarding fossilization and maximum L2 attainment. It is in this area that the study of errors is especially useful even though many of the methodological problems identified in the early days of contrastive and error analysis – such as the very definition of error – remain the same.

The systematic study of language-learner errors became the preoccupation of applied linguists in the late 1950s when Lado (1957) laid the foundations for the so-called Contrastive Analysis Hypothesis which claimed that the main difficulties in acquiring a foreign language were caused by differences between the learner's L1 and the target language. Such difficulties were considered predictable if a methodical contrastive analysis of the two languages in question was carried out, and consequently such pedagogical procedures could be designed as would target the problematic aspects in order to prevent language errors. The clearly aberrant belief that language teaching was primarily a question of error prevention and the gradual realization that learners did not behave exactly in the way contrastive analysts had predicted inevitably led to the decline of the methodology. The interest in errors, however, persisted. A new methodology, Error Analysis (EA) (Corder, 1967, 1974), studied errors not as obstacles in language learning but as evidence of learners' internal hypotheses regarding the target language system. Such

information would prove invaluable for the understanding of L2 acquisition and production. EA provided typologies of errors and linked them to the different L2 developmental stages whilst moving away from the questionable attempts to identify sources of errors. One of EA's most important contributions was the differentiation between systematic and performance errors, which proved useful to both language teachers and learners.

However, there were several methodological problems. The key one was the inability to provide a fool-proof definition of error. Definitions mostly juxtaposed errors with norms by claiming that errors were deviations from the norm, or even more loosely that errors are "unsuccessful bits of language" (James, 1998, p. 1). The problem instantly arose as to which norm to work with and in which different contexts. What might be incorrect in one particular situation could be entirely "normal" in another, and the whole picture might receive different contours in different varieties of the target language, a realization which led Dušková (1969) to point out that there was a continuum between deviation and acceptability, or, as Gilquin and De Cock (2011) observed, between error and dysfluency.

The waning interest in errors was rekindled in the 1990s with the development of computerised learner corpora, which resulted in many pedagogical applications such as usage notes and error warnings in learner dictionaries (e.g. Rundell, 2007), a dictionary of errors (Turton and Heaton, 1996) and Swan's and Smith's (2001) treatise of learner English around the world. The study of errors also prepares the ground for the development of accuracy studies within the CAF (complexity, accuracy and fluency) model of language proficiency and production (Housen et al., 2012).

1.2 Verbal categories in learner English

Verbal categories include number, tense and aspect, mood, and voice. Of these it is tense and aspect that are of particular interest to language teaching. Both tense and aspect are verbal categories of semantic nature. Tense is a type of deixis whose function is to provide temporal reference, or – in other words – relate the actual time of the event, action or state described by the verb to some other time, most commonly the time of speaking. Aspect expresses the stance the speaker takes to the progress of the described event and encodes among other such notions as completeness, continuity, sequentiality, intentionality, and iterativeness.

The use of tenses and aspect is a notoriously difficult area for the learners of foreign languages. English, despite its proverbial simplicity, is not an exception. This is especially true for learners whose own mother tongue's tense and aspect system is as different as in the case of Czech learners learning English. For the ease of teaching, pedagogical grammars of English traditionally subsume the category of aspect under the broader category of tense and differentiate between present, past, future, present perfect, past perfect and future perfect, and within each of these further distinguish between simple and continuous/progressive forms. Thus English textbooks and pedagogical grammars recognize 12 different tenses. Once they are labelled in this way, they are usually introduced separately and then comparisons are made between simple and continuous forms of each tense. This practice does not encourage the development of the awareness of aspect; teachers teach present simple, then present continuous, but they do not deal with the concept of continuity on its own. Similarly, they do not deal with the concept of

perfectiveness but teach the perfect tenses as separate entities not related to each other. Moreover, they frequently take sentences out of context and rather than develop in their students the feel for the broader notions and the roles they play in the construction of discourse, they "equip" the learners with often confusing rules such as "in sentences with *since* always use the present perfect". Consequently, many learners come to believe that the choice of a grammatical aspect depends on the presence of an overt feature (such as a particular adverb).

The basis for the distinction between the tenses derives from morphological markers but it does not deal with broader notions. Thus the present perfect is identified as a form consisting of an auxiliary *have* and the past participle of the lexical verb, and then rules are given for the different uses of the tense. The very system of the twelve tenses thus inherently encourages deductive and present-practice-produce approaches to the teaching of grammar at the cost of induction and discovery which promote the development of linguistic awareness and from a pedagogical point of view appear more logical.

The teaching of tense and aspect in English to beginners usually commences with the present simple forms of the verbs *be* and *have*. These are usually acquired rapidly with only occasional mistakes in using the form *have* for *has* in the 3rd person singular. Lexical verbs are gradually introduced to be used in the present simple. The acquisition of the third person *–s* ending is, however, slow and problematic despite its conceptual simplicity, and slips when the student fails to produce the form may occur even beyond the intermediate level. Interestingly, its acquisition has been shown to be slow even for children acquiring English as L1 when the plural *-(e)s* ending and the genitive *'s* tend to be acquired first (Clark, 2009; Ingram, 1989). This is probably caused by the relative infrequency of the ending (compared to the higher frequency of the plural marker, and the frequency with which we refer to personal possessions when communicating with little children, e.g. *Mummy's bag, Daddy's car* etc.) and by its superfluity from the communicative point of view.

The present continuous is introduced soon and it has been pointed out (Conrad, 2016) that its teaching and practice takes precedence over the present simple even though it is the latter that appears in English with greater frequency. The present continuous is generally taught as a structure with concrete rules for use whilst the concept of continuity and the reasons for making it explicit are usually ignored. This may cause problems for speakers of languages (such as Czech and German) which do not make this distinction.

As much as continuity as a broader concept is ignored, so is perfectiveness. Explanations tend to focus on individual types of usage and many "helpful" rules are introduced when the students are advised to choose the tense if the sentence contains a particular, overt feature. This approach is then further strengthened by decontextualized practice which makes use of these overt features, rather than by contextualized, discourse-based approaches.

It is of no surprise that studies of learner language find the areas of tense and aspect the most error-prone (Davydova, 2011; Dietrich, Klein, and Noyau, 1995; Eriksson, 2004, 2008; Götz, 2015; Gráf, 2015a; Granger, 1999; Hinkel, 2004; Rogatcheva, 2009, 2012; Salaberry and Shirai, 2002).

Other morphological categories of verbs – number, mood, and voice – with regard to spoken learner language and related errors have been studied less extensively. Numer-

ous studies have dealt with the use of active and passive voice in writing (e.g. Granger, 2013). Likewise, modality has drawn some attention (e.g. Guo, 2005) and learners' use of the conditional has been explored for example by Götz (2015). Whilst the current study explores errors in these categories as well, its focus is on tense and aspect; as is about to be shown, they vastly outnumber the other instances of grammatical inaccuracies.

2. Material and method

The material for our study draws on the Czech subcorpus of LINDSEI[1] (Gilquin et al., 2010). It will be further referred to as LINDSEI_CZ (Gráf, 2015b). LINDSEI is a multi-national corpus of advanced spoken learner English. At the time of writing it comprises 20 national subcorpora, each with at least 50 speakers' recordings with average duration of 15 minutes and corresponding orthographic transcriptions. LINDSEI was conceived as the spoken counterpart of the written corpus ICLE (Granger, 2009).[2] The speakers perform three tasks (a monologue on a preselected topic, a dialogue and a picture-based narrative). The speakers' proficiency had not been established prior to their recruitment. Instead, a method of institutional selection was adopted and LINDSEI recruited speakers from among students of English philology in at least the third year of their university course. It has been shown (Götz, 2015; Gráf, 2015) that this is a weak point in the corpus design as proficiency within the corpus can vary (Carlsen, 2012). Some of the national subcorpora (French, Taiwanese, Czech) have consequently engaged professional proficiency raters, but to date only the French subcorpus has been fully assessed, the work on the other two being in progress.

For this study spontaneous speech production has been chosen as it is here, under the pressure of online planning, on the assumption that the learners are likely to slip more often than in writing, where a much higher degree of accuracy might be expected as a result of the extended time for planning.

Table 1 provides an overview of LINDSEI_CZ as regards the number of speakers, the length of the interviews, the number of the tokens, and basic metadata.

Table 1. LINDSEI_CZ – the volume of recorded and transcribed data

Choice of topic for Task 1	Length of A & B turns[3] in tokens	Length of B turns only in tokens	Duration of A & B turns (hh:mm:ss)	Duration of B turns (hh:mm:ss)	Mean length of interview in tokens	Mean duration of interview (mm:ss)
Country = 22 Film/play = 18 Experience = 10	123,761	95,904 mean = 1,918 (SD = 407)	12:52:25	10:37:42	2,475 (SD = 386)	15:27 (SD = 2:14)

---

[1] The Louvain International Database of Spoken English Interlanguage
[2] The International Corpus of Learner English
[3] The phrase "A turn" and "B turn" denote utterances made by the interviewer and the interviewee, respectively.

135

The method used is Computer-aided Error Analysis (Dagneaux et al., 1998). As a first step, errors were identified[4] in the transcribed corpus and annotated using a system of descriptive, incremental tags described in the Louvain Error-tagging Manual (Dagneaux et al., 2008). A typical tag has a number of positions which are occupied by letters denoting the particular type of error. The first position is the most general (for example the letters L and G denote lexical and grammatical errors), the subsequent positions add further detail (e.g. V in the second position stands for verb, and in the third position for voice). Thus GVT stands for **G**rammar-**V**erb-**T**ense and GVV for **G**rammar-**V**erb-**V**oice errors. The system can be easily adapted to suit the design of the study and the researcher's intentions. Thus Kämmerer (2009) extended the original tag LS used for marking Lexical Single errors to LSP to mark Lexical Single Prepositional errors. The tags are entered in round brackets.

As a second step, the target hypothesis is entered after the erroneous form. The target hypothesis is a suggestion made by the error annotator as to what he assumes the speaker was intending to produce. In other words, it is what the annotator assumes to be the correct or more appropriate form. This serves simply as a guide to the annotator's way of thinking, and hence only one target hypothesis is suggested although different options might be possible. Target hypotheses are entered between dollar signs. The following examples present two utterances containing tagged errors and target hypotheses. Example (a) shows a grammatical error in the use of a verb tense, example (b) then a lexical phrase (LP) error. In all of the examples in this study the examples are directly lifted from the transcribed corpus LINDSEI_CZ including the various transcription conventions (such as the dot denoting a silent pause). No punctuation is used in the transcriptions.

(a) the whole time it **(GVT) hasn't rained $didn't rain$** one day
(b) I think that I would **(LP) say . truth $tell the truth$**

As a third step all GV (Grammar-Verb) errors were extracted using simple queries in the concordancer AntConc (Anthony, 2014). The examples were then sorted, further coded and analysed using the data functions of MS Excel.

3. Results

The error identification and tagging process revealed a total number of 250 errors involving morphological aspects of verb usage. This constitutes 19.2% of the total number of errors (n=1,301) identified in LINDSEI_CZ (see Gráf, 2015b). In the whole corpus only 4 speakers did not commit any of these errors.

As expected, errors in the use of tenses (including tense agreement) are most frequent, forming 69.2% of all verb errors. These errors were committed by 42 (84%) speakers. The second group of most frequent errors is made up of errors in the use of the conditional,

---

4   Problematic cases were compared with similar utterances in the parallel corpus of spontaneous native speech LOCNEC. The comparison revealed that, for example, the use of the past perfect is common in native speech even in situations where its use is not entirely necessary as the sequence of events is clear from the context.

with 24 errors (9.6%) and 14 (28%) erring speakers. The remaining types are considerably less frequent (see Table 2).

Table 2. Verb error frequencies, numbers of erring speakers and mean numbers of errors

Type	Count	%	Number of erring speakers (n=50)	Mean number of errors per speaker	SD
GVT – verb tense errors (including tense agreement errors)	173	**69.20%**	42 (84%)	**3.46**	3.21
GV*C – errors in the use of the conditional	24	**9.60%**	14 (28%)	**0.48**	1.04
GVM – erroneous forms	20	**8.00%**	11 (22%)	**0.4**	1.22
GVAUX – erroneous use of auxiliary or modal verb	17	**6.80%**	14 (28%)	**0.34**	0.68
GVN – verb number errors	14	**5.60%**	10 (20%)	**0.28**	0.60
GVV – erroneous use of voice	2	**0.80%**	2 (4%)	**0.04**	0.20
Total	250	**100.00%**	**46 (92%)**	**5**	4.05

The mean numbers of errors and the high standard deviations (SD) show that the performance of the individual speakers varies significantly: for some of the categories only a small proportion of the speakers were involved and in most categories there appeared speakers who committed more errors in the individual categories than the others. This is more clearly visible from Figure 1 which shows the relevant boxplots.

Table 3 shows the frequency of errors in the use of the individual tenses. As was expected, the most problematic tense is the present perfect; it was erroneously used in 37 cases (28.46% of all GVT errors), and not used when it was supposed to in 58 cases (44.61%). Altogether then the present perfect (wrongly used or wrongly not used) appears in 73% of all verb tense errors. The following largest group is formed by errors in the use or non-use of the past simple, which is incorrectly used in 69 cases, and incorrectly not used in 37 cases. There also appear 20 cases of erroneous use and 8 cases of non-use of the present simple. The past perfect was erroneously used in 1 case, and not used when it was preferable in 24 cases. There are 3 cases of errors in the use of the present continuous, and 1 instance of wrongly used future simple. Errors in other tenses such as the future continuous or perfect have not been identified.

3.1 Present perfect errors

As has been shown above, the present perfect appears to be the most problematic tense for the speakers in LINDSEI_CZ. The 37 examples of its incorrect use were produced by 20 (40% speakers). The 58 instances in which a different tense was used erroneously instead of the present perfect were produced by 28 (56%) speakers. Altogether 34 speakers (68%) either produced at least one instance of incorrect use of the present perfect or at least one instance when they erroneously used a different tense instead. The following section provides an overview and typology of these errors.

Figure 1. Boxplots showing the distribution of errors in the following categories: verb tense including tense agreement (GVT_A), conditional (GV_C), form (GVM), voice (GVV), number (GVN), auxiliary and modal verbs (GVAUX). The last boxplot (Total) provides a summary of all of the instances.
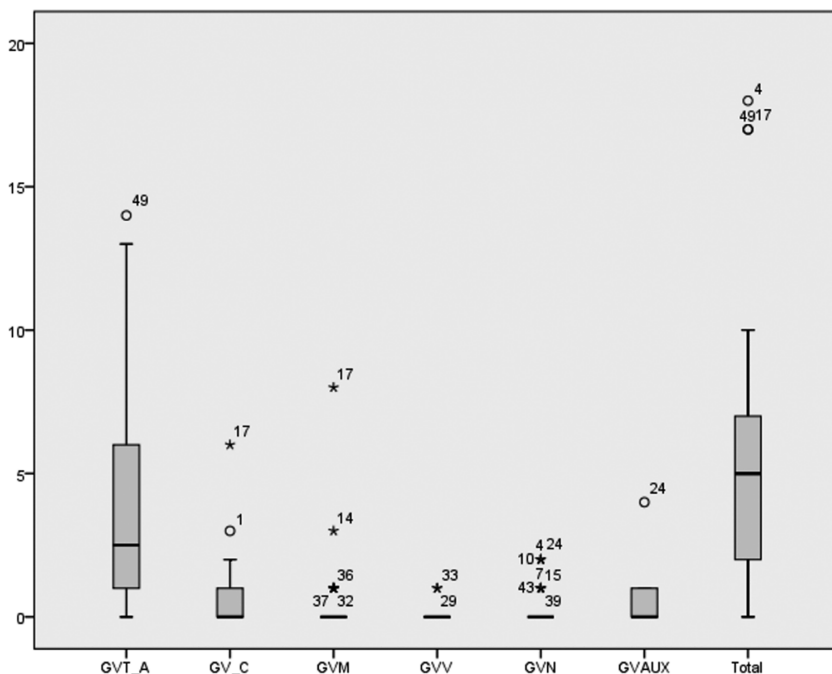

Table 3. Pivot table listing the frequency of erroneously used tenses (rows) and the hypothesised tenses (in columns)

Erroneously used/ Corrected	Past continuous	Past perfect	Past simple	Present continuous	Present perfect	Present perfect continuous	Present simple	Total
Future simple							1	1
Past perfect					1			1
Past simple	2	13			48	1	5	69
Present continuous					1		1	2
Present perfect		10	25				1	36
Present perfect continuous			1					1
Present simple		1	11	1	6	1		20
Total	**2**	**24**	**37**	**1**	**56**	**2**	**8**	**130**

138

3.1.1 Erroneous use of the present perfect

Let us first explore the instances in which the present perfect was used in place of a more appropriate tense (25 instances, 16 speakers). The tense was used despite the fact that specific time or occasion were explicitly mentioned as part of a larger narrative scheme thus requiring the use of the past tense or past perfect. A closer analysis of the examples reveals that there might exist a number of features which might trigger off the production of the present perfect even in situations which have no relation to the present. Such triggers are either typical situations in which the present perfect is used when it does refer to the present, or other lexical or grammatical features which the learner associates with the present perfect.

Such is the situation in which it is used to refer to recently completed actions in which the speaker expresses a sense of achievement (there were 4 such cases in our corpus). This is illustrated by examples (1) and (2) which were both part of a longer past-tense narrative.

(1) well first of all (GVT) I've learned $I learned$ (er) . to speak English very well there
(2) I think (GVT) it's encouraged $it encouraged$ me to to (er) do that

A similar situation (there were 3 cases in our corpus) may be observed in examples (3) and (4) in which the trigger is the sense of experience (which is even explicitly mentioned in example 4).

(3) after after high school […] (GVT) I've made $I made$ a wrong decision
(4) when I was seventeen . and (eh) (GVT) it's been $it was$ (em) great experience

Example (5) illustrates a type (there were 3 cases in our corpus) where the speaker expresses duration, which is often a concept learners associate and frequently practise in the present perfect.

(5) because (GVT) I've been $I was$ there for a year which was (eh) . a long time

Examples (6–8) illustrate how verbs commonly used in the present perfect in sentences referring to experience may act as its trigger even in sentences with past reference without any relation to the present. It may be hypothesised that such verbs are frequently used in sentences for present perfect practice. Our sample contains 7 such instances of the verb *see,* 2 cases of *visit*, and one of *meet*.

(6) (GVT) I've seen $I saw$ it (er) on my birthday
(7) (GVT) I've visited $visited$ (eh) Portugal . Lisbon (eh) this summer
(8) and secondly . (GVT) I've met $I met$. (er) many interesting people

Example (9) contains the intensifier *just* which is again frequently used in the present perfect and in practice sentences.

(9) they were incredibly (er) forthcoming . to us and (er) (GVT) it's just been $it was$ just wonderful

Example (10) appears to be another typical textbook and practice context for present perfect practice.

(10)   in my bachelor studies (GVT) I've been studying $I studied$ also (er) . also literature

In this group there are also cases, such as example (11), where no likely trigger can be identified.

(11)   and then (GVT) I've thought $I thought$ it would be great to to to be able to engage with

3.1.2 Failure to supply the present perfect

A large group of errors (58 instances, 28 speakers) revealing the complexity of the tense to the learners is made up of instances where the learners failed to produce the present perfect when it was due, replacing it by a different tense. Whilst in the previous section we could hypothesise about possible triggers (mainly teaching- or analogy-induced), the examples in this section are symptomatic of insufficient understanding or automatization of the rules, and some of the errors might be the result of L1 transfer. This is common for example in sentences which express duration at the present moment (see examples 12 and 13) which Czech expresses using the present simple, and also very commonly in complex clauses containing the conjunction *since* (see examples 14 and 15).

(12)   it is because I (GVT) teach $have been teaching$ the girl for about . three years already
(13)   but it it (GVT) it's $it's been$ open for just a few years
(14)   since that moment I . I (eh) (GVT) wished $have been wishing$ I could return .
(15)   ever since I was . thirteen or fourteen I I (GVT) had $have had$ a . list . of countries I w= I . want to . visit

Further errors in this group include those where the present perfect was not used for the result of a recent activity (ex. 16). Example (17) presents a common usage in American English and if it were not for the speaker's entirely British-like accent it would have to be excluded from our account.

(16)   that's how he . (em) how he (GVT) drew $has drawn$ her and how he probably sees her
(17)   like she (GVT) just came $has just come$ from her hairdresser's

Unlike in examples (6–8) which illustrated a common situation in which the speaker uses the present perfect to refer to experience which is located in a concrete situation or context, examples (18–21) show the opposite: here the students fail to use the present perfect when referring to a past experience whose effect is currently relevant.

140

(18) and I (GVT) was (er) in $have been to$ (eh) Germany twice
(19) it's written by a German author but . I (GVT) didn't read $haven't read$ it
(20) yes I (GVT) did $have done$ it before but you know this is like a sort of different experience because
(21) I've tried some translation I've I've I (GVT) did $have done$ some little little teaching

Considering the high proficiency of the speakers, these errors are somewhat surprising and clearly show that the automatization of the control of a complex tense which has no one-to-one parallel in the speakers' mother tongues is a slow process. It is, however, also possible that these instances are mere slips and we would have to take into account not only the erroneous instances in our corpus but also all of the correct ones. Similarly, we would not expect problems in sentences referring to an event started in the past and not finished at a present point in time, as this usage is usually taught early on in language courses and textbooks and revised frequently. Examples (22–23) show that errors do occur here as well, and we can only speculate if they are the result of L1 transfer, as Czech would use the past simple in the same situations.

(22) well I (GVT) always wanted to $have always wanted to$. (erm) .. so I hope . I can become a teacher but
(23) I don't think she (GVT) ever s= . stopped $has ever stopped$. as an au pair . I think she still . (eh) looks after

The last group of examples in which students incorrectly used a different tense other than the present perfect has altogether 31 examples produced by 18 speakers. They are sentences in which the present perfect refers to a past or recent event which has direct consequences at present. Their high incidence suggests (and the many years of my teaching experience support this view) that this usage of the present perfect poses great problems for the learners as they have to make a choice of the tense not only based upon the event's temporal reference but also on the judgement of the current relevance. The difficulty might arise from the fact that Czech does not make this distinction and simply uses the past tense for such examples.

(24) and he comes back .. and . all his arts collection everything is gone it (GVT) was $has been$ stolen
(25) in the past year (erm) . I I (GVT) changed $have changed$ a lot
(26) the picture is finished so he (GVT) showed $has showed$ her the the final version and she . seems . looks quite surprised
(27) I don't know why I just (GVT) didn't develop $haven't developed$ any relationship with it

3.2 Errors in the use of other tenses and aspect

Besides the group of 95 errors in the use of the present perfect, our corpus contains 35 examples of sentences containing erroneous use of other tenses produced by 20 speak-

ers. The largest group is made up of instances of erroneous use or non-use of the past perfect (altogether 23 errors committed by 14 speakers). Most frequently the past simple is used instead, which makes the actual sequence of events rather unclear. In example (28) the speaker is referring to problems she had experienced prior to the trip which she is describing, but the lack of the past perfect might imply the problems had been experienced during the same trip. Similarly, the use of the past perfect would have prevented the ambiguity as to the temporal interpretation present in example (29), where we are left wondering whether the portrayed person was unhappy with the result or the process of being painted.

(28) there had to be plenty of water because we (GVT) experienced $had experienced$ some problems with water
(29) so probably she's not happy with the way she (GVT) was $had been$ painted

There appear to exist a number of triggers leading to the failure to produce the past perfect. This is true of the cases in which it is replaced by the present perfect,[5] as in examples (30–33). Such triggers include the chunk *been to*, the verbs *read* and *see* and the adverb *just*, all of which are frequently used during present perfect practice.

(30) I tried to travel as much as I could because (eh) (GVT) I've never been $I had never been$ to Ireland before
(31) and it it was amazing I (GVT) I've read $I'd read$ the play before though so that might have made it
(32) it was my dream to see it because I (GVT) have seen $had seen$ many films and and serials about it
(33) I'd never been to Egypt . and I: . (GVT) I've $I'd$ just (erm) . been reading many books about it

Examples of errors in the use of other tenses are much less frequent and involve fewer speakers. Two speakers use the future simple in a conditional clause after the conjunction *if,* as in example (34). This is possibly caused by L1 transfer (cf. the Czech *jestli budu přijat*). The remaining examples are most likely slips which are difficult to categorize and explain, as in examples (35–36).

(34) hopefully I'd like to if if (GVT) I'll be $I am$ admitted
(35) I already (GVT) know $knew$ them very well
(36) but . I wasn't that interested I just . I (GVT) like $liked$ the social studies

Only four speakers erred in the selection of aspect, using continuous for simple, as in example (37), or simple where continuous would appear more natural, as in examples (38) and (39), which are actually harder to interpret as a result.

(37) while she oh she's actually (erm) having a squint $she actually has a squint$

---

5 The frequency of these examples is included in section 3.1.1.

(38) she didn't know (erm) what she (GVTA) signed $was signing$

(39) and they they told us when we (GVT) left $were leaving$ you know you should come to visit us . much more often

3.3 Erroneous sequence of tenses in reported speech and dependent clauses

In our corpus there are also 43 errors (committed by 22 speakers) in the use of tenses in dependent clauses following a main clause in the past tense. From a pedagogical point of view, this is another area of great difficulty for Czech speakers of English and one frequently practised and taught as the tense in Czech dependent clause is not affected by the tense in the main clause. In our system of classification, these errors form a sub-group of the grammar-verb-tense errors and are tagged as GVTA, where the A stands for *agreement,* which is a label commonly used in language textbooks (i.e. tense agreement).[6] Eleven examples occur in reported speech after verba dicendi. These include the verbs *ask, say,* and *tell* and also the verb *text* (as in *to write in a text message*) (see examples 40–42).

(40) when I came back my mum (er) told me that (er) she (GVTA) expected $had expected$ something like that to happen

(41) when I was standing here my father called me and asked me how I (GVTA) am was

(42) I just texted her (GVTA) I'm $I was$ coming home three days (er) (er) earlier

Other verbs include *think* (6 times), *feel* (5 times), *realize* (3 times) and single instances of *agree, decide, figure out, find out, check, know, make clear, see, suppose, understand, want to know/show*, and the copular *be (surprised)* and *be (shocked)* (see examples 43–46). Example (46) illustrates that such an error can actually affect the meaning and interpretation of the whole sentence; the one produced by our speaker would imply that the decision in question had not yet been made, whilst the opposite (as suggested by the target hypothesis) is actually the case.

(43) I thought that I (GVTA) can $could$. see the sea (eh) or ocean (eh) maybe it was just my impression

(44) I was there for four months and so in the middle of it I really felt I (GVTA) want $wanted$ to leave

(45) he made it quite clear that he (GVTA) doesn't $didn't$ really like the other boy

(46) the other women agreed that this (GVTA) was $had been$ entirely her decision

Further 2 errors have been recorded in the sequence of tenses in clauses of comparison, as exemplified in (47).

(47) when he would . paint her as he . (eh) (GVT) sees saw sees her

---

6   Another commonly used label is sequence of tenses, or in Czech časová souslednost.

3.4 Conditional errors

Besides tense errors, the recorded speakers also committed errors in the use of the conditional. There were a total of 24 of these errors, and they were committed by 14 speakers. They include mainly the wrong choice of mood in conditional clauses introduced by *if*, as in example (48), and instances of failing to use the perfect infinitive form in the if-clause when referring to a past event, as in example (49). Whilst the conditional is frequently practised in language classes, experience tells me that especially the latter error is fairly frequent even with advanced learners. This might be the result of L1 transfer, as contemporary informal Czech tends not to use the past conditional. This also applies to the frequent error in the failure to use the past perfect when referring to a hypothetical past event, as in example (50).

(48)   if . (er) the woman . (er) (GVAUXC) would be $were$. (er) a really good friend
(49)   and it would (er) (GVAUXC) cause $have caused$ me like (er)
(50)   if they (GVTC) stayed $had stayed$ over there and in the place where they were so they would have probably become

3.5 Other types of verb errors
(auxiliaries, number, voice, form)

For convenience sake, the remaining 4 groups of errors are dealt with in one section, including the group of 17 errors in the use of auxiliary and modal verbs (tagged GVAUX), examples 51–54), 14 errors in the verb number category (tagged GVN, examples 55–57), 2 examples of errors in the use of grammatical voice (tagged GVV, example 58).

(51)   we . weren't in a hotel .. so . I (GVAUX) 0 really looking $was really looking$ forward
(52)   all the weeks of the study . (eh) including holidays . (er) we (GVAUX) should have read $were supposed to read$
(53)   I always try to google it so that I (GVAUX) would 0 remember in the future
(54)   there is nothing to carry sound and so you . of course you (GVAUX) 0 can hear the voices of the characters
(55)   everyone in (erm) (eh) . in my surrounding . around me . (GVN) know $knows$ English
(56)   maybe the hair . the the her hair (GVN) are is different
(57)   an experience that (GVN) have has . taught me something
(58)   when the children are running to (eh) (GVV) get hidden $hide$ somewhere in the garden

The examples in these groups are so few that it is impossible to work out any patterns or areas of particular difficulty: in the GVN group there appear 9 instances of the wrong form of the third person singular verb (missing *-s*). Example (52) shows a fairly common error in translating the Czech *měli jsme* as *we should have done* rather than *we were supposed to do.* Transfer is most likely at play in example (53) where in the dependent

clause Czech uses conditional while English uses indicative, and in example (56) where the erroneous use of the plural form of the verb was most likely the result of the Czech equivalent for *hair* being the plural noun *vlasy*.

The last group of errors is difficult to categorize; they include odd instances of slips (tagged GVM) in the use of the correct form (example 59) or even in the use of negation (ex. 60). The low frequency of these errors shows that they are not systematic and do not require pedagogical intervention.

(59)　case . (er) it was very interesting for me to (GVM) found $find$ out that actually

(60)　yeah I am but I (GVM) no study $don't study$ (LP) English language

4. Discussion

One of the key concerns of this study was the exploration of spontaneous speech produced by advanced learners with the view to ascertaining the key types and frequency of verbal errors and thus identifying those grammatical aspects that would appear most problematic. The results have shown that despite their high proficiency nearly all of our advanced learners make occasional or even more regular slips in the area of verb usage, and this is especially true of the use of tenses. But even here there appear areas of particular difficulty, which is especially the use of the present perfect, and tense agreement.

The results are similar to Granger (1999), and especially to Götz's (2015) study of grammatical errors produced by the participants in the German subcorpus of LINDSEI. LINDSEI_CZ and LINDSEI_GE are very similar in size, the total number of errors identified (1,301 and 1,335 errors respectively) and their distribution. Götz identified 235 GVT errors, which make up 17.6% of the total number of errors in LINDSEI_GE. The proportion of GVT errors in the LINDSEI_CZ is 13.3%.

The major difference between the two corpora is the larger number of aspect (confusing simple and continuous) errors in the German subcorpus (30.3% of all GVT errors in German as opposed to 3.2% in Czech), and the disparity between the frequency of errors involving the present perfect (38% of all erroneous instances in Czech as opposed to 28% in German).

Similarly to our finding, Götz recognizes negative transfer from the L1 as one of the key factors causing the production of errors. A deeper comparison of these two similar studies cannot be performed at present, as Götz's classification of errors worked with a somewhat courser typology of errors (e.g. not dealing systematically with tense agreement errors).

Likewise, Davydova (2011) in her complex study of the present perfect produced by speakers of a variety of L1s finds similar causes for the difficulty these speakers experience – mainly the absence of the tense in the L1 (as in the case of Russian speakers) and the consequent attempt to express the notion in question using structures the speaker is familiar with from the L1.

The present study has also revealed that many of the identified errors appear to have been triggered off by the presence of some other component which the learners tend to associate particularly strongly with a different tense than the one which they ought to

use. These errors may be teaching induced as in the grammar presentation and practice components of language courses and textbooks such features are highlighted and often made the focus of intensive and repeated practice. Our findings here are comparable to Granger's (1999), who identified similar triggers in her study of advanced English writing by native speakers of French. Götz (2015), for comparison, does not deal with such triggers, and it must be admitted that they might be one of several possible explanations for these errors.

The findings of our study are limited in many ways. First and foremost, the very process of identification of errors is problematic (as, for example, different norms may apply to the so-called spoken grammar). The work on the double-checking of our corpus is not completed yet and consequently we cannot provide any information regarding inter-annotator agreement. Secondly, despite the laboriousness involved in their compilation, spoken learner corpora are usually much smaller than written ones. This means that we cannot be quite certain whether the recorded instances are examples of systematic or random behaviour. Besides, LINDSEI subcorpora have not yet been morphologically tagged (although the work has commenced now), which makes it virtually impossible to perform full-scale analyses of the uses of individual grammatical features, for example comparing the erroneous and correct use of tenses. Thus our study focuses only on errors and not on the correct usage. This is an area which must be addressed in future research. Last but not least, the production is affected by the design of the task: thus owing to its design, LINDSEI may not have provided the speakers with as much opportunity to use the future tense (or refer to the future using other grammatical constructions) whilst the opportunities to use the past or present perfect tenses are plentiful. Despite all this uncertainty, we find that our analyses are in agreement with our teaching experience and results published in similar studies elsewhere. They are also some of the most recent analyses of advanced English spoken by Czech learners and may thus provide an important point of comparison for future studies.

5. Conclusion

The analysis has shown that even at an advanced level there exist aspects of grammar which are particularly error-prone. These are especially the use of tenses (especially the present perfect), tense agreements and conditionals. Whilst in only a small number of instances the errors here led to reduced intelligibility or even altered meaning, advanced learners – who were in this case students of English philology and thus aspiring to become experts in the field – would no doubt like to think of their English speech performances as not being marred by such errors.

The majority of the errors could be explained either by the effect of negative L1 transfer or by the existence of various triggers or false associations. Our findings thus have pedagogical implications which point in two possible directions: we recommend that work with advanced learners should systematically target these error-prone areas not leaving the L1 aside, and we also call for a reconsideration of how these problematic areas are approached in the teaching process, whether in practice or in the explanation and presentation of grammar aspects.

We strongly advocate the following principles: 1) discourse-driven approaches to the presentation of grammar so that learners become aware of the importance and role of context upon the selection of grammatical features; 2) moving away from decontextualized sentence-based presentation and practice of new points of grammar so that learners become accustomed to choosing aspects of grammar depending on the prevailing notion (such as cause and effect) rather than by clinging to overt features (such as the various triggers); 3) data-driven learning so that learners may benefit from analysing typical performance problems of other students of the same level of proficiency and learn from each other's mistakes; and 4) using L1-informed approaches to presenting new language material as these might help – if they are well designed – to avoid transfer related problems.

Whilst our study is limited in the number of learners (n = 50) and the size of the corpus (c. 125,000 tokens), it analyzes the largest collection of advanced spoken English produced by native Czech speakers available to date. As such it might be of benefit to language teachers and also researchers who should not only extend this research to other areas of grammar and lexis but also make a full account of those features which these learners find unproblematic.

REFERENCES

Anthony, L. (2014) *AntConc (Version 3.4.3)*. Tokyo: Waseda University.

Carlsen, C. (2012) Proficiency level – a fuzzy variable in computer learner corpora. *Applied Linguistics* 33/2, 161–183.

Clark, E. V. (2009) *First Language Acquisition*. Cambridge: University Press.

Conrad, S. (2016) Using corpus linguistics to improve the teaching of grammar. In: Hinkel, E. (ed.), *Teaching English Grammar to Speakers of Other Languages*. New York: Routledge.

Corder, S. P. (1967) The significance of learners' errors. *International Review of Applied Linguistics* 5, 161–170.

Corder, S. P. (1974) Error analysis. In: Allen, J. P. B., and S. P. Corder (eds.), *Techniques in Applied Linguistics*. London/New York: Oxford University Press.

Council of Europe (2001) *Common European Framework of Reference for Languages. Learning, teaching, assessment*. Cambridge: University Press.

Dagneaux, E., S. Denness and S. Granger (1998) Computer-aided error analysis. *System: An International Journal of Educational Technology and Applied Linguistics* 26/2, 163–174.

Dagneaux, E., S. Denness, S. Granger, F. Meunier, J. Neff Van Aertselaer and J. Thewissen (2008) *Error Tagging Manual Version* 1.3. Unpublished manual. Université catholique de Louvain.

Davydova, J. (2011) *The Present Perfect in Non-native Englishes: a Corpus-based Study of Variation*. Berlin/Boston: De Gruyter Mouton.

Dietrich, R., W. Klein and C. Noyau (1995) *The Acquisition of Temporality in a Second Language*. Amsterdam/Philadelphia: J. Benjamins.

Dušková, L. (1969) On sources of errors in foreign language learning. *International Review of Applied Linguistics* 7, 11–36.

Eriksson, A. (2004) *Tense and Aspect in Learner Writing. Advanced Swedish Learners' Use of Tense and Aspect in English Argumentative Text*. Göteborg: University of Göteborg.

Eriksson, A. (2008) *Tense and Aspect in Advanced Swedish Learners' Written English*. Göteborg: Acta Universitatis Gothoburgensis.

Gilquin, G. and S. De Cock (2011) Errors and disfluencies in spoken corpora: Setting the scene. *International Journal of Corpus Linguistics* 16/2, 141–172.

Gilquin, G., S. De Cock and S. Granger (2010) *The Louvain International Database of Spoken English Interlanguage. Handbook and CD-ROM*. Louvain-la-Neuve: Presses universitaires de Louvain.

Götz, S. (2015) Tense and aspect errors in spoken learner language: Implications for language testing and assessment. In Marcus, C. (ed.), *Learner Corpora in Language Testing and Assessment*, 191–215. Amsterdam/Philadelphia: J. Benjamins.

Gráf, T. (2015a) *Accuracy and Fluency in the Speech of the Advanced Learner of English* (Ph.D. Thesis). Prague: Charles University.

Gráf, T. (2015b) *Korpus LINDSEI_CZ*. Prague: Charles University.

Granger, S. (1996) From CA to CIA and back: an integrated approach to computerized bilingual and learner corpora. In: Aijmer, K., B. Altenberg and M. Johansson (eds.), *Languages in Contrast. Text-based Cross-linguistic Studies*, 37–51. Lund: Lund University Press.

Granger, S. (1999) Use of tenses by advanced EFL learners: evidence from an error-tagged computer corpus. In: Hasselgård, H. and S. Oksefjell (eds.), *Out of Corpora. Studies in Honour of Stig Johansson*, 191–202. Amsterdam/Atlanta: Rodopi.

Granger, S. (2009) The contribution of learner corpora to second language acquisition and foreign language teaching: A critical evaluation. In: Aijmer, K. (ed.), *Corpora and Language Teaching*. Amsterdam/Philadelphia: John Benjamins.

Granger, S. (2013) The passive in learner English: Corpus insights and implications for pedagogical grammar. In: Ishikawa, S. (ed.), *Papers from LCSAW2013* 1, 5–16. Japan: Kobe University.

Granger, S., F. Meunier and G. Gilquin (eds.) (2015) *The Cambridge Handbook of Learner Corpus Research*. Cambridge: University Press.

Guo, X. (2005) Modal auxiliaries in phraseology: a contrastive study of learner English and native speaker English. In: Danielsson, P. and M. Wagenmakers (eds.), *Proceedings from the Corpus Linguistics Conference Series* 1, Birmingham. Available at: http://www.birmingham.ac.uk/research/activity/corpus/publications/conference-archives/2005-birmingham.aspx.

Hinkel, E. (2004) Tense, aspect and the passive voice in L1 and L2 academic texts. *Language Teaching Research* 8/1, 5–29.

Housen, A., F. Kuiken and I. Vedder (eds.) (2012) *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA*. Amsterdam/Philadelphia: J. Benjamins.

Ingram, D. (1989) *First Language Acquisition: Method, Description and Explanation*. Cambridge: University Press.

James, C. (1998) *Errors in Language Learning and Use: Exploring Error Analysis*. London/New York: Longman.

Kammerer, S. (2009) Error-tagging spoken features of (learner) language: The UCL Error Editor revised. In: *Papers presented at the 30th annual conference of the International Computer Archive of Modern and Medieval English (ICAME 30)*. Lancaster.

Lado, R. (1957) *Linguistics Across Cultures: Applied Linguistics for Language Teachers.* Ann Arbor: University of Michigan Press.

Merilainen, L. (2010) *Language Transfer in the Written English of Finnish Students*. Joensuu: University of Eastern Finland.

Rogatcheva, S. I. (2009) I've only found the answer a few days ago: aspect use in Bulgarian and German EFL writing. In: Prado-Alonso, C., L. Gomez-Garcia, I. Pastor-Gomez and D. Tizon-Couto (eds.), *New Trends and Methodologies in Applied English Language Research. Diachronic, Diatopic and Contrastive Studies* 103, 255–278. Berlin: Peter Lang.

Rogatcheva, S. I. (2012) Measuring learner (mis)use: Tense and aspect errors in the Bulgarian and German components of ICLE. In: Thomas, J. and A. Boulton (eds.), *Input, Process and Product: Developments in Teaching and Language Corpora*, 258–272. Brno: Masaryk University.

Rundell, M. (2007) *Macmillan English Dictionary for Advanced Learners*. Oxford: Macmillan.

Salaberry, M. R., Y. Shirai and American Association for Applied Linguistics (eds.) (2002) *The L2 Acquisition of Tense-aspect Morphology*. Amsterdam: J. Benjamins.

Swan, M. and B. Smith (eds.) (2001) *Learner English: A Teacher's Guide to Interference and Other Problems* (2nd ed.). Cambridge, New York: Cambridge University Press.

Turton, N. D. and J. B. Heaton (1996) *Longman Dictionary of Common Errors*. Harlow: Longman.

CHYBY V UŽÍVÁNÍ ČASŮ V MLUVENÉ ANGLIČTINĚ POKROČILÝCH MLUVČÍCH

Resumé

Cílem studie bylo prozkoumat mluvený jazyk pokročilých mluvčích angličtiny s ohledem na chyby a nepřesnosti, jichž se dopouštějí v užívání morfologických kategorií sloves. Zdrojem dat pro studii byl korpus pokročilé žákovské angličtiny LINDSEI_CZ, který obsahuje transkribované nahrávky 50 studentů vyšších ročníků anglistiky na FF UK. Metodou, který byla využita, byla počítačová analýza chyb (Computer-aided Error Analysis). Její výsledky byly zpracovány kvantitativně, aby mohly být rozpoznány ty gramatické jevy, které pro pokročilé studenty představují nejčastější problém. Ukazuje se, že i takto pokročilí studenti chybují v poměrně základních užitích časů a časové souslednosti, a to především v užívání předpřítomného času. Vedle toho studie identifikuje i řadu jiných typických problémů. Studie poukazuje na to, že tyto problémy mohou pramenit z negativního transferu z mateřštiny, ale mohou být rovněž důsledkem nevhodných technik využívaných při vyučování těchto jevů. Řada příkladů totiž obsahuje společné prvky, které získávají pozornost při výkladu a procvičování gramatiky na úkor pěstování hlubšího pochopení takových sémantických konceptů jako je průběhovost či „perfektivita" (perfectiveness). Rozhodování studentů při volbě správného gramatického jevu pak může být ovlivněno přítomností marginálního prvku (např. adverbia, které se hojně využívá v určitém jevu), jež pak slouží jako spouštěč chyby, spíše než vyhodnocením kontextu a situačního zasazení promluvy. Studie následně vyzývá k přehodnocení postupů užívaných při vyučování gramatiky, a to především k odklonu od dekontextualizovaného procvičování a výkladu gramatiky na základě striktních pravidel a k příklonu k diskurzně orientovanému procvičování a vytváření hlubšího pochopení složitých konceptů ovlivňujících volbu času a vidu.

Tomáš Gráf
Department of English Language and ELT Methodology
Faculty of Arts, Charles University
Tomas.Graf@ff.cuni.cz

ACTA UNIVERSITATIS CAROLINAE
PHILOLOGICA 1/2017