Elizaveta A. Smirnova, PhD in Philology,
Associate professor of Foreign languages department, NRU HSE Perm

Syntactic markers of an academic text: literature review

The ability to write a well-formed academic text is seen as key to success in educational settings across many subjects at modern universities (Wolfe, 2011). However, many L2 writers find this task difficult to accomplish. The aim of the project called “Software Development for Corpus Research in English Studies” my colleagues and I are currently involved in is to create a software tool capable to analyse and assess academic writing in English. Within the framework of the 2017 project, the approach to text analysis was developed and a pilot system was created. The software that was named Paper Cat can analyse a text against a set of academic discourse markers typical of academic writing such as abstract semantic verbs, logic connectors, intensifying adverbs, etc.

The second stage of the project aims at broadening the list of markers with syntactic ones in order to achieve a higher degree of accuracy in academic text assessment.

This work presents a review of syntactic markers of academic style which are mentioned in textbooks, manuals and articles on academic writing. I will make an attempt to systematise them and to speculate on the possibility to process them with the Gate software, i.e. use them as an extension to Paper Cat.

Based on the analysis of existing literature on the topic, I have classified syntactic markers of an academic text into four groups:

- sentence length;
- word order;
- communicative type of sentence;
- specific syntactic constructions.

A number of researchers consider sentence length to be a good indicator for measuring syntactic complexity of an academic text (see, for example, Lu 2011; Ortega 2003; Wolfe-Quintero et al. 1998), however the issue seems to be debatable. Dahme & Selfa (2017) studied sentence length in Catalan students' research reports written at high school and at university. Sentence length was calculated as the number of words per sentence, where a word was seen as any element between spaces. The authors found out that sentence length decreased from high school to university. It was explained by the fact that the university students use "compressed grammatical structures" which decrease the length of their texts (Dahme & Selfa, 2017: 9). In line with this, Wallwork (2016) in his recent manual "*English for writing research papers"* recommends learners to use shorter sentences (active rather than passive) in their texts, keeping the subject and verb close to each other without parenthical information in between and avoiding using *and, as well as, which* (Wallwork, 2016: 24-44). However, Siepmann et al. (2011) point out that academic texts can contain both short and long sentences (2011: 92), therefore denying the correlation between sentence length and 'scientificity' of a text.

The next important syntactic marker of an academic text is word order. Wallwork (2016) in his book devotes substantial attention to this issue. Thus, he suggests putting the most relevant subject at the beginning of a sentence, direct object before the indirect one, adjectives before nouns they describe (Wallwork, 2016: 22-29). Siepmann et al. (2011) also claim that "English has a strong preference for sentences beginning with the subject of the main clause" (2011: 99). They also state that in written English the most important information usually occurs towards the end of the clause (105). Swales and Feak (2004) mention inversion as a highlighting device which is used in academic texts to single out one result/fault/problem/virtue from many others, e.g. *Particularly prominent were functional strategies…* (2004: 269). Siepmann et al. (2011) call it presentative focusing on the communicative function of the construction. In sentences with

inversion the subject occupies the clause-final position, e.g. *Implicit in the theory is the presupposition that all gun owners are potentially violent* (Siepmann et al., 2011: 134).

A large body of work has focused on types of sentences used in academic discourse. For example, Swales, 1990; Teufel et al., 1999; Liakata et al., 2010 aimed at defining and annotating sentence types (called zones) in conference articles. Apart from statements which are obviously the most common type, there are questions and definitions. Siepmann et al. (2011) mention the restricted use of questions and exclamations in academic texts and the use of orders, in the form of imperative clauses, for introducing examples and inviting the reader to act, e.g. *Consider the following examples. See Butler (2003) for a fuller account. Note that we have not included passives in our analysis. Compare the analysis offered by Smits et al. (2007)* (Siepmann et al., 2011: 88). Swales and Feak (2004) believe that the only type of questions typical of an academic text is indirect questions, e.g. *Researchers are now investigating whether this recovery time can be accelerated* (2004: 133). Another type of sentence which can be seen as a marker of a scientific text is a definition. Louis & Nenkova (2012) describe them as noun phrases expressing the concept to be defined followed by a copular verb (is/are). The predicate of a definition consists of two parts: a noun phrase reporting the concept as part of a larger class (e.g. *an aqueduct is a water supply*) and a relative clause which enumerates unique properties of the concept, e.g. *An aqueduct is a water supply or navigable channel constructed to convey water* (2012: 1157 -1158).

Specific syntactic constructions that occur in academic discourse are:

- it-clefts whose pattern is It + BE + one focused constituent + that + clause, e.g. *It was only in October last year that the council decided to establish a new fund. It is precisely in cases like this that one needs to be particularly careful* (Siepmann et al., 2011: 113-114);

- pseudo-clefts whose basic formula is as follows: [Wh-…] + BE + one focused constituent, e.g. *What the government has failed to consider is the effect on old-age pensioners* (Siepmann et al., 2011: 117);

- the th-wh construction with the pattern Th- BE [Wh-…], e.g. *This is why linguistics is so important. This is where the scientist comes into his own: a true scientist is totally dispassionate* (Siepmann et al., 2011: 120);

- various types of clauses, such as:

1) the attitudinal clauses, e.g. *It is interesting to note that the only grammatical category not represented is the adverb* (Siepmann et al., 2011: 128);

2) adverbial clauses. The most common types of adverbial clause are those of time, place, reason, purpose, result, manner, contrast, condition and concession. The most common conjunctions associated with each meaning relation are given below:

time: after, before, when, until, as soon as, while

place: where, wherever

reason: because, as, since

purpose: in order that

result: so that

manner: as, as if, as though

contrast: whereas, while

condition: if, unless, on condition that, in the event that, provided (that), providing, supposing (that)

concession: although, even though, despite the fact that (Siepmann et al., 2011:149).

Among these types special attention is paid to conditional clauses (see, for example, Mazgutova & Kormoz, 2015). Thus, Warchal (2010) claims that conditional clauses, being able to perform a wide range of functions, are essential in academic writing where they are used to present logical argumentation and problem solving. Swales and Feak (2004) in their manual also mention unreal

conditionals which are used in academic texts for criticising previous research (Swales & Feak, 2004: 260);

3) relative clauses also appear frequently in academic writing (Byrnes & Sinicrope, 2008). Relative clauses are considered one of the most explicit types of noun modification, and their frequency is often used as evidence of syntactic complexity (Jucker, 1992);

4) non-finite clauses, namely participial clauses. They sometimes function as adverbial clauses with conjunctions, but the only conjunctions allowed are *when, before, after, while, since* (in the temporal sense)*, if, unless, although* and *as if*. Participial clauses tend to occur before, and not after, the main clause. Non-finite clauses are compact and their use decreases sentence length while increasing the complexity of a sentence. Compare the non-finite and finite forms in the following examples:

When considered from this angle, the difficulties appear insurmountable.

When the difficulties are considered from this angle, they appear insurmountable (Siepmann et al., 2011: 152);

5) infinitive clauses, e.g. *Universities should give students more opportunities to correct their mistakes* (Mazgutova & Kormoz, 2015: 5);

6) embedded clauses. Some scholars believe that syntactic complexity relies on prepositional phrases embedded in noun phrases, rather than on clausal subordination (Biber et al., 2011), e.g. *The conclusion I have drawn from my work is that the drug XX does work, since as demonstrated in the practical part, it stops the activity of the phosphorylated ERK protein, thus halting the process of cellular division, as I foresaw in my initial hypothesis*. Thus, another way of calculating syntactic complexity suggested by Dahme and Selfa (2017) is counting the number of words before the first finite verb (Dahme & Selfa, 2017: 10);

- non-agent subjects, e.g. *Bosworth's new book discusses the origins of obesity in both Europe and Asia*. In such sentences subjects express not people but circumstances such as time, place, purpose, reason and manner. Time and place adjuncts also frequently occur in initial position in order to provide "a spatio-

temporal framework for the information which follows, and are particularly useful when you want to organize whole stretches of text from a temporal or spatial point of view" (Siepmann et al., 2011: 122-126).

Discussion

The final issue I'm going to focus on is the possibility and rationale for detecting these syntactic markers of academic style with the help of our software.

The first syntactic marker that I have considered is sentence length. Even though it is technically possible to detect sentence length with the help of Paper Cat in the texts under study, it doesn't seem to be necessary for the assessment of students' works, as, according to the literature, this criterion cannot serve as an absolute indicator of quality of an academic text.

As for word order, since the corpora we are working with are syntactically annotated, it is possible to find sentences with inversion. However, without a preliminary corpus analysis it is difficult to say whether this marker is significant for our purposes or not. However, the other issues connected with word order, such as putting the most relevant subject at the beginning of a clause or the most important information at the end of a clause; seem to require semantic annotation which we didn't plan to provide.

Provided that Paper Cat will soon be able to perform word search, it will be possible to find different types of clauses via key words, i.e. conjunctions. It is also likely to involve a high degree of manual search because some conjunctions have different meanings and can be used in different types of clauses e.g. *since* can be used in adverbial and non-finite clauses).

References

Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly, 45*(1), 5-35

Dahme, A. M. P., & Selfa, M. (2017). Academic language in Catalan students' research reports across levels of study. *Corpus Linguistics and Linguistic Theory* (in press).

Liakata, M., Teufel, S., Siddharthan, A., & Batchelor, C. R. (2010). Corpora for the Conceptualisation and Zoning of Scientific Papers. In *Proceedings of LREC*.

Byrnes, H., & Sinicrope, C. (2008). Advancedness and the development of relativization in l2 german: a curriculum-based longitudinal study. In L. Ortega, & H. Byrnes (Eds.), *The longitudinal study of advanced L2 capacities* (pp. 109–138). NewYork: Routledge / Taylor & Francis.

Jucker, A. H. (1992). *Social stylistics: Syntactic variation in British newspapers.* NewYork: Mouton de Gruyter.

Louis, A., & Nenkova, A. (2012). A coherence model based on syntactic patterns. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning* (pp. 1157-1168). Association for Computational Linguistics.

Lu, X. 2011. A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly 45*(1). 36–62.

Mazgutova, D., & Kormos, J. (2015). Syntactic and lexical development in an intensive English for Academic Purposes programme. *Journal of Second Language Writing, 29*, 3-15.

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics 24*. 492–518.

Siepmann, D., Gallagher, J. D., Hannay, M., & Mackenzie, J. L. (2011). *Writing in English: a guide for advanced learners*. Tuebingen: UTB

Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Swales, J. M., & Feak, C. B. (2004). *Academic writing for graduate students: Essential tasks and skills* (Vol. 1). Ann Arbor, MI: University of Michigan Press.

Teufel, S., Carletta, J., & Moens, M. (1999). An annotation scheme for discourse-level argumentation in research articles. In *Proceedings of the ninth conference on European chapter of the Association for Computational Linguistics* (pp. 110-117). Association for Computational Linguistics.

Wallwork, A. (2016). *English for writing research papers.* New York: Springer.

Warchal, K. (2010). Moulding interpersonal relations through conditional clauses, consensus-building strategies in written academic discourse. *Journal of English for Academic Purposes, 9,* 140–150.

Wolfe, C. R. (2011). Argumentation across the curriculum. *Written Communication, 28*(2), 193-219.

Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity*. Honolulu, HI: University of Hawaii Press.