

Stručni članak
UDK: 81'324:811.112.2'367.3

Primljen 11.9.2011.
Prihvaćen 12.9.2012.

CONNECTIVES IN STUDENT WRITING - A LEARNER CORPUS STUDY

*Antonela Konjevod**

Filozofski fakultet u Zagrebu

This paper aims at showing the usefulness of (small) corpora for foreign language teaching. A small corpus of students' homework, 32 essays, i.e. 13,250 words, was compiled and analyzed in terms of the usage of German connecting expressions, i.e. connectives (such as *und*, *aber*, *nämlich*). The goal of the study was not just to gain insight into students' competence in connecting two stretches of discourse, but also to share with them the insights about their language use and to offer suggestions for their further language learning. Although the study has relevance only for the given group of students, it illustrates the possibilities of such student-centred discovery learning and gives teachers impulse to experiment with their own data. According to the data provided, the syllabus of a language teaching course can be modified or adapted, i.e. emphasis can be placed on certain aspects of language that have been neglected or should be practiced more in class.

Keywords: corpus-linguistic approach, small corpora, German connectives

1. INTRODUCTION

During the past two decades the corpus-linguistic approach has entered the area of foreign language teaching and opened up new

* Antonela Konjevod, Filozofski fakultet u Zagrebu, akonjevo@ffzg.hr

possibilities for the study of learner languages. The focus of researchers in corpus linguistics has shifted from a mere linguistic (descriptive) approach to a number of pedagogical ones. The usefulness of corpora for foreign language teaching was recognized after the publication of the first dictionaries and grammars that used authentic data from an English native speaker corpus (Sinclair, 1987; Quirk et al. 1985; Sinclair, 1990). The corpus material provided data about the language use as well as numerous authentic, every-day examples of language use. Foreign language learners were presented with the examples obtained from the corpus material, instead of being confronted with constructed examples (Reppen & Simpson, 2001: 106). This gave impulse to numerous empirical studies that were based on authentic corpus data and aimed at improving not only foreign language teaching but also at creating better teaching materials. To name only some of them, Granger (1997) compares participle clauses in native and non-native academic writing; Conrad (1999) illustrates corpus-based research with a study on linking adverbials in English; Hyland (2002) explored the use of directives in L2 student essays and Biber and Reppen (2002) the relationship between the grammar information presented in ESL-EFL materials and actual language use; while Fortanet (2004) studied the use of first person pronouns in academic speech.

Many other researchers demonstrated in their studies the importance of examining classroom interaction for the understanding of grammar features and tried to provide us with linguistic descriptions of classroom discourse: Biber et al. (2004) investigate the use of multi-word sequences in classroom teaching and textbooks; Csomay (2006) compares the language of American university classrooms to academic prose and face-to-face conversation; Cortes and Csomay (2007) used corpus-based methodologies to analyze recurrent word combinations, so-called lexical bundles, in university classroom talk; and Csomay (2007) compared teacher talk to student talk in American university classes.

It has been widely stressed that collecting and analyzing learner data, both native and non-native, is very useful for foreign language classes because "corpora constitute resources which, placed in the hands of teachers and learners who are aware of their potential and limits, can significantly enrich the pedagogic environment" (Granger, 2004: 136). Comparisons of native and non-native language contribute to students'

awareness of semantic nuances of words, problems in using certain structures, stylistic differences between certain structures, etc.¹ However, the application of the results of corpus-based research in the classroom still leaves a lot to be desired. First of all, many of the studies aimed at improving teaching remain at the level of lexical or lexico-grammatical description (Conrad, 1999: 2). Further, the use of native and non-native corpora in foreign language teaching is not available to everyone. The majority of teachers does not have access to corpora and lack funds to acquire and exploit large corpora. Teachers must be taught about the usefulness of certain corpora for their teaching and trained to use corpus-based approach in language teaching, or trained how to compile a small corpus, i.e. a language data sample of their own pupils' work (Granger, 2004: 125). That many of them lack this knowledge was demonstrated in the survey conducted by Mukherjee (2004): only about ten percent of some 250 English teachers were familiar with corpus linguistics.

Teachers who are motivated to apply corpus-based work in their classroom must also know that small corpora can be representative enough. Compiling large learner corpora is not always necessary. Actually, today the notion of a small or a large corpus has become rather relative (Sinclair, 2001). The size of the corpus is not crucial, the aim and methodology are what determine its relevance (ibid.). As pointed out by Granger (2004: 125) "small corpora that are compiled by teachers of their own students' work ...[can] draw attention to an individual or group profile of learner language use."

The goal of this paper was to compile a small² corpus of students' written homework and provide a linguistic characterization of a complex grammatical feature in German connecting expressions, i.e. connectives (such as *und*, *aber*, *nämlich*). By analysing essays of advanced German students in their 2nd year at university we attempted to investigate their

¹Cf. Altenberg and Granger (2001: 183-184): how students can be aware of underuse/overuse of relational verbs; Conrad (1999: 15-17): how they can be aware of register variation when using linking adverbials; Hyland (2007: 266-285): the importance of reformulations and exemplifications in academic writing; Guillot (2002: 30) stresses the contribution of corpus work for the meta-linguistic and meta-cognitive awareness of learners.

² Corpus-based research presumes determining the size of the corpus. According to conducted investigations, even a sample of 1000 words, with ten different texts representing a category of text, can give reliable source of a certain grammatical features (Conrad 1999: 3). For this study a corpus of 32 essays, 13,250 words, was compiled.

ability to connect two stretches of discourse and afterwards share with them the insights about their language use. Although the results of the study have relevance only for the given group of students and illustrate their difficulties in learning and using the German language, the study also illustrates the possibility of combining the use of corpus material with language instruction. In cases where the focus of instruction is a particular grammatical structure, corpus-based work can provide insight into the use of that particular structure (Reppen & Simpson, 2001: 107).

2. THE STUDY

2.1 Connectives

In this paper the term “connective” refers to the words that connect two discourse parts (Pasch et. al., 2003: 1), i.e. “signal the connections between passages of text (...) and are important devices for creating textual cohesion” (Conrad, 1999: 4). In German, a connecting word or *Konnektor* can be (among others) a conjunction, an adverbial or even a pronoun (Pasch et al., 2003). Traditional grammars divide connectives into several categories: additive, adversative, causative, concessive, conditional, disjunctive, temporal and others (cf. Duden 2005; Willkop, 2001, Helbig & Buscha, 2005):

- 1 additive: (e.g. *und, auch ferner, übrigens*)
- 2 adversative (e.g. *oder, sonst, aber, dennoch, dagegen, jedoch, allerdings, immerhin*)
- 3 causative (e.g. *denn, , deswegen, schließlich, daher, so*)
- 4 concessive (e.g. *obwohl, wenngleich, dennoch, trotzdem, nichtdestoweniger*)
- 5 conditional (e.g. *wenn, falls, sofern, bei*)
- 6 disjunctive (e.g. *oder*)
- 7 temporal (e.g. *dann, da, danach, so, nun, unterdessen*)
- 8 others (eg. summative *kurzum, jedenfalls*, explicative *so, nämlich, dass heißt*, etc.)

The choice of a connecting expression affects the word order of a German clause, and a connective can occur at different positions in the clause. This is one of the problems students of German as a foreign language are confronted with. Furthermore, the meanings of some categories overlap and students have problems in deciding which

connective is the “appropriate” one to make clear the connection between ideas. The meaning is mostly determined by or becomes clear in the context (Velčić, 1987: 21; Sommerfield et al.: 1998: 279). Furthermore, connectives are sometimes optional, i.e. relationships between sentences are also clear without a connective (Conrad, 1999: 5). When encouraged by teachers to use connectives in their essays, students tend to overuse them and the texts can sound awkward (ibid.). The avoidance or the overuse of such connectives is an interesting aspect for a corpus-based study. What can be investigated among other issues is whether certain connecting expressions are used more often than others, whether there are differences in usage within a certain group of students, or whether there are differences between native and non-native speakers’ use of connectives and what kind of differences these are.

Therefore this small study aims at establishing the overall frequency of connectives in the analyzed essays and also at investigating which connectives were used by the group of advanced German students who were participants in the research (the group specifics are described below). The aim was to provide insight into the students’ proficiency, i.e. whether they tend to overuse certain connectives and underuse others. The study also aimed at detecting whether the use of connectives is proportionally distributed through the entire corpus or whether only some of the students ventured to use them.

2.2. The corpus and method

In contrast to generalized, specialized and pedagogic corpora which consist of authentic texts stored electronically (Bennet, 2010: 14), the computer learner corpus (CLC)³ is a specialized corpus, also stored electronically, which contains texts by language students (ibid.). Its purpose is, for example, to establish the errors that students made or to develop goal oriented language instructions for the respective group (Bennet, 2010; Granger, 2004). The CLC used for this study is a collection of students’ essays that were produced as part of a written exam and submitted by students themselves electronically within a brief period of time (in the course of a week). The students, advanced German students in their 2nd year at university, Croatian native speakers, were asked to

³ The term “computer learner corpora (CLC)” used by Granger (2004: 124), and referred to as “electronic collections of spoken or written text produced by foreign or second language learners.”

write a short,400-word essay (a biography of a well-known German scientist, writer, composer, painter, etc. of their own choice) and they were encouraged to use connectives in the essay where appropriate. The use of connectives was explained and practiced during the lessons preceding the exam, and it was stressed that the essay would also be evaluated from the aspect of the usage of connectives. The students participating in the study had studied German as a foreign language for at least 6-9 years and their proficiency level was considered advanced.

The corpus consists of 32 texts, i.e. 13,250 words. The corpus was not annotated, i.e. no linguistic information was added to the electronic corpus of written data (Granger, 2004: 128). Accordingly, it was not possible to use ready-made tools, such as lemmatizers, part-of-speech taggers, or parsers. On the other hand such computer programs “have been trained on the basis of native speaker corpora, and there is no guarantee that they will perform as accurately when confronted with learner data” (ibid.). Since the corpus was very small, it allowed the manual identification of connecting expressions. Only those connectives were marked that linked two separate clauses and not those within a single clause (e.g. two subjects or objects in the same clause). After that, the AntCom concordance program⁴ was used to determine the frequency of connectives and their distribution within the group of students. The program consists of several tools which enable us to either see how words and phrases are used in the corpus, to see the position where search results appear, or to scan the entire corpus for 'N' length clusters (the N-grams Tool). This allows us to find common expressions in a corpus.

2.3. Results and discussion

Of the 48 connectives in Helbig and Buscha’s list of connectives (2005: 562-564), 14 appeared at least once, but 34 of them, i.e. 70%, did not occur at all.

One might be satisfied with the fact that students tried to use different semantic categories of connectives and thus stylistically

⁴AntConc is a freeware, multiplatform tool for carrying out corpus linguistics research and data driven learning created by Laurence Anthony, Ph.D. from the Center for English Language Education in Science and Engineering, School of Science and Engineering, Waseda University, 341 Okubo, Shinjuku, Tokyo 169, 8555, Japan.

improve their essays and add to the coherence of the text, but the breakdown of semantic categories provides insight into a rather one-sided preference. Table 1 shows the frequency and percentage of semantic classes out of the total number of words.

Table 1: *Connectives in the corpus by their meaning*

Connective	Frequency	Percentage of total words
Temporal	34	0.25%
Causative	11	0.08%
Adversative	23	0.17%
Additive	116	0.87%
Others (Summative)	1	0.0007 %
Total	185 connectives	1.39%

As evident, semantic categories of concessive, conditional and disjunctive connectives were not represented at all, whereas the summative category occurred only once. Since summative connectives are more typical in academic prose (Conrad, 1999: 7) than in biographical essays, which constitute the sample of this study, their omission is not disturbing. Another large discrepancy is evident between additive connectives and other semantic groups. The explanation for this might also be connected with the topic of the sample and viewed as an attempt to enumerate as many biographical facts as possible, rather than to explain their relationship by using different cohesion devices. This explanation, however, is not completely satisfactory, primarily in light of the distribution of single connectives (Table 3). The only additive connective used was *und*. According to the Common European Framework of Reference for Languages CEFR (Council of Europe, 2001) this connective is considered to be a "simple connective" or a "very basic connective" (along with the temporal *dann*) which is expected to be used at low levels of language learning. The range of different connectives should expand across proficiency levels, i.e. more advanced learners should be able to make more use of less frequent connectives than learners at lower levels, and learners should gain increased control of connectives as they progress. Accordingly, it would be realistic to expect that the students partaking in this study, who had been learning German

for at least 6-9 years, and whose proficiency level was considered advanced, should be able to make use of other additive connectives such as *auch, sowie, außerdem, darüber hinaus, nicht zuletzt*.

Although the CEFR does not predict greater connective density at higher levels (it predicts only a greater range), the frequency of connectives in this corpus can in general be interpreted as low, considering the fact that students were instructed to use them. This result is even more worrying when bearing in mind that 8 essays (25%) do not contain any other connectives than the additive *und* (in English: and). Table 2 shows the overall distribution of connectives (excluding the additive connective *und*) in the essays.

Table 2: *Distribution of connectives in essays*

Number of connectives per essay	Number of essays
1	9 (28%)
2	7 (21%)
3	2 (6.25%)
4	3 (9.37%)
5	3 (9.37%)

Out of 24 essays, 16 essays contain only one or two connectives. That means that half of the students ventured to use only one or two connectives other than *und*. The absence of connectives in a text will not lead to a complete loss of sense between clauses, i.e. the process of giving sense to a text can also be achieved through the sequencing of structures (clauses) (Velčić, 1987: 33), but connectives contribute to the meaning and purpose of a text in general. When only additive connectives are used to connect clauses, things are “put together” primarily to stress the cohesion of a text (Velčić, 1987: 39). However, when argumentative (causative and adversative) connectives are used, as well as those that refer to a certain situation or circumstance, such connectives, apart from giving cohesion to a text, also (explicitly) express the logical relation that exists between the connected structures, or they even modify this relation (*ibid.*). Thus, when analyzing connectives, questions about a text must inevitably be asked, about its purpose and the consequences for the informational value conveyed by the text (*ibid.*). This is something that should be clarified within the group after the results are presented.

The above results can also be interpreted in another way, namely that students do not possess a proper command of connectives, of their meaning and role in a text. Students are not aware of the stylistic and informational effects they could achieve by using certain connectives in a text. The breakdown into the number of occurrences of specific connectives might provide better insight into the students' command of the connectives used in the essays. Table 3 shows the overall occurrence of specific connectives in the texts.

Table 3: *Distribution of connectives in essays*

	Connective	Overall
Temporal:	dann	5
	da	4
	danach	23
	so	2
Causative:	nämlich	6
	deswegen	2
	schließlich	4
	daher	1
Adversative:	aber	9
	dennoch	2
	jedoch	5
	doch	7
Additive:	und	116
Summative:	kurz	1

These results show that the connectives mostly used by students are those taught at a very low level in foreign language teaching, i.e. *und*, *danach*, *aber* (in English: and, after that, but). Apart from the usage related to the task, one of the reasons why students preferred to use these “very basic connectives” might also be that they were rather sure which word order they should apply in the clause (which is one of major problems in studying German).

Since the corpus was very small, each essay could be read and the usage of connectives (other than the above-mentioned *und*, *danach*, *aber*) was compared with the overall style and vocabulary of the essay. Students who ventured to use different connectives throughout the whole essay had an overall better command of German throughout the entire essay. All those essays in which the usage of the additive *und*

prevailed (in three essays there were as many as 9) contained more grammatical and lexical mistakes than the essays with a wider range of connectives, which might lead to the conclusion that the avoidance of connectives is a sign of a lower language proficiency level.

3. CONCLUSION

On the basis of an electronically compiled corpus an attempt was made to detect difficulties in the language learning of a certain group of students, to provide students with insight into their specific problems and to find solutions for them. The results show that learners of German as a foreign language, even those at an advanced proficiency level, have difficulties with using connectives and achieving appropriate text cohesion. They underuse connectives in general. When using a connective, they prefer those which could be considered "safe ones" with regard to their meaning and the possible implications for the syntax (word order). Merely half of the students ventured to use different connectives and thus tried to add more information and improve their essays.

These results have certain pedagogical implications. By diagnosing students' strengths and weaknesses in using connectives it is possible to plan class activities in accordance with the results of the study. It is useful to present the results to students and thus raise their consciousness of the complexity of the problem established. Later on teachers could introduce appropriate exercises on connectives aimed at improving students' knowledge and proficiency.

One possibility is to start with native language corpora and propose exercises that would point out the syntactic and semantic complexity of connectives and provide insight into the cohesion possibilities of connectives in the mother tongue. Then similar exercises could be carried out in German. For example, students could be given small text passages without a single connective. After that, the same text could be presented with connectives. Students should compare the texts and discuss the differences, not only at the syntactic level, but also with regard to the content and meaning. They could be asked to make a list of sentence examples in which connectives appear and extend the list by adding more connectives which they had deliberately used in one of their texts, those which were unknown to them, or those the meaning of

which they were unsure of. The list could be used as an aid whenever they write a text. The more often they knowingly consult the list, the higher their awareness and the better their proficiency in using connectives.

This small study primarily served the internal purposes of a single group and aimed at showing the possibilities of investigating small computer corpora of learner texts. It also provided insight into insufficient mastery of connectives and their role concerning the cohesion, style and content of a text. After gaining insight into the results of the study, students became aware of their shortcomings in language proficiency and of the aspects of language that should be learned more intensively. Furthermore, students were able to compare their essays to those which were better (or more successful) with regard to the usage of connectives and, finally, they gained a better understanding of evaluation criteria. Such studies and their results allow students to be more directly involved with their learning process. "Data-driven learning" (Johns, 1991) gives foreign language learners a new perspective on learning: it "allows learners to problematize language, to explore texts and to authenticate discourse independently and collectively, adding to the reality of the corpus the reality of their own experience of it" (Gavioli & Aston, 2001: 244).

What has not been discussed and could be the subject of a future study is a comparison of native and non-native learner corpora as well as a comparison between non-native learners of different foreign languages with the purpose of establishing the differences or similarities in the usage of connectives.

REFERENCES

- Altenberg, B., Granger, S. (2001): The grammatical and lexical patterning of 'make' in native and non-native student writing, *Applied Linguistics* 22/2, 173-195.
- Bennet, G.R. (2010) An Introduction to Corpus Linguistics. In: *Using Corpora in the Language Learning Classroom: Corpus Linguistics for Teachers*. <http://www.press.umich.edu>, Michigan ELT.
- Biber, D., Reppen, R. (2002): What does frequency have to do with grammar teaching? *Studies in Second Language Acquisition* 24.
- Biber, D., Conrad, S., Cortes, V. (2004): If you look at . . . : Lexical bundles in university teaching and textbooks, *Applied Linguistics* 25/3, 371-405.

- Biber, D., Conrad, S., Leech, G. (2002): Longman student grammar of spoken and written English. Harlow: Longman.
- Conrad, S. (1999): The importance of corpus-based research for language teachers, *System* 27, 1-18.
- Cortes, V., Csomay, E. (2007): Positioning lexical bundles in university lectures, *Applied Linguistics*: Peter Lang.
- Csomay, E. (2006): Academic talk in American university classrooms: crossing the boundaries of oral-literate discourse?, *Journal of English for Academic Purposes* 5, 117-135.
- Csomay, E. (2007): A corpus-based look at linguistic variation in classroom interaction: Teacher talk versus student talk in American university classes, *Journal of English for Academic Purposes – Special Edition*.
- DUDEN. Die Grammatik (2005): Die Grammatik. Unentbehrlich für richtiges Deutsch. Ed. Duden-Redaktion (Duden Band 4). Mannheim.
- Fortanet, I. (2004): The use of 'we' in university lectures: reference and function, *English for Specific Purposes* 23, 45-66.
- Gavioli, L., Aston, G.(2001): Enriching reality: language corpora in language pedagogy, *ELT Journal* 55 (3), 238-246.
- Granger, S. (1997): On identifying the syntactic and discourse features of participle clauses in academic English: native and non-native writers compared In: Aarts, J., de Mönning, I., Wekker, H. (eds.), *Studies in English language and teaching*: Amsterdam&Atlanta: Rodopi, 185-198.
- Granger, S. (2004) Computer learner corpus research: Current status and future prospects, In: Connor, U., Upton, T. (eds.) *Applied Corpus Linguistics. A Multidimensional Perspective*: Amsterdam, Rodopi, 123-145.
- Guilot, M.-N. (2002): Corpus-based work and discourse analysis in FL pedagogy: a reassessment. *System* 30, 15-32.
- Helbig, G. & Buscha J. (2005) *Deutsche Grammatik. Ein Handbuch für den Ausländerunterricht*. Langenscheidt, Berlin-München-Wien-Zürich-NewYork.
- Hyland, K. (2002): Directives: Argument and Engagement in Academic Writing, *Applied Linguistics* 23/2.
- Hyland, K. (2007): Applying a Gloss: Exemplifying & Reformulating in Academic Discourse, *Applied Linguistics* 23/2: 266-285.
- Johns, T. (1991): Should you be persuaded: two examples of data-driven learning. In: Johns, T. & King, P. (eds.), *Classroom concordancing*, *ELR Journal* Vol. 4: Birmingham, Birmingham University Press, 1-16.
- Mukherjee, J. (2004) Bridging the gap between applied corpus linguistics and the reality of English language teaching in Germany, In: Connor, U. & Upton, T. (eds.), *Applied corpus Linguistics. A Multidimensional Perspective*: Amsterdam&New York: Rodopi, 239-250.
- Pasch, R. et al. (2003): *Handbuch der deutschen Konnektoren. Linguistische Grundlagen der Beschreibung und syntaktische Merkmale der deutschen Satzverknüpfung (Konjunktionen, Satzadverbien und Partikeln)*. Berlin, De Gruyter.

- Quirk, R., Greenbaum, S., Leech, G., Svartvik, J. (1985): *A comprehensive grammar of the English language*. London, Longman.
- Reppen R. & Simpson, R.: (2001): *Corpus Linguistics*. In: Schmitt, N. (ed.) *An Introduction to Applied Linguistics*: London, Arnold, 92-111.
- Sinclair, J. (1987): *Collocations: a progress report*, In: Steele, R., Threadgold, T. (eds.), *Language topics. Essays in honour of Michael Halliday*: Amsterdam&Philadelphia, John Benjamins, 319-331.
- Sinclair, J. (1990): *Collins Cobuild English grammar*. London, Collins.
- Sinclair, J.(2001): *Preface*, In: Ghadessy, M., Henry, A., Roseberry, R.L. (eds.) *Small corpus studies and ELT*: Amsterdam, John Benjamins.
- Sommerfield, K.E., Starke, Günter, Hackel, Werner (1998): *Einführung in die Grammatik der deutschen Sprache*. De Gruyter.
- Velčić, M. (1987) *Uvod u lingvistiku teksta*. Zagreb, Školskasknjiga.
- Willkop, E.-M. (2001): *Linguistische Analyseverfahren von Texten*. In: Helbig, G. et al. (ed.) *Deutsch als Fremdsprache. Ein internationales Handbuch*. 1. Halbband: Berlin, New York, de Gruyter, 314-324.

UPORABA KONEKTORA – ANALIZA KORPUSA UČENIČKIH PISANIH RADOVA

Cilj je ovoga rada ukazati na mogućnosti primjene korpusne analize i korisnost analize (maloga) korpusa u nastavi stranoga jezika. U tu je svrhu načinjen malen korpus učeničkih pisanih radova i analizirana je uporaba njemačkih konektora (poput *und*, *aber*, *nämlich*). Analizom se željelo utvrditi koliko su učenici upotrebljavali konektore za povezivanje rečenica i koje su konektore koristili. Dobiveni rezultati bili su korisni s dva aspekta: kao informacija učenicima o njihovoj uspješnosti/neuspješnosti u korištenju određenoga leksika, odnosno gramatičke kategorije, i kao informacija nastavniku o uspješnosti/neuspješnosti usvojenosti određene nastavne jedinice. Premda su rezultati ove analize bili relevantni samo za dotičnu skupinu učenika, ovim se radom želi pokazati mogućnost učenja (i poučavanja) stranoga jezika u kojemu je učenik u središtu procesa učenja i u kojemu se strani jezik uči uvidom. Nastavnici mogu primjenom korpusne analize dobiti povratnu informaciju o tome koji aspekti stranoga jezika nisu usvojeni i koji se trebaju dodatno uvježbavati na nastavi.

Ključne riječi: analiza korpusa, mali korpusi, njemački konektori

