This is the peer reviewd version of the followng article:

Semantic Sequences and the Pragmatics of Medical Research Article Writing / Mazzi, Davide. - 203:(2015), pp. 353-368.

Lang
Terms of use:
The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

12/10/2024 09:30

(Article begins on next page)

DAVIDE MAZZI

Semantic Sequences and the Pragmatics of Medical Research Article Writing

1. Introduction

In today's medical profession, there is increasing recognition that doctors should develop clinical as well as research-oriented skills. In this respect, the medical discourse community demands that they become ever more active members of the discipline in both doing and writing about the research they undertake (Swales 2004). One implication of such a complex training process is that doctors are expected to familiarise themselves with the communicative requirements of medical research texts, and more precisely with their key pragmatic aspects.

This may request that doctors acquire sound knowledge of the relationship between the discourse structures of the discipline and its underlying epistemology (Dahl 2004; Engberg 2010). This is an overall challenging task in terms of language-teaching methodology and syllabus design, whereby the need for a grammar addressing fluency as well as context-based accuracy goes hand in hand with the expectation that "descriptions of language will be based on quantities of authentic data rather than on a course writer's intuitions and/or language prejudices" (Hunston/Francis 1998: 45).

It is within such an educational setting that specialised corpora (Hunston 2002; McEnery/Hardie 2011; Gabrielatos et al. 2012) deserve to be acknowledged as a source of significant insights to socialise medical students along with practitioners into the distinctive communicative practices of scientific communities. This aspect should become central when it comes to the prolific output of corpus investigation to show the recurrence of co-occurring items in text. In

that regard, the adoption of corpus-based approaches to the study of naturally-occurring language has shed light on remarkable discourse regularities, among which the tendency of words to go together and make meaning by virtue of their combination has been a favourite subject of investigation over the last two decades.

Co-occurring items have been variously termed. For instance, Sinclair (1996) talks about 'units of meaning' as longer sequences to be described in terms of collocation, colligation and semantic preference. These respectively denote firstly the regular co-occurrence of words; secondly, the co-occurrence of grammatical choices; and thirdly, "the restriction of regular co-occurrence to items which share a semantic feature" (Sinclair 2004: 142) as in the case, for instance, of an adjective co-occurring with nouns from the lexical field of sports.

Likewise, in their project aimed at a corpus-driven pedagogic grammar, Hunston/Francis (1998) look at the close association between 'verb patterns' and meaning in the 250-million-word Bank of English; Biber et al. (1999: 990) conduct a cross-register investigation of 'lexical bundles', i.e. "sequences of word forms that commonly go together in natural discourse" regardless of their idiomaticity; and Wray (2002) discusses 'formulaic expressions or sequences' as linguistic units composed of multiple words, which she analyses in the light of different frames of interpretation, e.g. individual motivations for achieving novelty and pragmatic notions of shared knowledge between speaker/writer and listener/reader. Co-occurring patterns have been variously ascribed to such widespread phenomena as Sinclair's (2004) idiom principle, Hoey's (2005) lexical priming and Goldberg's (2009) construction grammar, and they represent the primary focus of the present chapter as well.

In this vein, the aim of this study is to focus on yet another suitable candidate for the description of regularity in text, i.e. 'semantic sequences'. These are defined by Hunston (2008: 271) as "recurring sequences of words or phrases […] more usefully characterized as sequences of meaning elements rather than as formal sequences", and they are analysed here as a clue to the main aspects related to the presentation and discussion of research findings in specialised journals. In particular, semantic sequences were studied on the basis of a corpus of authentic research articles, the genre with the

lion's share in specialised medico-scientific knowledge dissemination. In Section 2, the materials on which the research is based are discussed and analytical criteria are clarified, whereas Section 3 is dedicated to the presentation of findings to be discussed in Section 4.

2. Materials and methods

The study was undertaken on a small synchronic corpus of 280 authentic medico-scientific research articles (RAs) taken from fourteen specialised journals.[1] The first strand of corpus texts (140) dates back to the period June/September 2007, and it was later supplemented with a second, more recent strand including just as many texts covering the same time span (June-September) from the year 2012. Altogether, the corpus consists of 984,304 running words.

From a methodological perspective, the study began with the computer-assisted (Scott 2009) corpus-based retrieval of the most frequent semantic sequences at two main levels. First of all, sequences were identified by starting with 'closed-class' grammar words: these "would seem to be of no interest to the discourse analyst at all" yet may be taken as a sound basis to identify patterns that define "the conventional meanings and values expressed in corpus texts" (Groom 2010: 61). Secondly, sequences were detected by starting with a grammar pattern acknowledged by the literature

1 The specialised journals were selected on the basis of chiefly exogenous criteria, i.e. the feedback provided by the staff of the Ph.D. School in Clinical and Experimental Medicine of the University of Modena and Reggio Emilia (Italy). In particular, the following publications were included: *British Journal of Dermatology* (BJD), *Journal of the American Academy of Dermatology* (JAAD), *Blood Cells, Molecules and Diseases* (BCMD), *Proceedings of the National Academy of Sciences of the United States* (PNAS), *Cancer Research* (CR), *British Journal of Haematology* (BJD), *Artificial Organs* (AO), *Proteome Science* (PS), *Clinical Chemistry* (CC), *Journal of Pharmaceutical and Biomedical Analysis* (JPBA), *Science* (SC) and *Current Opinions in Genetics and Development* (COGD).

(Halliday/Matthiessen 2004; Hunston 2008) to be both outstandingly frequent and significant in scientific writing. This is the 'N *that*' pattern, where a noun "indicates the epistemic status of the proposition" expressed in the upcoming *that*-clause (Hunston 2008: 278).

The second stage of the investigation lay in the concordance-based (Stubbs 2001) quantitative and qualitative study of both the most frequent 3/6-grams including the selected grammar words, and the most common 'N *that*' patterns highlighted in the corpus. As a result, the analysis focused on items such as *of patients with*, *in response to* and *depending on the* on the one hand, and *the possibility that*, *the hypothesis that* as well as *the finding that* on the other.

As such, the above elements were taken as a basis for the isolation of semantic sequences in context. More specifically, the collocational, colligational and semantic-preference patterns displayed by the items were a clue to the semantic sequences and the main discourse functions these were observed to perform across research-article sections ('Introduction', 'Methodology', 'Results' and 'Discussion'), from a convincing presentation of research findings to expressions of alignment of current research with a body of mainstream literature.

3. Results

The study of semantic sequences in context tends to reveal interesting patterns related to their distribution and functions across research article sections. For this reason and for the sake of clarity, the use of the most frequent sequences is illustrated with a view to the RA section they were observed to typically occur in, with §3.1 devoted to the 'Introduction', §3.2 to 'Methods', §3.3 and §3.4 dealing with 'Results' and 'Discussion' respectively.

3.1. Semantic sequences in RA Introductions

The 'Introduction' is by far one of the most extensively studied sections of RAs. This is largely due to its considerable rhetorical complexity (Swales/Feak 2012), which all too often has a two-fold cause. First of all, a remarkable variety in the steps through which its structure may be formed, with a range of different options allowing writers to situate their contribution within the broader research field either with a sense of continuity, or with a more challenging attitude pointing to the limitations of prior research. Secondly, the wealth of information for potential inclusion, from a broad review of items of previous research to more or less substantial anticipations of the contents of the rest of the paper.

In this context, the occurrence of semantic sequences in this RA section appears to be primarily linked to two inter-related phenomena. On the one hand, their use is correlated to the writer's attempt to make topic generalisations: this facet is distinctive of the first part of several scientific introductions, where authors survey their well-established research tradition through wide knowledge-claims about the state of the art of clinical practice, the implementation of diagnostic tools or laboratory techniques, to name but a few. From a discursive point of view, this function is best served by the three-word prepositional cluster *for patients with*, embedded within semantic sequences of the kind ['Current knowledge on available therapeutic options' + *for patients with* + 'Disease'] highlighted in (1) below:[2]

(1) **Currently**, chemotherapy, radiation, and surgery **are the only choices** <u>for patients with</u> **basal breast cancers**, but all show poor outcomes[(6)]. The need for a specific targeted therapy for basal breast cancer remains urgent. (CR_16)

2 In all numbered examples, underlining and bold typeface are mine. More specifically, the former is utilised to identify the node of the semantic sequence, whereas bold typeface marks the other elements that typically characterise the meaning pattern defined by the sequence. In addition, the source of each example is reported in brackets with both journal initials and an anonymised file number (e.g. CR_1 for the first corpus text from *Cancer Research*).

On the other hand, the same cluster may signal that writers are uncovering fresh and genuine evidence that will be further explored in the rest of their paper. In this case, they provide more than a simple overview of present-day knowledge; rather, in anticipating the core elements of their original contribution, they begin by shedding light on new and attractive therapeutic options clinicians are likely to benefit from in the near future. This is forcefully expressed by means of the sequence ['New treatment strategy' + *for patients with* + 'Disease'] instantiated in (2):

(2) For patients who fail initial therapy and are either refractory or relapsed, there is no standard therapy. **Novel agents** such as thalidomide, bortezomib, and lenalidomide **offer new treatment options** <u>for patients with</u> **AL amyloidosis**. (BL_14)

Taken together, the two sequences examined so far amount to 26.9% of the 63 occurrences of the expression *for patients with*, i.e. the fifth most common 3/6-word cluster retrieved for *with* at a corpus level.

3.2. Semantic sequences in 'Methods'

In his seminal work on research genres, Swales (2004) notes the somewhat decreasing rhetorical effort channelled into 'Methods' owing to the overall standardised nature of experimental procedures. This means that in a large number of published papers, researchers tend to replicate well-established methods that do not need full elucidation at the outset. Nonetheless, the fact remains that the methodological section is still required to play a key-role in the reconstruction of the relevant research setting. With the aim of providing accurate information about the criteria behind patient recruitment or data collection, the type of procedure applied – e.g. statistical or other – and the devices/facilities that make for the necessary research apparatus, scientific contributors often avail themselves of two prominent semantic sequences encompassing the top cluster of both *for* and *on*.

The first one is ['Methods/Tools' + was + *used for the* + 'Research goal'], which concerns 55.7% of the 61 entries of *used for the* and is designed to reconcile the adoption of a certain method with the key research goal this was elected to pursue (cf. 3 below). The second sequence suggests a close parallel with the first: accordingly, ['Procedure' + *on the basis of* + 'Underlying criteria'] emphasises the writer's care in relating an experimental procedure with the criterion or purpose underlying its choice. Of note, the sequence applies to 22.4% of the corpus tokens of *on the basis of*. Moreover, the first element in the pattern ('Procedure') is invariably expressed through verbs sharing a recognisable semantic preference of 'procedural requirement' – e.g. *recruit*, *assess*, *sort*, *exclude* or *select* – and employed in the passive voice, as can be seen in example (4):

(3) Two different sampling **devices** were <u>used for the</u> **dissolution experiment**. (JPBA_6)

(4) This PET imaging agent **was developed** <u>on the basis of</u> **a 14-mer peptide** identified by phage display technology.[(15)] (CR_19)

3.3. Semantic sequences in 'Results'

'Results' are by definition a most eagerly awaited section of RAs. Its interest predictably lies in the presentation of the findings of the study, which by all accounts deserves to be seen as a truly creative endeavour on the writer's part. In spite of the distinctive traits 'Results' have been observed to exhibit on a cross-disciplinary basis (Thompson 1993; Brett 1994; Williams 1999), there is convincing evidence that the section may yet display a number of repetitive phraseological patterns. In the case of medico-scientific articles, semantic sequences serve three major functions: pointing out, firstly, the salient feature(s) of the observed phenomenon; secondly, the description of cause-effect relationships in the data; and finally, an assessment of what brings about noticeable changes in the items under investigation.

To begin with, the effort to draw the readers' attention to any feature writers consider as conspicuous about the clinical data they scrutinised, is well orchestrated through two semantic sequences. Perhaps not surprisingly, they both begin with verbs sharing a clear semantic preference of 'sight' (e.g. *see*, *detect*, *observe*, *identify* and *highlight*). Still, a peculiarity of the first sequence is that it displays a marked preference for the passive voice of such verbs, whereas the second one suggests a stronger inclination towards the active form.

The first sequence, i.e. ['Key-element' + Verb of sight $_{(passive)}$ + 'Location' + *of patients with* + 'Disease'] concerns 5% of the 184 occurrences of the node phrase *of patients with*, and it points out what was detected as significant in the blood, plasma, serum or skin of patients affected by a certain illness (cf. example 5 below). Likewise, the second is [Verb of sight + *as a potential* + 'Determinant' + *for*/*of* + 'Disease']: it reveals the element isolated as the potential trigger (e.g. *biomarker for* or *risk factor of*) of a critical condition, and it applies to 29.4% of the corpus tokens of *as a potential*, as documented in (6):

(5) Because low-to-moderate avidity EphA2-specific **CD8+ T cells have been detected in the peripheral blood** <u>of patients with</u> **renal cell carcinoma or prostate carcinoma** [19, 20], levels of circulating CD8+ T cells could also be amplified by vaccination for improved immune targeting of EphA2+ tumor cells in vivo. (CR_14)

(6) In this work, five novel proteins associated with alcohol induced hepatic fibrosis were detected. **We have** specifically **identified annexin A3 and annexin A6** <u>as a potential</u> **biomarker for** predicting **alcohol-induced liver cirrhosis**. (PS_13)

Establishing cause-effect relationships in the data is the second fundamental aspect emphasised by semantic sequences in 'Results'. In this respect, the two sequences concerned involve the paradigmatically interchangeable presence of the node prepositional phrases *in the pathogenesis of* and *in the regulation of*, and their role is to show what elements are implicated in a certain biological or pathophysiological process (e.g. *cell adhesion* and *inflammatory states* in 7 and 8, respectively). To mention but two of the combinations

embedding the above phrases, ['Element' + be involved + *in the regulation of* + 'Process'] occurs in 39.5% of the hits for *in the regulation of*, whereas ['Element' + play + a critical/central/important role + *in the pathogenesis of* + 'Process'] is attested in 30.5% of *in the pathogenesis of*, as illustrated in (7) and (8) below:

(7) In the current study, three cytoskeletal proteins, vimentin (intermediate filaments), tubulin (microtubules) and actin (microfilaments) were identified with enhanced S-nitrosylation under rosuvastatin treatment. In addition, TPM, which regulates actin movement, was also found to be S-nitrosylated. **These proteins are involved** in the regulation of **cell shape, cell adhesion and migration**. (PS_20)

(8) As a prototypic damage-associated molecular pattern (DAMP) molecule, **HMGB1 plays a central role** in the pathogenesis of **many inflammatory states** released following tissue damage or injury, and is found in the serum-including cancer [38, 42] as well as other settings. (CR_11)

The last function carried out by semantic sequences in the findings section of scientific articles essentially looks like a minor variant of that documented in (7) and (8) above. It is represented by the writers' retrieval of first-hand empirical information about what changes vis-à-vis a certain type of input or determinant. In 17.5% of the entries of *in response to*, for instance, the phrase is part of the sequence ['Element' + Verb of 'change' + *in response to* + 'Input']. At the same time, 23.3% of *depending on the* can be observed to be part of the sequence ['Element' + Verb of 'change' + *depending on the* + 'Determinant']: here, a factor (e.g. *molar mass*, *molecular weight* or *the mode of injury*) is shown to explain the variation of some entity under the researchers' looking-glass, as it were. What gives both sequences their characteristic flavour is the colligation between the node and verbs sharing a typical semantic preference of 'change' – e.g. *change*, *alter*, *differentiate*, *vary*, *differ*, *degrade* – as can be appreciated in (9) and (10):

(9) **Several genes** involved in proliferation, metabolism or structure **were** also **altered** in response to **FLT3 signalling** including BCL7A, CLIC1, MRPL12, ATAD2, VIM and UBE2J1, as were genes involved in receptor-mediated signalling (IL1RAP, CDC42EP3, PLAUR, CD180 and GHRH). (BJH_3)

(10) Fig. 4 demonstrates that **the major products** formed **vary** depending on the **sample solvent**. (JPBA_19)

3.4. Semantic sequences in 'Discussion'

The 'Discussion' section is a key-moment in the process of specialised knowledge dissemination: at this stage, the specifics of experimental findings to which readers are often re-oriented in the opening segment of the section are expected to be contextualised within the broad disciplinary framework outlined in the literature review. In this vein, discussions should be a convenient place to establish the news value of research findings presented earlier on, and it is interesting that the writer's interpretation of data and consolidation of the research space occupied in the 'Introduction', very often leads to the expression of authorial stance. This can be defined as a dimension of discourse including features which refer to the ways writers present themselves and convey their opinions and commitments (Hyland 2005). As such, stance denotes the strategies through which "writers intrude to stamp their personal authority onto their arguments or step back and disguise their involvement" (McGrath/Kuteeva 2012: 163).

Following Biber et al. (1999) and Conrad/Biber (2000), it can be argued that the most frequently occurring semantic sequences in discussions are highly instrumental in expressing two kinds of stance. The first one is attitudinal stance, conveying attitudes, feelings and value judgments about the entities reported in the text. The sequences that suit this purpose are characterised by the collocation of *in patients with* and *the finding that* with evaluative markers through which writers express a markedly positive stance on the overall significance of the findings recorded in 'Results'. In [Evaluative marker + *in patients with* + 'Disease'], which concerns 7.4% of *in patients with*, evaluation may take the form of stance adverbials (cf. 11 below), nouns (e.g. *efficacy*, *interest*) or adjectives that show the benefits of a procedure, tool, therapy or predictor investigated in the current study (*effective*, *useful*). By contrast, [*the finding that* + 'Element' + Evaluative marker], applying to 14.7% of *the finding that*,

overwhelmingly involves the choice of evaluative adjectives (e.g. *encouraging*, *novel*) to promote the putative validity and genuineness of research findings, as shown in example (12):

(11) **Interestingly**, no DSRAD mutation has been found in patients with **DUH**.[29] (JAAD_6)

(12) The finding that **CCR5 antagonists block FBS-induced invasion** is **novel** and suggested that CCR5 activation contribute to the production of metastasis in vivo where different chemotactic and growth signals are present. (CR_16)

The second kind of stance is epistemic, related to the expression of degrees of certainty attached to the truth-value of the attendant proposition. On the one hand, 25.5% of the corpus hits of *the possibility that* are embedded in the sequence ['Re-statement of results' + Hedge + *the possibility that* + 'Implication']: in this co-text, the N *that* pattern co-occurs with hedges – e.g. *we cannot rule out*, *these data do not exclude* – the aim of which is to more tentatively draw the implications foreshadowed by the findings briefly summarised earlier on (cf. 13). At the top end of the epistemic scale, on the other hand, 24.4% of the occurrence of *the observation that* is part of the sequence ['Re-statement of findings' + Booster + *the observation that* + 'Back-up finding']. Here, the three-gram invariably collocates with boosters (e.g. *is consistent with*, *fits with*, *is upheld by*) that strengthen authorial claims about the convincing nature of findings or even more so, their mutual consistency. These aspects are emphasised in passages (13) and (14), respectively.

(13) **Our results**, which **suggest that rhBMP-7 protects the heart by inhibiting TGF-1-induced EndMT**, are consistent with a study showing that, in a mouse model of ischemic injury, rhBMP-7 injected into the heart stabilizes the microvasculature. [33] **We cannot exclude** the possibility that rhBMP-7, in addition to inhibiting EndMT, might also elicit anti-inflammatory effects, promote myogenesis or prevent apoptosis of endothelial cells.[34] (NM_6)

(14) **Figure 1A also shows that** although p53 is important in determining the outcome of AAV infection, **the downstream transcriptional target** of p53, p21, **is similarly important**. **This is highlighted by** the observation that

> **HCT116 p21-/- cells**, despite expressing functional p53 protein, **also undergo apoptosis** following AAV infection. (CR_8)

What passages such as (14) emphasise is that the idea of consistency in findings is highly-prized by the parent scientific community of professional writers. Corpus evidence suggests that this is more than a mere first impression, as is documented by two semantic sequences evenly distributed across 'Results' and 'Discussion', i.e. ['Finding' + Verb of 'continuity' + *our hypothesis that*] and ['Finding' + Verb of 'continuity' + *the notion that*]. These demonstrate that the solidity of findings in today's big science is frequently attributed to their consistency with either the initial hypothesis formulated in the 'Introduction', or helpful suggestions from authoritative sources in the literature.

Consistency between results and the initial working hypothesis is a regularity of *our hypothesis that*, where 58.2% of the occurrences of the pattern are inscribed in the sequence ['Finding' + Verb of 'continuity' + *our hypothesis that*]. Moreover, consistency between findings and suggestions from the relevant literature is the norm in 68.7% of the entries of *the notion that*, where the pattern is incorporated in the sequence ['Finding' + Verb of 'continuity' + *the notion that*]. The meanings conveyed by the two patterns are effectively communicated by the colligation between the N *that* patterns and verbs sharing a definite semantic preference of 'continuity', such as *confirm*, *support*, and *be consistent*/*in keeping*/*in line with* exemplified in (15) and (16) below:

(15) **These results were consistent with** <u>our hypothesis that</u> extracellular Prxs induce the IL-23−IL-17 inflammatory pathway through TLR2 and TLR4 and promote post-ischemic inflammation and infarct volume growth. (NM_14)

(16) On the other hand, similar treatment to the normal-glucose cultivated keratinocytes did not significantly alter the pSTAT-1 expression (see Fig. S3) and increased the hBD2 levels by 4·6 ± 14·7%, which is not statistically significant. **These results validated and strengthened** <u>the notion that</u> AGE formation in the high-glucose cultivated keratinocytes contributed to hBD2 reduction in our experimental conditions. (BJD_20)

4. Conclusions

The study of co-occurring items has become a prominent feature in present-day corpus and discourse studies. Whether we take Erman/ Warren's (2000) claim that at least 55% of the words in English texts are part of pre-fabricated multi-word units, or Biber et al.'s (1999) finding that 21% of words in academic prose occur in a recurrent lexical bundle, there is more than a reason to believe that co-occurrence patterns warrant proper consideration in descriptive accounts of both general and specialised language.

With reference to medico-scientific knowledge dissemination, this chapter has shed light on the wide array of meanings semantic sequences can be observed to codify across RA sections, from writers' generalisations about the state of the art of clinical practice, the implementation of diagnostic tools or laboratory techniques in introductions (cf. §3.1), to the expression of authorial stance in consolidating research findings in discussions (see §3.4). In spite of the specificity of each method used to identify semantic sequences – i.e. by starting with a grammar word or the N *that* pattern – there is remarkable continuity between the one and the other, insofar as they provide a useful link between lexis and grammar. Accordingly, evidence showed that not only do closed-class words and N *that* maintain a regularity in terms of their attested corpus use, but more interestingly, words sharing homogeneous patterns typically fall into groups based on mutually agreed, institutionalised and purposeful aspects of meaning inherent in the genre investigated here.

In this respect, data indicate a wide range of well-established, pre-constructed forms in use within the medical research community, when the draft of scholarly papers comes to the rhetorical turning points implied by a convincing presentation of research findings – whether it is accuracy in investigating cause-effect mechanisms, or consistency with the literature. As Hyland (1998) puts is, whereas individual factors may inform writers' choices in the step-by-step creation of an RA, key-meanings ultimately arise from the interaction between writers themselves and readers in specific social

circumstances. As a result, writing as a medical researcher implies being capable of talking to the expert members of the relevant discourse community in ways they find most effective, which may eventually take the form of the reiterated semantic sequences examined here.

The qualitative and quantitative findings in Section 3 suggest that the capability of corpus-based teaching of highlighting semantic sequences may provide a large amount of information about disciplinary ideologies. If, as a matter of fact, the study of individual word forms in context is well suited to the description of the subject matter of a discipline, the phraseologies at the heart of semantic sequences are indicative of its epistemology, because they are less linked to the subject matter than they are to well-established meanings and functions (Hunston 2008).

On these grounds, a sequence-based approach to word behaviour is likely to prove an essential tool for language instructors in devising awareness-raising activities based on authentic written text rather than intuition, especially at the top end of specialised-knowledge dissemination represented by high-profile RAs to be submitted to scholarly journals. With a view to practitioners' needs, sequences should indeed be seen as a firm basis for vocabulary building, so that the availability of lists of widespread sequences stands out as a resource for teachers to avert the chaos of individual word use through "an order based on meaning" (Hunston/Francis 1998: 70). Hence, the opportunity arises of including sequences such as those accounted for in the present study in a reference book providing learners with a more apt description of the language they will be expected to master, by contributing to their overall accuracy and fluency.

References

Biber, Douglas / Johansson, Stig / Leech, Geoffrey / Conrad, Susan / Finegan, Edward 1999. *Longman Grammar of Spoken and Written English*. London: Longman.

Brett, Paul 1994. A Genre Analysis of the Results Section of Sociology Articles. *English for Specific Purposes* 13, 47-60.

Conrad, Susan / Biber, Douglas 2000. Adverbial Marking of Stance in Speech and Writing. In Hunston, Susan / Thompson, Geoff (eds) *Evaluation in Text. Authorial Stance and the Construction of Discourse*. Oxford: Oxford University Press, 56-73.

Dahl, Trine 2004. Textual Metadiscourse in Research Articles: a Marker of National Culture or of Academic Culture?. *Journal of Pragmatics* 36/10, 1807-1825.

Engberg, Jan 2010. Knowledge Construction and Legal Discourse: The Interdependence of Perspective and Visibility of Characteristics. *Journal of Pragmatics* 42/1, 48-63.

Erman, Britt / Warren, Beatrice 2000. The Idiom Principle and the Open Choice Principle. *Text* 20/1, 29-62.

Gabrielatos, Costas / McEnery, Tony / Diggle, Peter J. / Baker, Paul 2012. The Peaks and Troughs of Corpus-based Contextual Analysis. *International Journal of Corpus Linguistics* 17/2, 151-175.

Goldberg, Adele 2009. The Nature of Generalization in Language. *Cognitive Linguistics* 20/1, 93-127.

Groom, Nicholas 2010. Closed-class Keywords and Corpus-driven Discourse Analysis. In Bondi, Marina / Scott, Mike (eds) *Keyness in Text*. Amsterdam: Benjamins, 59-78.

Halliday, Michael A.K. / Matthiessen, Christian 2004. *An Introduction to Functional Grammar*. London: Arnold.

Hoey, Michael 2005. *Lexical Priming: A New Theory of Words and Language*. London: Routledge.

Hunston, Susan 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Hunston, Susan 2008. Starting with the Small Words. Patterns, Lexis and Semantic Sequences. *International Journal of Corpus Linguistics* 13/3, 271-295.

Hunston, Susan / Francis, Gill 1998. Verbs Observed: A Corpus-driven Pedagogic Grammar. *Applied Linguistics* 19/1, 45-72.

Hyland, Ken 1998. Boosting, Hedging and the Negotiation of Academic Knowledge. *Text* 18/3, 349-382.

Hyland, Ken 2005. Stance and Engagement: A Model of Interaction in Academic Discourse. *Discourse Studies* 7/2, 173-192.

McEnery, Tony / Hardie, Andrew 2011. *Corpus Linguistics*. Cambridge: Cambridge University Press.

McGrath, Lisa / Kuteeva, Maria 2012. Stance and Engagement in Pure Mathematics Research Articles: Linking Discourse Features to Disciplinary Practices. *English for Specific Purposes* 31, 161-173.

Scott, Mike 2009. *WordSmith Tools 5.0*. Liverpool: Lexical Analysis Software.

Sinclair, John 1996. The Search for Units of Meaning. *Textus* 9/1, 75-106.

Sinclair, John 2004. *Trust the Text. Language, Corpus and Discourse*. London: Routledge.

Stubbs, Michael 2001. *Words and Phrases: Corpus Studies on Lexical Semantics.* Oxford: Blackwell.

Swales, John 2004. *Research Genres. Explorations and Applications*. Cambridge: Cambridge University Press.

Swales, John / Feak Christine 2012. *Academic Writing for Graduate Students. Essential Tasks and Skills*. Ann Arbor: The University of Michigan Press.

Thompson, Dorothea K. 1993. Arguing for Experimental "Facts" in Science: A Study of Research Article Results Sections in Biochemistry. *Written Communication* 10, 106-128.

Williams, Ian A. 1999. Results Sections of Medical Research Articles: Analysis of Rhetorical Categories for Pedagogical Purposes. *English for Specific Purposes* 18/4, 347-366.

Wray, Alison 2002. *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.