Constructing the global from the local: On the FSP status of keywords in academic discourse

Renata Pípalová (Praha)

ABSTRACT

Attaching a set of keywords has become a norm, or a convention at least, in most academic publications. Surprisingly, these prominent items, encoding the Global Theme of academic discourse and fulfilling numerous other functions, have yet to receive adequate linguistic attention. In syntactic terms they tend to be rather uniform, realized mostly by isolated nouns or by noun phrases (Pípalová 2017). This paper seeks to explore their in-text use (iteration) and FSP standing in authentic research articles. The work, which is part of a larger study, is framed in terms of two objectives. Firstly, it looks at varying frequencies of keyword items and at their distribution across the diverse parts of research articles. Secondly, examining both their explicit and implicit realizations, the paper strives to verify their thematic status in individual sentences. Established on a specialized corpus of recent academic articles drawn from peer-reviewed international journals, the paper attempts to balance quantitative and qualitative research and to correlate the operation of keywords at micro- and macrotextual levels. The results of the study should enrich FSP research and may also have practical relevance for Academic writing courses.

KEYWORDS

corpus, academic discourse, global theme, FSP theme, keyword, keyword iteration, communicative (distributional) field, FSP pattern

DOI

https://doi.org/10.14712/18059635.2019.2.5

1. INTRODUCTION

It is well-recognized that in Research Articles (RAs) a custom has developed whereby the authors identify the Global Theme in a dual manner — by title, and by a set of keywords (KWs; for more, see e.g., Pípalová 2017). This practice, intriguing in its own right, was considered ideal for the purposes of a research project aiming to explore the interrelationships between diverse themes operating at various levels in a hierarchy.

Unlike academic titles, whose numerous aspects have been researched thoroughly for over 30 years, the part played by keywords has been largely neglected. Most of the scholarly work devoted to the subject discusses their role in automatic information retrieval or in the context of the thematic concentration of texts by means of frequency methods (see, e.g., Čech 2016). Linguistic studies proper, however, are very rare. Todd (2011) explores the identification of topic boundaries and topic keywords by informants, using four methods of analyzing topics, viz. topical structure analysis, given-new progression, lexical analysis, and topic-based analysis. In Pípalová (2017), keywords are analysed in terms of the multiple functions they serve in academic dis-

course and studied from the syntactic and FSP standpoints. Attention is paid only to KW types, i.e., KW items establishing KW sets. The paper shows their largely nominal character, inasmuch as they are mainly realized either as syntactic nouns or noun phrases (NPs). Since the (pre)modifier as a rule precedes the head noun and tends to be rhematic, most KWs are conventionally arranged counter to the scale of rising CD (see Firbas 1992: esp. 66–87).

The present study focusses on the KWs exclusively. Rather than investigating KW items (as types) enlisted in KW sets, it assesses their actual use in the bodies of the RAs. Given their role as (one of the two) embodiments of the Global Theme, the KWs were selected by the authors of the RAs themselves as highly prominent concepts, indispensable for the coherent reception of the academic discourse in question. Accordingly, this paper, which represents only part of a larger study, has two specific goals. Firstly, it aims to explore the quantitative aspects of in-text KW iteration. Secondly, it seeks to examine the FSP standing of their tokens in the particular bodies of the RAs. In this context it might be mentioned that the FSP analysis is performed at various levels, ranging from main clause down to phrase level.

2. METHODOLOGY

2.1 DATA

2.1.1 CORPUS

For the purposes of the current study, a specialized corpus was compiled from RAs recently published (01–12/2017; 01–06/2018) in two peer-reviewed, international linguistic journals (*Language & Communication; Journal of Pragmatics*). The RAs incorporated in the corpus had to satisfy two criteria. Firstly, they had to exhibit the same conventional structure (see below). Secondly, the author (or at least one of the co-authors) had to be affiliated with a university in the inner circle English-speaking countries (Kachru). Interestingly, such conditions were satisfied by 6 RAs in *Language & Communication* and by 14 RAs in *Journal of Pragmatics*, i.e., by 20 RAs altogether. A complete list and bibliography is appended at the end of this paper. For obvious reasons, it was decided that only relevant parts of the RAs would be scrutinized. Thus, passages which proved to be irrelevant for the design of the research (e.g., examples, tables, notes, pictures and decriptive legends, bibliography, and the like) were disregarded.

With respect to quantitative parameters, the corpus assembled amounted to 150,000 words approximately. More specifically, it comprised 149,426 words, plus an additional 182 words included in keyword sets (KW sets), i.e., 149,608 in total. The corpus featured 108 KW items/types in KW sets. Since the 20 RAs that formed the basis of the study exhibited between 4–6 KW items in their KW sets, the mean figure was 5.4 KW items per KW set. With the individual KWs ranging between 1–4 graphic words, the mean length turned out to be 1.69 (graphic) words per KW type in the KW set. These findings are in accord with those published elsewhere (Pípalová 2017).

As might be expected, apart from the subgenre of the KW sets located by convention in both the source journals post-initially, i.e., after the name of the author/s and

their affiliation, each RA comprised seven other parts/subgenres/sections: Title, Abstract, Introduction, Method(ology), Results, Discussion and Conclusion, here referred to as the "body of the RA" for ease of reference. The corpus contained 140 such segments, in all of which the iteration of the KW items was explored. In this connection it should be noted that in the corpus the divisions referred to sometimes received slightly different labels. For example, the Methodology subgenre was occasionally identified as Method; Data and Method; Materials and Methods; Methods, Participants and Setting, etc. As a headline for the Results subgenre, Results and Data Analysis; Analysis; Findings; and so forth, were used. On one occasion, too, the Discussion subgenre was marked Discussion of Findings.

2.1.2 BROAD AND NARROW CORPORA

It should be noted that in this study, a distinction will be systematically made between two types of corpora differing in size, viz. the Broad Corpus and its constituent — the Narrow Corpus. The Broad Corpus comprises the above 20 RAs (approximately 150,000 words) and serves chiefly to uncover diverse tendencies using quantitative analysis. The Narrow Corpus, on the other hand, is constituted by two RAs (approximately 14,000 words) randomly selected from the Broad Corpus. Delimited largely for ease of manipulation, the Narrow Corpus is designed for quantitative as well as qualitative analysis.

2.2 PROCEDURE

2.2.1 PROCEDURE REGARDING BROAD CORPUS

In order to determine the size of the individual RA sections (which may be considered subgenres sui generis) in word terms, the AntConc corpus software was employed. The same software was likewise used to establish the share of KWs (iterations) in the sections themselves. The concordance function of the software was also found suitable for exploring direct iteration of KW items and their distribution across RA sections. The analysis was conducted in order to gain insight into the tendencies which govern KW iteration in current RAs. It was expected that the findings based on the Broad Corpus would contextualize the results subsequently drawn from the Narrow Corpus.

2.2.2 PROCEDURE REGARDING NARROW CORPUS

Delimited largely for ease of manipulation and for manageability purposes, the Narrow Corpus was designed for quantitative as well as qualitative analysis. The first step involved the delineation of KW clusters. In fact, in addition to direct iterations of KW items, various other notions semantically (and at times also formally) closely related to the KW in question were deemed crucial for the coherent perception of the RA. Hence networks of mutually interrelated items, including not only direct explicit iterations of KWs but also their morphological variants, allolexes, derived and implicit tokens, as well as pronominal and elliptical replacements, were taken into account. Although at times the boundaries of the KW clusters turned out to be rather fuzzy, nevertheless it was vital to delineate them on principled grounds for the sake of fur-

ther quantification. It should be stressed that the tokens gained were carefully assessed in their co-text, so as to incorporate only those relevant (and to exclude their non-relevant counterparts, e.g., their homonyms). Once the clusters were identified, the AntConc tool was used to explore their token distribution across the individual RA sections. Lastly, after this analysis was conducted, the relative weighting of individual KW clusters in the RA was perfomed, since the clusters showed considerable variation in terms of range and the number of tokens.

2.2.3 FSP PROCEDURE

The aim of the second step was to examine the FSP standing of the KW cluster tokens detected. It should be recalled that according to Firbas (1992: 17), Svoboda (1968; 1989: 76–7), Dušková (2015: 335–349), and others, every sentence, clause, or phrase represents a so-called Communicative (Distributional) Field (CF) of Communicative Dynamism (CD). Thus, there is a delicate hierarchy of distributional fields. The FSP analysis carried out in the present study was accomplished at several levels of a hierarchy simultaneously, namely at:

Main clause (MC) level (CF0)
Subordinate clause (SC) level (CF1)
Phrase (PH) level (often CF2) (cf. Svoboda 1968)

For simplicity's sake, only T-R (i.e., theme-rheme) distinction was systematically marked, thereby disregarding the subtler functions on a scale devised by Firbas (1992: esp. 66–87, i.e., Theme proper, Diatheme, Transition proper, Transition, Rheme, Rheme proper). Nevertheless in harmony with Firbas (1992: ibid), in the present study Transition functions were assigned to Non-theme, i.e., here marked as the Rheme.

To achieve methodological transparency, a simple notation scheme was devised that would capture the values of the analyses performed at three distinct levels simultaneously, namely at Main Clause level (MC), at Subordinate Clause level (SC), and at Phrase level (PH). Admittedly, the scheme may seem somewhat simplified inasmuch as it not only disregards a variety of sub-functions (e.g., Diatheme, Theme proper), but also in that it does not employ indices to take account of potential recursive embeddedness (e.g., the number of subordinate matrix clauses). For the purposes of the present paper, however, the author preferred to keep the patterns uncomplicated in order to better deliver her point, since it was felt that introducing indices to mark degrees of potential embeddedness would not affect the results to any significant extent.

The procedure can be illustrated using the following sentence: *Empirical research found that some domains within the cognitive function are significantly more language-dependent than others*, which includes a token of the keyword(s) (identified by the author of the source text) "cognitive function." To establish its FSP standing at diverse levels, a simple notation scheme was developed. In what follows, the level of analysis is particularized initially, and in the example itself the respective Communicative Field (CF), (see Svoboda 1968), within which the FSP functions are assigned, is underlined and the boundaries of its T and R functions are indicated by bracketing.

MC LEVEL (CF0): *Empirical research (T0) found that some domains within the cognitive function are significantly more language-dependent than others, …(R0)*
SC LEVEL (CF1): *Empirical research found that some domains within the cognitive function (T1) are significantly more language-dependent than others,…(R1)*
PH LEVEL (CF2): *Empirical research found that some domains (T2) within the cognitive function (R2) are significantly more language-dependent than others, for example…*

From the above analyses (conducted within CF0, CF1, CF2) we can characterize the FSP standing of this token of KW "cognitive function" using the R — T — R pattern. This means that the token takes the rhematic function (R0) at main clause (MC) level, thematic (T1) at subordinate clause (SC) level, and rhematic (R2) at phrase level (PH). In other words, this pattern ranks among the mixed categories, inasmuch as at least one function is thematic and at least one is rhematic. Since in this particular case the rhematic functions prevail, the pattern reveals a considerable amount of KW dynamicity. An overview of all the patterns gained from the data is provided below.

In line with Svoboda (1968: 72) it is held here that the communicative impact decreases with the levels of inferiority in the hierarchy. "As for the differences in CD between CU's (i.e., communicative units, note inserted by RP) occurring within a CF of first or even more inferior rank, they may be as great as those between the CU^0's within a CF^0. Examined from the viewpoint of the nearest superior rank, however, they may seem rather small or even irrelevant." Hence R1 bears more dynamism than R2 and T0 is less dynamic than T1.

2.3 HYPOTHESES

Before the research was carried out, the author had two hypotheses, to wit: as embodiments of the Global Theme, KWs will be largely employed in Thematic functions at diverse FSP levels. The thematicity of the KW items will grow with the flow of the RA text, reflecting the increasing degree of their boundedness/contextual dependency/activation.

3. RESULTS

3.1 BROAD CORPUS RESULTS

Overall, the Broad Corpus featured 20 KW sets, comprising 108 KW items (types), with 5.4 KW items per KW set on average, which is in harmony with some earlier findings (Pípalová 2017). Of the KW items employed in the KW sets, less than half were realized solely by syntactic nouns (i.e., N-pattern, 45.37%), whereas over half (54.63%) by noun phrases (NPs, i.e., composed of a head noun together with its modification). Generally speaking, in NPs the modifier (attribute) may take the form of a premodifier, a postmodifier or a pre- and postmodifier. In the data, among the KW items realized by NPs, premodification clearly prevailed, corresponding to 94.93% of instances. More specifically, simple premodification of the AN pattern (displaying a single premodifying unit) accounted for 76.28% of all, while premodification of the AAN pattern (exhibiting two or more pre-modifying units in succession) consti-

tuted 18.65%. In contrast, postmodification (whether with or without premodification) amounted to only 5.07% of all.

As for the iteration of KWs in the body of the RAs, tokens of KWs realized as syntactic nouns (N-pattern) by far outnumbered tokens of KWs realized by phrases (NP-pattern). Indeed, there were 49 KW types of the N-pattern in KW sets, matching 1,394 tokens in the RA data. Hence the mean figure for this group was 28.45 tokens/type. On the other hand, there were 59 KW types of the NP-pattern in KW sets, which corresponded to 660 tokens in the RA bodies. Consequently, this group yielded 11.19 tokens/type on average.

In fact, the Broad Corpus analysis uncovered no correlation between the number of KWs in a KW set and RA length. Neither did it confirm any correlation between the number of KWs in a set and the number of KW tokens traced in the RA. It would appear that RAs of similar lengths and identical numbers of KWs in sets may produce very different token results. For example, RA No. 14 which embraced 8,245 words, featured 6 KW types in the KW set, matching 153 KW tokens in the body of the RA. RA No. 15, whose body of the text exhibited 8,417 words, was marked by 6 KWs in the set, correlating with 220 tokens of direct iterations. RA No. 20, totalling 8,767 words, had 6 KWs in the set and 148 tokens in the body of the RA. Last of all, RA No. 16 was made up of 6,656 words, its KW set involved 4 KWs and the RA body only 1 token.

With respect to the results based on the investigation of the Broad Corpus (149,426 words, disregarding KW sets; 108 KW types), direct iteration of KWs had a relatively negligible share in constituting RAs. In actual terms, there were only 2,054 tokens assembled in all. With the average length of a KW item at 1.69 graphic words, the tokens taken together established 1.86% of all the corpus words (i.e., 2,786 graphic words). Hence, on average, each KW type was matched to 19.01 tokens (direct iterations). Since the mean number of KW items per KW set was 5.4, each RA featured approximately 102.7 KW tokens in the body. Naturally, these being merely the mean figures, individual RAs varied in the frequency of KW tokens, the top frequencies being 102 and 95 tokens /KW type. Conversely, 12 KWs were found not to yield any hits at all, and 9 KWs produced only a single hit each. One can therefore posit a scale ranging from high-frequency KWs all the way to low-incidence ones. Presumably, the former end of the scale features indexical/identifying KWs (having a clear impact on both coherence and cohesion (texture) of RAs), whereas the latter end embraces framing KWs (which indisputably enhance the coherent reception of the RA, although owing to their scarcity they cannot but have little impact on its texture).

As was mentioned above, all the RAs scrutinized displayed seven distinct sections/subgenres in the body of their texts. In quantitative terms, the most sizeable subgenres were found to be Results (constituting 34.79% of the data), followed by Introduction (29.98%) and Discussion (17.83%), i.e., altogether making up 82.6% of the RAs in question. The remaining 17.4% of the corpus was composed of four relatively shorter RA subgenres, namely Title, Abstract, Method/ology, and Conclusion.

Notwithstanding these quantitative considerations, the range, density and frequency of KW tokens in sections varied considerably. Generally speaking, the greatest variety (range) of KW items (as types) was detected in Introductions (4.5 KWs), Discussions (3.4 KWs), and Abstracts (3.15 KWs). In fact, Introductions were prone to

trigger almost the whole range of KW items in the set. Moreover, they represented the subgenre with the top frequency of KW tokens (accounting for 38.02% of all token findings), followed by Discussions with over a fifth (21.38%), and Results which comprised less than a fifth of all (19.81%). It follows that the relatively longest subgenres also displayed the greatest share of the findings. In contrast, the highest KW density (here understood as the number of KW tokens in graphic word terms/section words) was detected in the shortest parts/subgenres. Indeed, the KW tokens put together constituted 9.8% of (graphic) words composing the Titles; 3.15% of those formulating the Abstracts; and 2.23% of the Conclusions.

Seldom did tokens of all KW items in a KW set appear in a single RA section. In fact, this was true of only 7.86% of all RA sections, especially Introductions. Not surprisingly, the top-frequency KWs tended to be spread across many RA sections. Tellingly, of the twelve KW items employed at least once in all the sections of the respective RAs (including the Title), eleven ranked among the first two most frequent KW items in the particular RAs. Conversely, 15 KW items, usually low-frequency ones (whose incidence ranged between 1–8 tokens), were mostly confined to one section of the particular RA only.

In the Broad Corpus a noticeable tendency emerged, with one or two dominant KWs in the KW set, frequently matched with a hit already in the Title, exhibiting considerable frequency(ies). As a rule, those remaining were much less crucial statistically, their figures revealing a proportional decline. The following data may illustrate the point:

RA (1): 5 KWs; respective number of hits 45, 22, 3, 1, 1;
RA (6): 4 KWs; respective number of hits: 75, 30, 22, 21;
RA (14): 6 KWS; respective number of hits: 71, 45, 15, 11, 8, 3;

It is noteworthy that the underlined KWs had a token already in the RA title. Such remarkable frequencies may point to the prominent status of the particular KWs in the coherent reception of the RA. Furthermore, they may also suggest affinity with the top layer of the content aspect of the Global Theme (for more, see Pípalová 2008: 99–112; 2017).

3.2 NARROW CORPUS RESULTS

3.2.1 *RESULTS OF QUANTITATIVE ANALYSIS*

To reiterate, the Narrow Corpus embraced two RAs, amounting to 13,426 words (without the KW sets). Their KW sets comprised 6 KW items each and in the bodies of the RAs grouped together they were matched to 245 KW tokens (direct iterations). Hence the tokens detected (378 in graphic word terms) accounted for 2.81% of the words employed in the Narrow Corpus. However, when the analysis was extended to take account also of the tokens of KW clusters, 669 tokens were traced, corresponding to 964 graphic words, i.e., 7.18% of the data. It follows that the inclusion of KW clusters yielded nearly three times higher findings compared to direct iterations only.

When the AntConc Word list function was employed, an interesting finding surfaced, namely that KW sets are not always composed of the top-frequency content items in the data. More specifically, RA No. 7 contained 6 KWs, featuring altogether 10 different (graphic) words. The results reveal a discrepancy, since certain KW items are missing from the first 10 most frequent content words (*bilingualism, development, lifespan*), while some top-frequency content words are not integrated in the KW set (*English, social*). In RA No. 17 the chasm is even more glaring. The KW set exhibits 6 items, featuring 9 different (graphic) words. Some of the nine top-frequency content words in the RA are missing in the KW set (*information, patron, client, librarian*), whereas others featured in the KW set do not establish the top-frequency ones in the RA (*hyperlinking, encounters, conversational, analysis, institutional, interaction, multitasking*). Hence KWs do not seem to coincide with the top-frequency content units in the discourse. At this point it may be of interest to remark that the direct iterations of the same number of the top-frequency content words in the two RAs would constitute 9.02% of all the graphic words in the Narrow Corpus (i.e., a three times higher result than was yielded for the direct iterations of the KWs).

With regard to RA No. 7, it contained 7,971 words in all and its 6 KWs were found to have 173 direct iterations, comprising 287 graphic words, i.e., 3.6% of the RA words. However, when the research was extended to integrate other related tokens in KW clusters, 360 tokens of KW clusters were identified, featuring 589 graphic words, i.e., constituting 7.39% of the RA words. This means that the extension of the research yielded approximately twice as many tokens in the RA. Tellingly, the respective numbers of the hits of these KW items are as follows: 83, 73, 56, 51, 46, (+43), 6.

RA No. 17, on the other hand, turned out to be shorter, amounting to 5,718 words altogether. Its 6 KWs were matched to 72 direct iterations, corresponding to 91 graphic words, i.e., 1.59% of the graphic words in the RA. After the research was extended to include other items in the KW clusters, 309 tokens of KW clusters were traced (375 graphic words), constituting 6.56% of the RA words. In this RA, the integration of KW cluster tokens in the analysis increased the findings approximately four times. With regard to the number of hits, the RA perfectly epitomises the tendency reported above, since there is a conspicuous decline in incidence after the top frequency KWs: 181, 60, 28, 25, 13, 2.

3.2.2 *RESULTS OF QUALITATIVE ANALYSIS*

3.2.2.1 RA NO. 17

The six KWs in the set of RA No. 17 were as follows: *hyperlinking; service encounters; chat; conversation analysis; institution interaction; multitasking.*

In terms of the range of items forming a KW cluster, the KWs varied immensely. There were some which produced a rich cluster of tokens, where the related and derived items clearly outnumbered direct iterations (*hyperlinking*). There were others where the clusters were more constrained, and still more which did not form any clusters at all (*multitasking*). More specifically, when the direct iteration analysis was widened to incorporate KW clusters, the first KW *hyperlinking* was matched to a striking range of tokens: (*hyperlinking, hyperlink, hyperlinks, hyperlinked, linking, links, link, linked, 0,*

them, they, their, there, it, its, which). A somewhat more restricted range corresponded to the KW *chat*: (*chat, chats, chat-based, chatting, which*) and *conversation analysis*: (*Conversation analysis, CA, conversation analytical, it*); etc. However, *multitasking* did not establish any KW cluster, being only iterated directly, producing a mere two tokens.

Looked at from a different standpoint, some KW items had a very rich cluster of realizations (*hyperlinking*), where the direct iteration (9) was clearly outnumbered by derived and implicit realizations (150). Some KW items tended, primarily, to recur directly (*chat*: 42 instances, as against 18 derived and implicit ones). Other KWs recurred only directly (*multitasking*). Interestingly, single-word KWs tended to form much richer KW clusters (243) than NP ones (66). Presumably the single-word KW items turned out to be more flexible in actual use.

3.2.2.2 RA NO. 7

When the findings drawn from RA No. 7 were taken into account, the 6 KWs were as follows: *acculturation*, *bilingualism*, *communicative function*; *cognitive function*; *inner speech*; *language development across lifespan.* In this RA a unique phenomenon was observed, namely three KWs were conceptually interrelated, were often dealt with jointly in the body of the RA, and regularly formed a group (*communicative function*; *cognitive function*; *inner speech*). Furthermore, the text even featured their umbrella term (i.e., the text-specific hyperonym — *language functions*). Nevertheless, of these three, *inner speech* was given by far the greatest attention, which is attested to by the statistics. Due to this peculiarity, the frequencies of KW cluster tokens in RA No. 7 seem to be more levelled.

With regard to the implicit/grammatical realizations of KWs (cf. Halliday and Hasan 1985:75) in the Narrow Corpus, they were found to be generally rather rare, which appears to be in line with the relative formality and explicitness of academic discourse. In fact, two distinct classes could be distinguished. On the one hand, there were grammatical realizations with lexical involvement (definite articles, demonstratives), which accounted for 22.42% of the tokens. On the other, there were grammatical tokens without lexical involvement (pronominals, ellipsis), which proved to be generally very scarce, amounting to only 7.63% of all. This would seem to corroborate the affinity between KWs and their explicit realizations.

Admittedly, there may be some measure of subjectivity involved in the delimitations of the KW cluster boundaries. Occasionally, the researcher found herself in two minds concerning which items were to be integrated in the cluster and which seemed to be associated with the respective KW only rather loosely and hence could be disregarded. Indeed, some cases represented rather moot points. For example, *bilingualism*, with no direct iterations, was matched to a variety of related items in its cluster (*bilingual, bilinguals, ø, they, their, themselves, who*). Nonetheless, the delimitation of the actual bounds of this cluster turned out to be challenging, for the cluster proved to be fuzzy. The question arose whether such items as *Participants, Respondents*, etc., characterized as bilingual individuals in the RA, were to be embraced in the cluster as well. The decision was made to exclude the items from the respective KW cluster, since they were held to activate rather the "Research" cognitive frame. It follows, then, that in some KW clusters the existence of peripheral zones is inevitable.

3.3 RESULTS OF FSP ANALYSIS

3.3.1 *OVERVIEW OF PATTERNS*

As mentioned above, in this paper, multiple parallel analyses were performed at diverse levels in the hierarchy, namely at MC (CF0) level, SC (CF1) level and PH (CF2) level. Since sometimes the results at all three levels were aligned within thematic or rhematic functions, they posed no difficulty for the researcher and called for no further elaboration. For example:

(1) *Hyperlinks* ***(T — — T)*** *as* proposed responses *appeared frequently in both datasets when an element of uncertainty was expressed by the client.* (No. 17)
(2) *This study also provides empirical evidence that the intention to remain in the host country indefinitely is linked to higher level of L2 internalization in inner speech* ***(R — R — R)*** *and cognitive function* ***(R — R — R)***. (No. 7)

On occasion, however, the analyses conducted at the three levels produced disparate results. Naturally, such specimens, referred to here as "mixed category" items, deserved more space. They were puzzling, since the KW tokens in question were constructed as thematic at some level(s) of analysis, and as rhematic at an/other(s). What follows is an overview of all such patterns derived from the data.

Overview of mixed category cases (with th and rh functions at distinct levels of analysis) detected in the corpus

Pattern A1: T — — R
Pattern A2: T — R — R
Pattern A3: T — R — —
Pattern B1: R — T — —
Pattern B2: R — T — R
Pattern B3: R — T — T
Pattern C1: R — — T
Pattern C2: R — R — T

3.3.2 *FSP PATTERN ANALYSIS AND EXEMPLIFICATION*

In this section the results of the FSP analysis conducted at the three levels will be briefly discussed and illustrated by an example drawn from the data. Note: whenever a level yields no results (i.e., there is no pertinent subordinate clause or phrase), the pattern features a double dash to reflect this.

Mixed category cases: Group A

Pattern A1: T — — R
At CF0 level, the token is thematic (T0), there is no subordinate clause, but at CF2 level, the token bears the rhematic function (R2). In other words, the KW represents the dynamic part (R2) of the NP realizing complex theme in CF0

(3) *Language use in inner speech score was obtained by calculating the mean of the individual scores.* (No. 7)

Pattern A2: T — R — R
Part of complex theme (T0) in CF0, embracing a subordinate clause; at CF1 level, the item in question bears the rhematic function (R1); more specifically, it represents the dynamic part (R2) of the NP realizing R1

(4) *The degree of L2 use following migration, which reflects the extent of L2 internalization in inner speech and cognitive function, was found to be tightly linked to the degree …* (No. 7).

Pattern A3: T — R — —
Part of complex theme (T0) in CF0, embracing a subordinate clause; at CF1 level, the item in question bears the simple rhematic function (R1);

(5) *However, when the professional suggested a link, some clients denied that the requested information was available there.* (No. 17)

Mixed category cases: Group B

Pattern B1: R — T — —
Submerged theme, i.e., in CF0 part of rheme (R0); more specifically, at CF1 (subordinate clause) level, simple theme (T1)

(6) *We noted that hyperlinks pointed to a variety of resources.* (No. 17)

Pattern B2: R — T — T
Submerged theme, i.e., in CF0 part of rheme (R0); more specifically, at CF1 level thematic (T1); in CF2 it takes the thematic function (T2) within the complex NP realization of T1

(7) *The findings show that sequential bilinguals are capable of developing inner voice.* (No. 7).

Pattern B3: R — T — R
Submerged theme, i.e., in CF0 part of rheme (R0); more specifically, at CF1 level thematic (T1); within CF2, the item in question represents the dynamic (R2) part of the NP realization of T1

(8) *Emic results revealed that highly acculturated bilinguals associated internal translation from…* (No. 17)

Mixed category cases: Group C

Pattern C1: R — — T
Part of rheme (R0) in CF0; no pertinent subordinate clause; the rheme (R0) is realized by a complex phrase (CF2) in which the item in question takes a thematic function (T2)

(9) *Respondents consisted of 149 highly educated L2-competent Polish—English bilinguals who relocated to the UK in young adulthood...* (No. 7)

Pattern C2: R — R — T
Part of rheme (R0) in CF0, elaborated by a subordinate clause (CF1) in which the item in question takes the rhematic function (R1); in the phrase (CF2) realizing the rheme (R1) the KW token carries the thematic function (T2)

(10) *Excerpt 7 illustrates the lengthy Type 2 collaborative navigation episodes that can occur in the library chats.* (No. 17)

Hence the degrees of dynamism increase from A patterns via B patterns to C patterns. Indeed, at CF0, A patterns feature thematic functions, while both patterns B and C display rhematic functions. However, B patterns differ from C patterns at CF1 level, where the former convey thematic functions and are therefore less dynamic. As a result, dynamicity is most pronounced at CF0, but its impact is less prominent at CF1 level, and much less crucial at CF2 level.

3.3.3 *FSP CORPUS FINDINGS*

Applying this procedure to the data made it possible to uncover the delicate tendencies governing the FSP functions taken by the individual KW cluster tokens at diverse levels of the analysis. The following tables summarize the findings gained from the two RAs composing the Narrow Corpus.

KW cluster (RA No. 7)	TH only %	Mixed %	RH only %	Total %
acculturation	2.50	6.39	14.17	23.06%
bilingualism	3.34	6.95	5.27	15.56%
communicative function	1.67	1.94	9.17	12.78%
cognitive function	1.11	1.94	11.67	14.72%
inner speech	1.39	4.72	14.17	20.28%
language development across lifespan	0.83	0	0.83	1.67%
(*language functions*)	0.83	1.67	9.44	11.94%
Total KW clusters	10.83%	24.72%	64.45%	100%

TABLE 1. FSP functions (at three distinct levels of analysis, i.e., in a hierarchy of CFs) carried by individual KW cluster tokens in RA No. 7.

KW cluster (RA No. 17)	TH only %	Mixed %	RH only %	Total %
hyperlinking	17.79	24.27	16.50	58.56%
chat	0.65	8.74	10.03	19.43%
service encounters	1.95	2.26	3.88	8.09%
institutional interaction	2.27	1.62	5.18	9.06%
conversation analysis	1.29	1.62	1.29	4.21%
multitasking	0.32	0	0.32	0.64%
Total KW clusters	24.27%	38.51%	37.22%	100%

TABLE 2. FSP functions (at three distinct levels of analysis, i.e., in a hierarchy of CFs) carried by individual KW cluster tokens in RA No. 17.

The tables show that thematicity of KW cluster tokens attested to at all levels of the analysis turned out to be generally rather rare. It certainly represented the least likely option in both the RAs under scrutiny. Most of the findings demonstrate pronounced degrees of dynamicity at diverse levels of the FSP analysis. However, as stems from the results, the two RAs adopted somewhat distinct styles. Generally speaking, RA No. 7 employed the KW tokens in less dynamic functions than did RA No. 17. Furthermore, on closer inspection, one can distinguish fine grades of dynamism differentiating the individual KWs in the KW set, since some KWs tend to be less dynamic than others (presumably epitomizing the content core/the top layer of the Global theme, see Pípalová 2008: 99–112; 2017). For example, as is clear from the figures, in RA No. 17 *hyperlinking* turned out to be much less dynamic than *chat*.

Irrespective of all the differences detected, the tendency to employ KW tokens in relatively dynamic functions holds, since the Mixed category or Rhematic functions account for the decisive proportion of findings.

In order to find out whether thematicity grew with the flow of the RAs, all the KW cluster tokens in the particular sections were analyzed at all the diverse FSP levels and the outcomes are summarized in the following tables. It should be borne in mind that the subgenres differed in length and frequency of KW tokens, as follows from the Total figures.

KW cluster (RA No. 7)	TH only %	Mixed %	RH only %	Total %
Title	0	0	0.55	0.55%
Abstract	0.27	0.84	4.17	5.28%
Introduction	9.17	7.50	11.94	28.61%
Methodology	0.28	3.89	4.72	8.89%
Results	0	2.77	13.33	16.11%
Discussion	1.94	6.67	21.94	30.56%
Conclusion	0	1.94	8.05	10%
Total KW clusters	11.67%	23.61%	64.72%	100%

TABLE 3. FSP functions (at three distinct levels of analysis, i.e., in a hierarchy of CFs) carried by all KW cluster tokens in individual sections of RA No. 7.

KW cluster (RA No. 17)	TH only %	Mixed %	RH only %	Total
Title	0.32	0	0.97	1.29%
Abstract	1.94	2.59	3.56	8.09%
Introduction	6.15	11.33	11.97	29.45%
Methodology	2.27	4.21	3.88	10.36%
Results	9.71	12.62	9.39	31.72%
Discussion	3.56	6.47	3.56	13.59%
Conclusion	0.32	1.29	3.89	5.50%
Total KW clusters	24.27%	38.51%	37.22%	100%

TABLE 4. FSP functions (at three distinct levels of analysis, i.e., in a hierarchy of CFs) carried by all KW cluster tokens in individual sections of RA No. 17.

These tables indicate that despite the growing activation of the KWs across the RA sections, the KW tokens do not tend to carry more thematic functions in later RA sections.

In order to uncover the delicate tendencies still more accurately, it was deemed worthwhile to take a close look at the mixed category cases, constituting 23.61% of RA No. 7 and 38.51% of RA No. 17. Obviously, A-patterns, thematic at CFo are less dynamic than B and C Patterns, rhematic at CFo. B-patterns are in turn less dynamic than C-patterns at CF1 level.

KW clusters (RA No. 7)	A patterns %	B patterns %	C patterns %	Total mixed
Title	0	0	0	0%
Abstract	0	2.35	1.18	3.53%
Introduction	11.76	8.24	11.76	31.76%
Methodology	9.41	5.88	1.18	16.47%
Results	7.06	4.7	0	11.76%
Discussion	16.47	5.88	5.88	28.24%
Conclusion	0	7.06	1.18	8.24%
Total KW clusters	44.71%	34.11%	21.18%	100%

TABLE 5. Distribution of KW cluster tokens in mixed category FSP patterns across individual sections of RA No. 7 (23.61% altogether).

KW clusters (RA No.17)	A patterns %	B patterns %	C patterns %	Total mixed
Title	0	0	0	0%
Abstract	0.84	4.2	1.68	6.72%
Introduction	8.41	11.76	9.24	29.4%
Methodology	6.72	1.68	2.52	10.92%
Results	2.52	11.76	18.50	32.78%
Discussion	6.72	5.88	4.2	16.80%
Conclusion	0.84	1.68	0.84	3.36%
Total KW clusters	26.05%	36.98%	36.98%	100%

TABLE 6. Distribution of KW cluster tokens in mixed category FSP patterns across individual sections of RA No. 17 (38.51% altogether).

All in all, the KWs scrutinized were shown to exhibit a considerable degree of dynamicity which was not found to decrease with the flow of the RA. Hence the findings do not confirm the hypotheses posited at the outset of this study.

4. DISCUSSION AND CONCLUSION

4.1 BROAD CORPUS CONCLUSIONS

This paper investigated the Broad and the Narrow Corpora of RAs recently published in two renowned journals. In the Broad Corpus, subjected to quantitative scrutiny, direct iteration of KWs was found to have only a marginal share in the total wording of the RAs (2.32% in graphic word terms). KW items were shown to vary in frequency, density, and range across the individual RA sections. Not surprisingly, the highest frequencies were obtained in the longer RA subgenres (Introduction, Results, Discussion). The highest range of KW types was located in the shorter sections of the RAs (Title, Abstract, Conclusion, Methodology). As for the construction of KW types in KW sets, NP pattern was preferred to syntactic noun realizations. Conversely, when it came to KW tokens (direct iterations) in the bodies of RAs, it was the N pattern that prevailed.

The research also revealed uneven frequencies of individual KW items in RAs. A conspicuous tendency was uncovered where, as a rule, one or two of the statistically most prominent KW items, frequently featured in the Title itself, achieved a dominant share among KW cluster tokens. In contrast, there were some KW types which had no, or only one, token in all the RA.

4.2 NARROW CORPUS CONCLUSIONS

The findings based on the Narrow Corpus demonstrate that KWs do not coincide with the top-frequency content words in the RA. Direct iterations of KW tokens were found to constitute less than 3% of the graphic words. However, when tokens of KW clusters were taken into account, this generally raised their share to approximately 7% of the Narrow Corpus. Nevertheless, clustering of KWs was found to be very disproportionate. Some KWs produced very rich clusters (which often boosted their relative weight among the other items in the KW set), while others failed to form clusters at all, being only directly iterated. Hence tokens of KW clusters may, but need not, outnumber direct iterations. As was pointed out, there may also be a degree of subjective assessment involved in delimiting the boundaries of some fuzzy KW clusters. Some KW items in KW sets achieve the highest frequencies in the body of the RAs, which presumably testifies to their centrality for RA cohesion as well as for the perception of its coherence. It would be worthwhile to examine the affinity between such dominant KWs and the top layers of the content aspect of the Global Theme (see Pípalová 2008: 99–112; 2017). Conversely, some KWs with few tokens may serve solely the framing function. Rather than partaking in texture, they facilitated and enhanced a coherent reception of the RA.

4.3 FSP ANALYSIS CONCLUSIONS

Together with article titles, KWs were here treated as identifications of the Global Theme of the RA (for more, see Pípalová 2017). The research showed that largely rhe-

matic or mixed (T/R) results prevailed both when examining individual KW clusters and when considering all KW tokens in a RA section. Most surprisingly, perhaps, in this research, neither of the hypotheses was confirmed:

> The KW tokens examined were not found to be employed primarily in thematic functions at MC, SC, PH levels.
> KW tokens were not found to grow in thematicity across the RA (despite their activation gradually growing).

4.4 TENTATIVE INTERPRETATION

Presumably, a number of reasons might be adduced to account for the pronounced degree of dynamicity discovered. Firstly, lexical, i.e., fuller, more explicit and prosodically heavier realizations were convincingly shown to prevail. Conversely, implicit, grammatical items, especially of a pronominal or elliptical nature, were confirmed as marginal. Even though academic discourse generally tends to be lexically dense, this striking tendency cannot but confirm the prominent status of the KWs, since in the overwhelming majority of cases they were prone to be expressed rather fully. Hence an affinity arises between this tendency and end-weight and end-focus principles (cf. e.g., Greenbaum and Quirk 1990: 397–400).

Secondly, although KWs are crucial and indispensable for the coherent reception of academic discourse, their tokens proved to be relatively sparse in the data. Thus, rhematic functions, at least at some level(s) of the analysis, may represent a way of underscoring their communicative impact and upgrading their status in the hierarchy of themes. It appears then that their marginal statistical share was counterbalanced and cognitively boosted by their somewhat more dynamic standing at the various FSP levels examined, on account of which they lent themselves more conspicuously to the attention of the recipient.

Thirdly, owing to the fact that the RAs displayed between 4–6 KWs in their KW sets, in cases where tokens of several KWs showed in the same excerpt, as a rule only some were thematized:

(11) *Hyperlinking* ***(T — —)*** *is deeply entangled with service encounters* ***(R — —)*** *as a form of social interaction taking place online.* (No. 17)

However, this was not always the case. In the following example, all the KW tokens bear rhematic functions:

(12) *We first review the literature on hyperlinking* ***(R — — R)*** *and service encounters* ***(R — — R),*** *describe our use of CA* ***(R — — R)*** *methods to analyze our data, illustrate the three types of hyperlinks* ***(R — — T)*** *identified in chats* ***(R — — R)*** *(…RP)* (No. 17)

Fourthly, academic discourse, aiming at a precise, multiaspectual and comprehensive account of the subject matter, inevitably opts for formulation accuracy and complexity. Hence it was common to find the KWs themselves employed in phrases, where, as

a rule, they were used as more dynamic modifiers of general academic expressions, usually abstract ones:

(13) *The overall goal of CA **(T — — R)** is to examine recordings of naturally occurring conversations in order to…(RP)* (No. 17)
(14) *Felix-Brasdefer (2015) identified the following elements of the sequential structure of service encounters **(R — — R)**:(…RP)*
(15) *The results showed that acculturation **(R — T — R)** level had a significant effect on the extent of L2 use in inner speech, **(R — R — R)** cognitive **(R — R — R)** and communicative functions. **(R — R — R)*** (No. 7)

Fifthly, since academic discourse strives to communicate highly complex, intellectually demanding and finely graded content, structural complexity represents a norm. Moreover, hypotaxis prevails over parataxis, allowing for hierarchizing of the knowledge in a precise and subtle way. Of the variety of the subordinate clauses, the highly intellectual and frequently abstract subject matter calls for particularly abundant use of nominal content clauses, through which the actual content, embodied also by the KWs, is often communicated, see, e.g., Examples 6, 7, 8 and 15 above.

It should be noted, however, that nominal content clauses are special in that they in fact embody a mismatch between form and function: their subordinate status contrasts sharply with the crucial content they deliver. This fact alone may appear to relativize to some extent or even blur some differences between patterns. Hence, providing such structures are employed, superficially the differences between patterns, e.g., A and B, may look negligible, e.g. (Example 6) *We noted that hyperlinks pointed to a variety of resources.* (No.17) and its semiparaphrase: *In our analysis, hyperlinks pointed…* However, on balance, the verb-inclusive clause realization (*we noted*) is fuller, explicit and more foregrounded, which allows the producer to identify accurately the type of act (see Hyland 2000: 20–40) as a Discourse act, whereas in the semiparaphrase (*in our analysis*) this becomes a Research act, backgrounded by nominalization. Tellingly, rather than **In our note, hyperlinks pointed…,* we may use *Our note that hyperlinks pointed…*, although solely as part of a complex To. Besides, whereas the original (*We noted that hyperlinks pointed…*) seems to serve reminding (intratextuality) functions, the semiparaphrase (*Our note that…*) may presumably point to different ends (e.g., commenting, evaluating). It is perhaps needless to add that the use of *note* would again background the respective act. Hence despite the fuzzy area of nominal content clauses it appears that the actual pattern choice does indeed reflect the producer's particular, finely graded communicative needs.

Furthermore, it is well recognized that complex and sophisticated academic content in RAs is seldom just simply presented. Rather, it is constantly negotiated with the recipient, interpreted, assessed, framed, contextualized within the existing research, etc., all of which adds to the complexity of academic formulation. One layer of the resources implementing such functions represents academic metadiscourse (e.g., Hyland 2017: 20) "a heterogeneous array of features which assist readers not only to connect and organize material but also to interpret it in a way preferred by the writer and with regard to the understandings and values of a particular discourse community."

Indeed, at MC level, many thematic functions (To) were not undertaken by the KW tokens themselves, but were often employed to denote the author, other researchers, the present RA, its parts, elements of the research frame (the findings, etc.), abstract rhetors and the like, or were made available for other purposes, such as suggesting intertextuality, framing content, reporting other discourse or evaluating it, see, e.g. Example 15 and a variety of others:

(16) *Grosjean (2010) points out that in the case of bilinguals* ***(R — T — R)*** *the first language may not necessarily be the stronger or dominant language…* (No. 7)
(17) *This study also adds an acculturation* ***(R — — R)*** *perspective to Dewaele (2015a, 2006b) by showing that inner speech* ***(R — T— —)*** *does contain the highest ratio of L1 use in comparison to cognitive* ***(R — R— —)*** *and communicative functions* ***(R — R— —)*** *even in highly acculturated* ***(R — R—*** *R****)*** *bilinguals****. (R — R — T)*** (No. 7)
(18) *Excerpt 7 illustrates the lengthy Type 2 collaborative navigation episodes that can occur in the library chats****. (R — R— T)*** (No. 17)
(19) *This negotiation may even lead to Type 3 hyperlink* ***(R — T)*** *(a supplemental resource)*. (No. 17)
(20) *Both the datasets come from web-based chat* ***(R — R)*** *services which connect (…RP)* (No. 17)
(21) *This implies the link* ***(R — T — —)*** *is not only a proposal to the client, but also an instant resource for the professional.* (No. 17)

Since numerous thematic functions at MC level served such needs, the actual content, epitomized by the KWs, had to be assigned more dynamic functions, frequently resulting in submerged theme cases:

(22) *We found that hyperlinks* ***(R —T — —)*** *are embedded in and work as responses to information or advice requests.* (No. 17)

Conversely, there are also some good grounds for employing KWs in a far less dynamic capacity. For example, irrespective of whether at CF0 or CF1 thematic or rhematic, in some specimens the KW item expresses a core notion whose diverse aspects, types, and categories are gradually discussed. Hence at CF2 level, such KW tokens adopt thematic functions in the respective NP.

(23) *Type 2 hyperlinks* ***(T — — T)*** *were the most common in the counselling chats.* ***(R— — T)*** (No. 17)

Furthermore, the relatively remarkable degrees of thematicity attested to in some earlier subgenres, especially in Introductions, rather than in later sections in the RAs, follow, among other concerns, from the frequent need to provide definitions of crucial concepts initially. Moreover, such paragraphs tend to be organized paradigmatically.

(24) *Inner speech* ***(T — —)*** *is understood as private language use which fulfils needs other than social and interpersonal communication…(RP).* (No. 7)
(25) *Language development across the lifespan* ***(T — —)*** *is a growing area of research which focusses…(RP)* (No. 7)

Thus, it appears that each discourse type category may have its own unique FSP tendencies. Hence the various degrees of thematicity and rhematicity assigned to individual KW tokens at different levels, especially in the mixed category patterns, reflect the delicate shades of the author's communicative needs, which strive to strike the right balance between ensuring clarity and facilitating felicitous negotiation of meaning in the interest of a coherent reception of the discourse. Diverse frequencies, distributions, degrees of dynamicity and FSP patterns detected for each KW cluster seem to be indicative of the cluster's role and its relative standing in encoding the Global Theme. Naturally, since for reasons of manageability the data explored were rather limited, far more research would be required to verify the tendencies suggested.

SYMBOLS AND ABBREVIATIONS

AN	Head noun preceded by a modifier
Ant Conc	https://www.laurenceanthony.net/software/antconc/
AAN	Head noun preceded by two or more modifiers
CF	Communicative Field
FSP	Functional Sentence Perspective
KW	Keyword
MC	Main clause
NP	Noun phrase
N	Noun
PH	Phrase
RA	Research Article
R	Rheme
RP	Renata Pípalová
SC	Subordinate clause
T	Theme

REFERENCES

Čech, R. (2016) *Tematická koncentrace textu v češtině*. Praha: Ústav formální a aplikované lingvistiky UK.

Daneš, F. (ed.) (1974) *Papers on Functional Sentence Perspective*. Praha: Academia.

Daneš, F. (1989) Functional sentence perspective and text connectedness. In: Conte, M.-E., J.S. Petöfi and E. Sözer (eds) *Text and Discourse Connectedness*, 23–31. Amsterdam / Philadelphia: John Benjamins Publishing.

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

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

Ghadessy, M. (ed.) (1995) *Thematic Development in English Texts.* London and New York: Pinter.

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

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

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

Hyland, K. (2006) *English for Academic Purposes. An Advance Resource Book*. London and New York: Routledge.

Hyland, K. (2009) *Academic Discourse: English in a Global Context.* London and New York: Continuum.

Hyland, K. (2017) Metadiscourse: what is it and where is it going? *Journal of Pragmatics 113*, 16–29.

Pípalová R. (2008) *Thematic Organization of Paragraphs and Higher Text Units.* Praha: UK PedF.

Pípalová R. (2017) Encoding the global theme in research articles: syntactic and FSP parameters of academic titles and keyword sets. *Acta Universitatis Carolinae, Philologica 1/ Prague Studies in English*, 91–113.

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]. Praha: Státní pedagogické nakladatelství.

Todd, R.W. (2011) Analysing discourse topics and topic keywords. *Semiotica*, Vol. 2011, 251–270.

CORPORA

Broad Corpus

1. Al-Gahtani, S. and C. Roever (2018) Proficiency and preference organization in second language refusals. *Journal of Pragmatics* 129, 140–153.
2. Ames, K. (2018) 'Ironic detachment': Locals laughing 'at' the local on commercial breakfast radio. *Journal of Pragmatics* 123, 1–10.
3. Catedral, L. (2018) 'That's not the kind of church we are': Heteroglossic ambiguity and religious identity in contexts of LGB exclusion. *Language & Communication* 60, 108–119.
4. Chaidas, D. (2018) Legitimation strategies in the Greek paradigm: A comparative analysis of Syriza and New Democracy. *Language & Communication* 60, 136–149.
5. Dewaele, J.M. and L. Salomidou (2017) Loving a partner in a Foreign Language. *Journal of Pragmatics* 108, 116–130.
6. Duncan, D. (2017) Australian singer, American features: Performing authenticity in country music. *Language & Communication* 52, 31–44.
7. Hammer, K. (2017) Language through the Prism: Patterns of L2 Internalisation and Use in Acculturated Bilinguals. *Journal of Pragmatics* 117, 72–87.
8. Hammer, K. (2017)They speak what language to whom?! Acculturation and language use for communicative domains in bilinguals. *Language & Communication* 56, 42–54.
9. Loureiro-Rodríguez, V., M. I. Moyna, and D. Robles (2018) Hey, baby, ¿Qué Pasó?: Performing bilingual identities in Texan popular music. *Language & Communication* 60, 120–135.
10. Morrison, M. E. (2018) Beyond derivation: Creative use of noun class prefixation for both semantic and reference tracking purposes. *Journal of Pragmatics* 123, 38–56.
11. Myketiak, C., S. Concannon and P. Curzon (2017) Narrative perspective, person references, and evidentiality in clinical incident reports. *Journal of Pragmatics* 117, 139–154.

12. Ramírez-Cruz, H. (2017) No manches, güey! Service encounters in a Hispanic American intercultural communication setting. *Journal of Pragmatics* 118, 28–47.
13. Reynolds, E. (2017) Description of membership and enacting membership: Seeing-a-lift, being a team. *Journal of Pragmatics* 118, 99–119.
14. Robles, J. S., S. DiDomenico and J. Raclaw (2018) Doing being an ordinary technology and social media user. *Language & Communication* 60, 150–167.
15. Secova, M. (2017) Discourse-pragmatic variation in Paris French and London English: Insights from general extenders. *Journal of Pragmatics* 114, 1–15.
16. Shvidko, E. (2018) Writing conference feedback as moment-to-moment affiliative relationship building. *Journal of Pragmatics* 127, 20–34.
17. Stimmel, W., T. M. Paulus and D. P. Atkins (2017) "Here's the Link": Hyperlinking in Service-Focused Chat Interaction. *Journal of Pragmatics* 115, 56–67.
18. Su, D. (2017) Significance as a lens: Understanding the Mandarin *ba* construction through discourse adjacent alternation. *Journal of Pragmatics* 117, 204–230.
19. Vergis, N. (2017) The interaction of the Maxim of Quality and face concerns: An experimental approach using the vignette technique. *Journal of Pragmatics* 118, 38–50.
20. Zhao, L., N. Dehé and V. A. Murphy (2018) From pitch to purpose: The prosodic--pragmatic mapping of [I + verb] belief constructions in English and Mandarin. *Journal of Pragmatics* 123, 57–77.

Narrow Corpus

7. Hammer, K. (2017) Language through the Prism: Patterns of L2 Internalisation and Use in Acculturated Bilinguals. *Journal of Pragmatics* 117, 72–87.
17. Stimmel, W., T. M. Paulus and D. P. Atkins (2017) "Here's the Link": Hyperlinking in Service-Focused Chat Interaction. *Journal of Pragmatics* 115, 56–67.

Renata Pípalová
Department of the English Language and Literature
Faculty of Education, Charles University
Magdaleny Rettigové 8, 110 00 Praha 1
ORCID 0000-003-0171-0963
e-mail: renata.pipalova@pedf.cuni.cz