Izvorni znanstveni članak
UDK: 37.091.3:811=111

Primljen: 19. 2. 2019.
Prihvaćen: 24. 6. 2019.

CORPATT: A SCALE FOR MEASURING ATTITUDES TOWARDS CORPUS USE

Sanja Marinov[*]
Ekonomski fakultet Sveučilišta u Splitu

This study reports on the development of an instrument for measuring learners' attitude towards the use of a language corpus in language teaching. The use of corpora in teaching is known as data-driven learning (DDL) and research has shown that students can benefit from corpus consultation to develop their language learning competences because it presents an important source of real target language discourse (Luo, 2016; Mizumoto, Chujo & Yokota, 2016). In addition to studying the effects of DDL, research has also shown interest in learners' perceptions of such a learning experience, an issue addressed in a number of studies in a range of contexts and with a wide range of methods.

This study aims at adding to this research and filling the gap by providing an instrument that measures the actual attitude towards such learning experience rather than asking students about the perceived benefits and problems of using corpora.

The participants were students majoring in Tourism at the Faculty of Economics, Business and Tourism, University of Split whose attitude is measured using the newly developed instrument. We also compare the results of the two sub-samples: students of the undergraduate study programme of tourism and students of the professional study programme of tourism.

Students' attitude towards the experience was assessed using a 41-item questionnaire that was submitted to item analysis which resulted in the reduction of the instrument to the current 24 items. The principal components analysis indicated two components as sub constructs on this instrument: (i) the cognitive-behavioural component, and (ii) the affective component of the measure of attitude. Internal consistency of the two multi-item scales proved to be high indicating that a psychometrically valid scale for measuring attitude towards corpus use was constructed. Finally, we found that the students showed an average attitude towards the experience of corpus use and no significant difference was found between the two sub-samples compared using the independent samples t-test. Limitations of the study and suggestions for further study are emphasised at the end of the paper.

Keywords: *corpus, data-driven learning, attitude, cognition, affect, questionnaire*

1. *INTRODUCTION*

The use of corpora in teaching is known as data-driven learning (DDL), independently of whether learners engage with corpora in a direct or an indirect manner (Leech, 1997; Römer, 2011), i.e. whether they browse the corpora themselves or they work on the pre-processed and edited corpus

[*] smarinov@efst.hr

data. The wealth of research on implementation of DDL has shown that almost every context is unique and will thus influence the results of the intervention as well as the feedback received from the learners. The differences in the type of engagement with corpora, its position in the course syllabus, the amount of time allocated for such an intervention, and the learners' proficiency level should be considered when interpreting the results.

DDL researchers have already established that students can benefit from consulting corpora to improve their L2 writing skills because it presents an important source of real target language discourse (Luo, 2016; Mizumoto, Chujo and Yokota, 2016; Thurstun & Candlin,1998). Empirical research into DDL has also established that coping with a large amount of data in concordance output of a large corpus may be an obstacle and a demotivator in proceeding with corpus use. This fact makes small specialised corpora particularly appealing, especially in English for Specific Purposes (ESP) courses. A number of advantages of such corpora were brought to our attention by Aston (1997): they are more easily manageable, more fully analysable, easier to become familiar with, easier to interpret, construct and reconstruct, they are more clearly patterned and their limits are clearer. Constructing such a corpus is not a problem any longer because easy access and availability of texts afforded by modern technologies enable not only experts but also students to choose texts from a variety of genres, registers, text types, domains and styles rather than to rely on general texts. Exploring specific genres in ESP courses is of particular importance (Flowerdew, 2005) because students need to acquire the linguistic conventions specific to a genre that are not always easily found in grammars or dictionaries (Flowerdew, 1993).

When designing the task we drew on the above findings and beliefs, yet deeply aware of contextual limitations, i.e. the position of DDL in the English language in Tourism 1 and 2 course syllabus (in which the respondents were enrolled), i.e. the amount of time available for introducing the approach and practising the skill. However, we do believe that this is a more realistic scenario in many contexts where it is not possible to dedicate the whole course to DDL or to practice the approach at great length before students get the grasp of it and get to realise its full potential.

The outcomes of any corpus use will, among other necessary skills and contextual constraints, depend on the users' willingness to engage in its usage. This paper describes the process of developing a scale for measuring attitudes towards corpus use. The measure was developed based on students' experience gained in an English for Specific Purposes (ESP) course where they engaged in compiling and browsing a small specialised corpus.

Although learners' feedback on their experience with corpus consultation has already been addressed in a number of studies, we have developed a new

measure of attitudes towards corpus use (CORPATT). This paper describes the process of developing the instrument and explains what differentiates it from other ways of collecting feedback on DDL experiences. Finally, using the newly developed instrument we measure the attitude of the given sample and compare the results of the sub-samples of respondents: students of the undergraduate study programme of tourism (USPT) and students of the professional study programme of tourism (PSPT).

The literature review presents an overview of DDL studies that have collected learners' feedback on the experience detailing the characteristics of the contexts in which the intervention was introduced, the methods used to obtain feedback and the respondents' most prominent attitudes.

2. *LITERATURE REVIEW*

Second language acquisition has paid a lot of interest to the beliefs about language learning because the outcomes of language learning do not depend only on the materials used and techniques applied but also on internal processes of each individual (Stevick, 1980). Language learners approach language learning with their own pre-existing beliefs about the nature of language learning and about how a language should be learned that have a major influence on their behaviour in the learning process (Horwitz, 1987; Dörnyei, 2005).

DDL researchers and practitioners have long recognised that learners' beliefs play a major role in learning and have thus, along with studying the effects of DDL, also shown interest in learners' perceptions of such a learning experience. This section offers an overview of some of the findings related to these perceptions. The literature review presented here is by no means exhaustive. It serves as an illustration of the variety of contexts in which DDL has been performed and a range of approaches used to collect the users' feedback.

Some studies have asked for participants' feedback on indirect use of corpora, i.e. they did not browse a corpus but worked with a series of corpus-based exercises. For example, in Thurstun and Candlin's (1998) study a set of exercises was developed using a corpus of academic texts to address a number of frequent vocabulary issues in academic writing. The material was meant for classroom use and independent learning. In all other studies reviewed here the participants engaged with corpora in a direct manner.

The literature review has shown that DDL is being implemented at various levels of language proficiency which is either defined in terms of the levels of the Common European Framework for Languages (CEFRL) or approximately: e.g. Mizumotoetal's (2016) participants were at A1/A2 level and Geluso and Yamaguchi's (2014) at A2/B2 level while Liu and Jiang's (2009) report that their participants were intermediate to upper intermediate

and Yoon and Hirvela's (2004) were intermediate and advanced. Occasionally the participants of different levels of language proficiency were compared and differences in their feedback noted. Comparing an intermediate and an advanced English as a Second Language (ESL) academic writing course Yoon and Hirvela (2004) found that the intermediate class students had a more positive attitude which was explained by the fact that they were engaged in more hands-on and in-class corpus activity while the advanced group was encouraged to explore corpora on their own. Also, the intermediate students may have perceived themselves more like language learners than the advanced students and were therefore more open to new opportunities for improving their language skills.

Studies on DDL can be large-scale involving a large number of participants such as, for example, Liu and Jiang's (2009) study which involved 236 students and 8 instructors; 160 students in an English as a Foreign Language (EFL) course at a large Chinese university and 76 students in an English as a Second Language (ESL) course at two universities in the United States, Aşik, Şarlanoğlu-Vural, and Akpinar's (2016) research that involved 126 Turkish pre-service English teachers, or Mizumoto et al.'s (2016) study that included 267 respondents. There are also small-scale studies such as Yoon and Hirvela's (2004) with 22 participants responding to a questionnaire and 4 participants taking part in interviews, Geluso and Yamaguchi's (2014) study with 30 respondents completing a questionnaire or Luo's study (2016) with 48 participants in the DDL activities, out of which 10 were interviewed for the feedback.

The contexts in which DDL is implemented vary to a great extent. Apart from students' language proficiency, there are issues of participants' majors, the type of course, amount of time dedicated to DDL within the course syllabus, type of corpus used and type of exercises students engage in. Geluso and Yamaguchi (2014) contextualised their study of DDL in a semester-long optional course in the Department of International Communication at a private foreign language university in Japan. The first three weeks were invested in introducing corpus work using the Corpus of Contemporary American English (COCA), a large general corpus. Students explored the corpus autonomously to improve their language use in a series of "speaking journals" they were asked to prepare in the course of the semester. The same corpus was used in the Aşiketal's (2016) research on DDL which studied the attitudes and beliefs of 126 Turkish pre-service teachers majoring in English. They were introduced to DDL as a part of a one-semester long obligatory *Lexical Competence* course which included vocabulary learning strategies, affixes, synonyms and antonyms, collocations, denotation and connotation, and idioms. Not all participants in DDL are language majors but may come from different academic programmes. For example, Yoon and Hirvela's

(2004) participants were enrolled in science-related academic programmes. They received instructions on how to use the Collins COBUILD Corpus, another large general corpus, and were involved in weekly exercises using it. Similarly, the participants of Luo's (2016) study were non-English majors from a science and engineering university in western China. In two parallel English classes taking the *College English* course where the teacher met with students twice a week, the author compared the effects of using the British National Corpus web and the Baidu search engine on students' writing accuracy, fluency and complexity. Liu and Jiang's (2009) large cohort of students were engaged in the use of the British National Corpus (BNC) and the BNC Baby for one semester. Mizumoto et al's (2016) students were science and engineering majors at low level of English proficiency so the DDL syllabus was developed to meet the requirements of such beginner-level learners. Bilingual corpora and parallel concordances were used to enable learners to understand the target language concordance lines and thus overcome the often reported difficulty in interpreting them. This DDL intervention was part of a compulsory English course at the students' university and it extended over a period of three months.

Feedback data on the experience with using DDL have been collected using a range of quantitative and qualitative instruments: while some relied on qualitative data only (e.g. interviews in Luo's study, 2016) others opted for a quantitative instrument (e.g. Geluso and Yamaguchi, 2014). In Luo's (2016) study the DDL intervention was followed by an interview with 10 students from the experimental group to find out about their attitudes towards the BNC web. Geluso and Yamaguchi (2014) obtained feedback using a 44-item questionnaire consisting of items adapted from previous studies as well as items newly designed by the researchers. The items were merged into multi-item scales based on theoretical considerations (difficulty in using corpora; positive impact of using corpora; effectiveness of presentation and delivery of coursework; completing speaking journals and incorporating phrases; and attitudes and beliefs about DDL and its potential).

A qualitative study sometimes precedes a quantitative study in order to collect preliminary data used in designing quantitative instruments as in Mizumoto et al. (2016). The authors developed and validated a psychometric scale to measure learners' perceived preferences and benefits of DDL because they had noticed a lack of such an instrument in empirical studies. The initial pool of items for this questionnaire was based on the authors' long-standing involvement with teaching in the given context as well as on previous studies describing students' experiences with DDL. At the end of each course, the students' open-ended responses to questions about perceived preferences and benefits of DDL were studied, categorised and coded and a text analysis was conducted. Although the students reported both benefits and drawbacks

of DDL, the authors decided to focus on the perceived preferences and benefits and excluded the drawbacks. The procedure resulted in 18 items of a questionnaire which was administered after a three-month DDL intervention. Item analyses included: (i) calculating the item-total correlations to determine whether the figures were over 0.3; (ii) performing exploratory factor analysis to investigate which items belonged together; (iii) analysing Cronbach's α levels to verify the internal consistency of the subscales. Two factors were identified: 1 – Clarity, measuring the advantages of DDL in clarifying the authentic use of target structures and 2 – Autonomy, measuring the extent to which the learners embrace the autonomy of learning in this approach. Two items were excluded from the final version of the questionnaire. Internal consistency reliability (Cronbach's α) for Clarity was $\alpha = 0.91$ and for Autonomy $\alpha = 0.81$. The Pearson correlation coefficient between the two was 0.60 which suggests that the two sub-scales measure similar constructs under an overarching theme (perceived preferences and benefits of DDL).

The third possibility is combining two or more instruments for eliciting students' opinions - mixed methods studies. Yoon and Hirvela's (2004) study investigated students' behaviours in corpus use as well as their perception of strengths and weaknesses of corpora as a second language writing tool. The quantitative data on students' experience with using corpora were collected using a 62-item questionnaire: 23 items collecting information about students' personal background, computer use, and dictionary use and 42 items eliciting their perceptions of the strengths and weaknesses of corpus pedagogy. The items looked into advantages of corpus use, problems/difficulties of corpus use, students' response to corpus use in writing instruction, solving problems by using the corpus in writing, and overall evaluation of corpus use. The authors reported a high level of reliability of the instrument ($\alpha = 0.96$). The qualitative data were collected through interviews to help interpret and support the data obtained by the questionnaire.

Aşik et al.'s (2016) mixed methods approach included a questionnaire and a focus group interview. The questionnaire consisted of 43 items, some adapted from previous studies on using corpora and some designed by the researchers to address the particular nature of the study. The authors reported high reliability of the scale ($\alpha = 0.92$). Eventually, the items were grouped under five categories: (i) learners' attitudes towards the effectiveness of DDL tasks in terms of lexical awareness and development; (ii) attitudes towards the effectiveness of DDL tasks for developing lexical awareness and proficiency; (iii) positive aspects of corpus use; (iv) difficulties encountered during corpus use; (v) attitudes towards the delivery of task-based DDL instruction. Fifteen randomly selected participants took part in a focus group which lasted for 30 minutes and was moderated by the researchers.

A study can use a wider range of instruments with the purpose of triangulation as in Liu and Jiang (2009) where the following instruments were adopted: (i) students' work (corpus search assignments, grammar exercises, written reports and reflections on their findings); (ii) instructors' work and observations (teaching logs, lesson plans, sample teaching activities, reflection journals); and (iii) students' and instructors' post-study questionnaires.

In the reviewed studies, the participants generally agreed that the DDL approach was helpful for studying vocabulary whether they just studied the concordance output in an indirect way (Thurstun & Candlin, 1998) or were faced directly with a corpus. For example, the students felt that this approach to language learning increased their knowledge of collocations, helped them acquire new phrases and discover new ways of using familiar vocabulary (Geluso & Yamaguchi, 2014). Aşik et al. (2016) found positive attitudes towards DDL instruction in terms of raising awareness of synonyms and collocations but not so much positivity towards word frequency, idioms and vocabulary learning strategies. Regarding the use of corpora in foreign language writing, Yoon and Hirvela's (2004) data indicated that the corpus approach was found to be beneficial to the development of L2 writing skills and increasing confidence towards L2 writing. Luo (2016) recognised a positive attitude towards using corpora in the revision stage of writing, e.g. raising awareness of the importance of collocations and learning words in chunks and context rather than in isolation. Finally, triangulation of the data collected in Liu and Jiang's (2009) study showed that the use of corpora can raise students' language awareness, help acquire more lexico-grammatical rules, increase their understanding of the importance of context, help develop critical understanding of grammar, and promote discovery learning.

On the downside there seem to be three major issues, reflecting both the technological as well as the psychological barrier (Luo, 2016): (i) the lack of skill and experience in reading concordances (Geluso & Yamaguchi, 2014; Luo, 2016;Thurstun & Candlin, 1998); (ii) coping with a large number of unknown words and the complexity of language found in the corpus output (Liu & Jiang, 2009;Thurstun & Candlin, 1998); (iii) the lack of skill, experience and motivation for an inductive approach to learning, i.e. for discovery learning (Aşiketal, 2016; Liu & Jiang, 2009; Luo, 2016).

Since the purpose of this study was to develop yet another instrument for measuring students' attitudes towards DDL the following paragraphs will first describe the participants and the DDL intervention, then the process of creating the item pool for the questionnaire, along with the rationale for this approach, and finally the procedure of designing it and verifying its validity.

3. *AIM*

The main aim of this study is to design an instrument (CORPATT) that can qualitatively measure students' attitudes towards corpus use, as one of their individual differences. Using the newly designed instrument we will also be able to answer two ensuing research questions: 1) What is our respondents' attitude towards corpus use? 2) Is there any difference between the two subsamples (students of the USPT and PSPT)?

4. *METHOD*

4.1. *Participants*

The study participants were 1st year students at the Faculty of Economics, Business and Tourism, University of Split in Croatia majoring in Tourism. The sample can be further broken down into two subsamples: one group enrolled in the PSPT (n=60), while the other comprised students of the USPT (n=70). The former focuses more on professional while the latter on academic skills. Students of professional study programmes are commonly enrolled as students with a weaker academic background and often less ambitious career plans, while students of undergraduate programmes enter the university programmes with a stronger academic record and more ambitious career plans. In both programmes English for Specific Purposes (ESP), or more precisely English for Tourism, is an obligatory course in the first year. The language proficiency ranges from B1 to B2 (according to the school leaving exam the students take prior to their enrolment at university). Throughout the course syllabus DDL was introduced mostly in the indirect way with a variety of exercises, mostly vocabulary-oriented, included in the regular teaching material. With both groups direct DDL was used only once: with undergraduate students of Tourism in their 1st semester as part of the English in Tourism 1 syllabus and with the professional study programme in the 2nd semester as part of the English in Tourism 2 syllabus., i.e. as a teaching intervention this study follows from.

4.2. *Materials/instruments used*

In order to measure students' attitudes towards the direct use of corpus we first introduced a teaching intervention which involved corpus use and was related to a written assignment concerning the topic of accommodation. The corpus that was used was a small specialised corpus containing 30 Airbnb texts prepared in advance by the lecturer and later extended by the texts collected by the participants. Each student ended up browsing a corpus of different size. The concordancing programme used for browsing the corpus was AntConc (Anthony, 2018) which is freely available and easily downloadable to a large number of computers at once.

Upon completion of the task involving corpus use the participants were administered a questionnaire attempting to measure their attitude. More details about both are given in the following section.

4.3. *Procedure*

4.3.1. *Teaching intervention*

The activities involving corpus use were split into 6 stages:

Stage 1 – Students were asked to write a draft of the advertisement (real or imagined) of the accommodations they are renting on Airbnb following the paragraph structure suggested by Airbnb (The space; Guest access; Interaction with guests; The neighbourhood; Getting around). The drafts were written under test conditions to ensure the authenticity of students' own production. The students were told it was a draft and, as such, would not be graded.

Stage 2–Students were introduced to the idea of a language corpus and the basics of browsing it. The introduction was organised as a hands-on activity in the IT lab with students following and repeating the steps presented by the lecturer and described in the handout. Students were introduced to the possibilities of creating word lists, creating and sorting concordances, using a collocation tool, and the possibility of accessing the wider context in which a particular instance of language was used.

Stage 3 – Students were assigned to select more texts to add to the original 30 files.

Stage 4–The class met once again in the IT lab to add their texts to the existing corpus and/or to exchange their texts to help each other create bigger corpora. They also individually practiced using the options presented the week before looking for any language items that may have been of interest considering the written assignment they had.

Stage 5 –In the following two weeks students were asked to write a "Corpus search journal" with at least 4 entries. In this individual, out-of-class assignment each entry had to contain information about (i) the purpose of the search; (ii) AntConc options they used in performing the search; (iii) the outcome of the search; and (iv) the selected useful language. This part of the activity was optional, but students were granted up to 5 points for it (out of the maximum 80 that they get for this course).

Stage 6–Students were asked to reconsider their original drafts and make any necessary changes to the text using any resource they thought was suitable (dictionary, Google translate, corpus, ...). This part of the activity was also optional, and students were granted up to 5 points for it.

Finally, students were asked to give their feedback about the experience by completing a questionnaire. This was done on a voluntary basis with a

total of 76 responses submitted. It is important to note that the questionnaire we are presenting was based on such an intervention but can, nonetheless, fit other similar contexts. Following is the account of the procedure of creating the questionnaire item pool.

4.3.2. *Creating an item pool and administering the questionnaire*

The main aim of creating this questionnaire was to produce a psychometrically valid instrument that would measure users' attitude towards corpus use. The measure obtained can potentially be correlated with a variety of other measures of students' individual differences such as age, study programme, proficiency level, language strategies used, self-regulation of learning, motivation, amount of experience with corpus use, etc. Although the item pool does stem from the perceived benefits, preferences and drawbacks of using corpora in language learning the idea is not to describe these but to get the actual measure of attitude. Unlike Mizumoto et al. (2016), who excluded the drawbacks and focused only on benefits, our approach required the negative statements to be included as well because they are all part of an attitude which is defined as "a psychological tendency that is expressed by evaluating a particular entity with some degree of favor or disfavor" (Eagly & Chaiken, 2007:598). This evaluation includes cognitive, affective and behavioural response (Eagly & Chaiken, 2007; Hilgard, 1980), where cognitive stands for knowledge, beliefs and values of the entity, emotional for feelings towards the entity and behavioural for behavioural intention regarding the entity (Zvonarević, 1981).

In constructing the item pool, we considered the literature on students' feedback about corpus use as well as the three-year experience with classroom observation and task outcomes on the same task with different groups of students. No items from previous questionnaires were borrowed because the main principle applied in item construction was to comply with the "attitude architecture", consisting of the three principal characteristics of attitude (cognition, affect, behaviour). The items were worded accordingly. An additional criterion was to aim at a balance between positively and negatively worded items.

The procedure resulted in 41 items as shown in Table 1 with answers provided on the 5-point Likert scale (see also the Appendix for the final version of the questionnaire) ranging from "I strongly disagree" to "I strongly agree". Responding to the questionnaire was voluntary and anonymous. The questionnaire was administered using the Moodle platform, regularly used as a virtual learning environment at the Faculty of Economics, Business and Tourism in Split. A total of 76 students responded to the questionnaire, out of which 43 students of the PSPT and 33 students of the USPT Students were asked to use codes in order to enable matching the questionnaire answers to

other parts of the task (written assignment and the "Corpus use journal") in some future studies.

4.4. *Data analysis*

The content of the students' drafts and journals was not analysed because that was beyond the scope of this study and not directly relevant for current research questions. The analysis focused on the items of the questionnaire where we tried to determine which of the items better met the criterion of internal consistency, i.e. whether they measured the same construct – in this case an attitude. After calculating the item-total correlation we eliminated the "weak items" and with the remaining ones performed the factor analysis to find out if we could discern between various factors that constituted the given attitude. A word of warning is due here considering the size of the sample and its adequacy for factor analysis. Given the number of items on this questionnaire a sample of 76 is relatively small for a factor analysis but we decided to perform it given the strong internal consistency that was recorded. We are aware that with a larger sample the results may be different. Finally, the t-test was applied to find out if there were any significant differences between the two subsamples. The software package used for statistical analysis was version 23.0 of SPSS.

5. *RESULTS*

5.1. *Questionnaire item analysis*

Multi-item scales are valid only if the items within a scale measure the same construct, i.e. if they meet the criterion of internal consistency. Each item on a scale should correlate with the total score and with other items (Dörnyei & Taguchi 2010).Accordingly, the first step in conducting the item analysis was to calculate the item-total, i.e. correlations between each particular item and the total score, which provided us with information about which of the items correlated better with the total score. Table 1 shows that all correlations except one (item number 9) are significant. Most correlations are significant at the 0.01 level, while two are significant at the 0.05 level (items number 7 and 22). Cronbach's Alpha (α = 0.95) indicates high internal consistency of the initial item pool.

The following criteria were imposed when selecting the items to retain in the final version:

- aim at a questionnaire of about 20 items;
- exclude all items with non-significant item-total correlation;
- retain the items with the highest item-total correlation;
- achieve a balance of positively worded and negatively worded items.

Table 1. Reliability analysis of the original item pool (questionnaire)

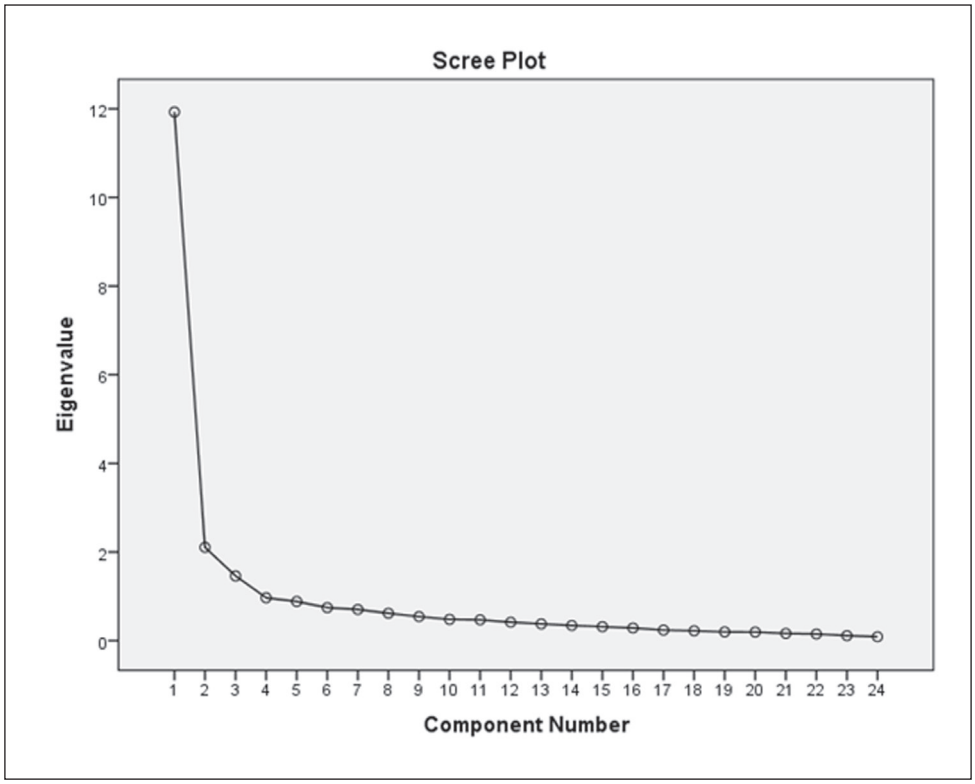
	Items of the original version of the questionnaire	Item-total correlation	Corrected item-total correlation	Cronbach Alpha if item deleted
1	**Browsing a corpus is useful.**	**0.71****	**0.69**	**0.95**
2	Browsing a corpus is complicated.	0.34**	0.28	0.95
3	The results of corpus search give a clear picture of how to use a word in context.	0.43**	0.41	0.95
4	I access and browse the corpus on my own initiative.	0.48**	0.44	0.95
5	When browsing the corpus I do not always find what I am looking for.	0.40**	0.37	0.95
6	I believe that studying the corpus examples is a good way of learning a foreign language.	0.57**	0.54	0.95
7	The problem with corpus search is that we are left to cope on our own.	0.26*	0.22	0.95
8	**Browsing a corpus makes me an active participant in learning.**	**0.61****	**0.58**	**0.95**
9	When I browse a corpus I invest a lot of time in studying particular examples.	-0.13	-0.16	0.95
10	**Browsing corpora seems to be more efficient than some traditional types of exercises.**	**0.76****	**0.74**	**0.95**
11	**When I look into corpus data I feel helpless.**	**0.69****	**0.67**	**0.95**
12	**Using a corpus helps me practice the words/phrases that I recognise but I do not use actively.**	**0.62****	**0.59**	**0.95**
13	**I believe that by using a corpus I develop autonomy in language learning.**	**0.69****	**0.66**	**0.95**
14	**Browsing a corpus is tiring.**	**0.75****	**0.73**	**0.95**
15	**Studying corpus examples is not an efficient method for acquiring a foreign language.**	**0.80****	**0.79**	**0.95**
16	In a corpus I easily find language solutions (words, phrases) that I need.	0.54**	0.52	0.95
17	**In a corpus I easily find ideas that I need.**	**0.62****	**0.59**	**0.95**
18	**Using corpora inspires creativity.**	**0.76****	**0.73**	**0.95**
19	I think that by browsing a corpus we spend a lot of time on studying individual examples.	0.41**	0.38	0.95
20	**I do not need such a detailed view of the language.**	**0.69****	**0.67**	**0.95**
21	**I am delighted by the number of examples I see in a corpus.**	**0.69****	**0.67**	**0.95**
22	When working with corpora I lack the help of a lecturer.	0.26*	0.22	0.95
23	**I love seeing a large number of useful examples of language use in one place.**	**0.67****	**0.64**	**0.95**
24	**I am confused by the large number of data I see when browsing a corpus.**	**0.51****	**0.48**	**0.95**
25	**I like browsing a corpus.**	**0.70****	**0.67**	**0.95**
26	**Using a corpus frustrates me.**	**0.69****	**0.66**	**0.95**

	Items of the original version of the questionnaire	Item-total correlation	Corrected item-total correlation	Cronbach Alpha if item deleted
27	**I use a corpus only because I have to.**	**0.69****	**0.67**	**0.95**
28	**It is more fun to look at corpus examples rather than at course book exercises.**	**0.71****	**0.68**	**0.95**
29	**Browsing a corpus is exhausting.**	**0.75****	**0.73**	**0.95**
30	**Browsing a corpus makes me feel insecure.**	**0.71****	**0.69**	**0.95**
31	When I notice something useful in a corpus I am proud of myself.	0.54**	0.51	0.95
32	**Browsing a corpus is time-consuming.**	**0.60****	**0.57**	**0.95**
33	**Using a corpus is fun.**	**0.80****	**0.78**	**0.95**
34	Browsing a corpus is tiring because it forces me to use a dictionary a lot.	0.50**	0.48	0.95
35	**I would love to use corpora in the future.**	**0.79****	**0.72**	**0.95**
36	I am considering compiling a corpus myself.	0.54**	0.52	0.95
37	**Using a corpus enables us to work at our own pace.**	**0.65****	**0.62**	**0.95**
38	When using a corpus I constantly feel I have to ask someone whether what I am doing is good/correct.	0.39**	0.34	0.95
39	When using a corpus I do not know where to start from.	0.51**	0.48	0.95
40	I have difficulties mastering the technical aspects of corpus use.	0.46**	0.43	0.95
41	I use a corpus more frequently than I am asked to.	0.39**	0.36	0.95

Note: The 24 items in bold are those selected for the final version of the questionnaire.

The analysis of the items and correlations with the total score resulted in a 24-item questionnaire containing 14 positively worded and 10 negatively worded items. All correlations of individual items with the total score are higher than 0.60 ($r > 0.60$) except one (item 24, $r = 0.51$) which was estimated as important and kept to meet the condition of balancing the negatively and positively worded items.

Next, exploratory factor analysis was performed to establish which items belonged together and thus create sub-constructs within the questionnaire itself. The principal component analysis generated a rotated component matrix which suggested 3 components. Item analysis did not produce any meaningful connection between the items assigned to particular components. Difficulty of interpreting the factors and the inspection of the Cattell's scree test (see Picture 1) which levels off after two factors, indicated that the number of factors could be reduced to two.



Picture 1. Cattell's scree test for the 24 selected items

The repeated procedure generated two components that can be clearly distinguished: (i) cognitive-behavioural and (ii) affective, i.e. while the 1st factor is more about beliefs, values and behavioural intentions of using corpora, the 2nd better depicts the feelings this type of work creates.

Table 2 shows the results of the principal component analysis fixed to two factors. The two-component solution explained 58% of the total variance.

Table 2. Outcome of the principal component analysis, fixed to two factors (Extraction method: principal Component Analysis; Rotation Method: Varimax with Kaiser Normalization; Rotation converged in 3 iterations)

Rotated Structure matrix for PCA with Varimax Rotation of a Two Component Questionnaire			
Items		Rotate Component Coefficients	
		Component 1	Component 2
13	I believe that by using a corpus I develop autonomy in language learning.	**0.84**	0.15
12	Using a corpus helps me practice the words/phrases that I recognise but I do not use actively.	**0.79**	0.07
37	Using a corpus enables us to work at our own pace.	**0.78**	0.15
1	Browsing a corpus is useful.	**0.72**	0.28
35	I would love to use corpora in the future.	**0.71**	0.37
8	Browsing a corpus makes me an active participant in learning.	**0.68**	0.20
18	Using corpora inspires creativity.	**0.68**	0.38
15	Studying corpus examples is not an efficient method for acquiring a foreign language.	**0.67**	0.46
33	Using a corpus is fun.	**0.67**	0.52
20	I do not need such a detailed view of the language.	**0.67**	0.29
23	I love seeing a large number of useful examples of language use in one place.	**0.65**	0.31
21	I am delighted by the number of examples I see in a corpus.	**0.64**	0.38
17	In a corpus I easily find ideas that I need.	**0.60**	0.29
10	Browsing corpora seems to be more efficient than some traditional types of exercises.	**0.60**	0.50
28	It is more fun to look at corpus examples rather than at course book exercises.	**0.53**	0.49
29	Browsing a corpus is exhausting.	0.31	**0.80**
26	Using a corpus frustrates me.	0.21	**0.79**
32	Browsing a corpus is time-consuming.	0.10	**0.77**
14	Browsing a corpus is tiring.	0.36	**0.73**
30	Browsing a corpus makes me feel insecure.	0.29	**0.70**
27	I use a corpus only because I have to.	0.34	**0.64**
24	I am confused by the large number of data I see when browsing a corpus.	0.10	**0.63**
11	When I look into corpus data I feel helpless.	0.34	**0.62**
25	I like browsing a corpus.	0.42	0.56

Note: Major loadings for each item are bolded.

The cognitive-behavioural component consists of 15 items while the affective component contains 9 items. The pool of cognitive-behavioural items includes several statements that feature affective wording but can be interpreted as cognitive-behavioural since the enthusiasm may simply acknowledge the value and usefulness of such view of the language. (e.g. I love seeing a large number of useful examples of language usage in one place. OR I am delighted with the number of examples in the corpus.).

The mean inter-item correlation for both sub-scales is high: 0.54 for the cognitive-behavioural component and 0.50 for the affective component.

Internal consistency of each of the subscales was examined using Cronbach's Alpha coefficient which proved the internal consistency of both sub-scales to be high: α = 0.95 for the cognitive-behavioural scale and α = 0.90 for the affective scale. The internal consistency of the entire scale is α = 0.95. According to Dörnyei and Taguchi (2010) a well-constructed attitude scale consisting of 10 items should have α = 0.80, for short scales of only 3-4 items α = 0.70 is acceptable, while α = 0.60 indicates a problem. The Pearson correlation coefficient between the two scales was statistically significant at 0.01 level of significance and it is high (r = 0.72) which indicates that the two sub-scales measure similar sub-constructs of the underlying construct of attitude.

Table 3 shows the reliability analysis of the final questionnaire containing two sub-scales. No increase of internal consistency can be achieved in either of the scales by eliminating any of the items.

Table 3. Reliability analysis of the final 24-item questionnaire

	Items	Corrected item-total correlation	Cronbach Alpha if item deleted
	Cognitive-behavioural sub-scale		
1	I believe that by using a corpus I develop autonomy in language learning.	0.70	0.94
2	Using a corpus helps me practice the words/phrases that I recognise but I do not use actively.	0.61	0.94
3	Using a corpus enables us to work at our own pace.	0.65	0.94
4	Browsing a corpus is useful.	0.69	0.94
5	I would love to use corpora in the future.	0.75	0.94
6	Browsing a corpus makes me an active participant in learning.	0.61	0.94
7	Using corpora inspires creativity.	0.73	0.94
8	Studying corpus examples is not an efficient method for acquiring a foreign language.	0.78	0.94
9	Using a corpus is fun.	0.82	0.94
10	I do not need such a detailed view of the language.	0.66	0.94

	Items	Corrected item-total correlation	Cronbach Alpha if item deleted
11	I love seeing a large number of useful examples of language use in one place.	0.66	0.94
12	I am delighted by the number of examples I see in a corpus.	0.70	0.94
13	In a corpus I easily find ideas that I need.	0.61	0.94
14	Browsing corpora seems to be more efficient than some traditional types of exercises.	0.75	0.94
15	It is more fun to look at corpus examples rather than at course book exercises.	0.69	0.94
	Affective sub-scale		
16	Browsing a corpus is exhausting.	0.73	0.88
17	Using a corpus frustrates me.	0.64	0.89
18	Browsing a corpus is time-consuming.	0.54	0.89
19	Browsing a corpus is tiring.	0.72	0.89
20	Browsing a corpus makes me feel insecure.	0.65	0.89
21	I use a corpus only because I have to.	0.64	0.89
22	I am confused by the large number of data I see when browsing a corpus.	0.46	0.90
23	When I look into corpus data I feel helpless.	0.63	0.89
24	I like browsing a corpus.	0.65	0.90

Being a measure of attitude, the questionnaire has been named CORPATT and its full version is presented in the Appendix. The items 1, 2, 3, 5, 6, 8, 9, 10, 11, 12, 13, 18, 22, 23, 24 belong to the cognitive-behavioural component while items 4, 7, 14, 15, 16, 17, 19, 20, 21 belong to the affective component.

5.2. *Students' attitudes - descriptive analysis*

The analysis of this data set showed that students' attitudes towards the experience was average ($M = 3.15$, $SD = 0.68$) with the mean of the cognitive-behavioural component slightly higher ($M = 3.28$, $SD = 0.73$) than that of the affective component ($M = 2.94$, $SD = 0.74$).

The two sub-samples were compared to see whether the difference in students' educational background, academic record and choice of study programme is reflected in the expressed attitudes (Table 4).

The independent samples t-test was run to determine whether there were differences between the attitudes of the two sub-samples. The results showed that there was homogeneity of variances, as assessed by Levene's test for quality of variances. We were thus also able to determine the effect size of the obtained results by calculating the Cohen's d value.

The students of the USPT had a more positive attitude towards corpus use (M = 3.23, SD = 0.70) than the students of the PSPT (M = 3.09, SD = 0.67), a statistically non-significant difference, M = 0.14, 95% CI [-0.18, 0.45], $t(74)$ = 878, p = 383, d = 0.20

The cognitive-behavioural aspect of attitude of the students of the USPT was also higher (M = 3.36, SD = 0.73) than that of the students of the PSPT (M = 3.22, SD = 0.73), a statistically non-significant difference, M = 0.13, 95% CI [-0.20, 0.47], $t(74)$ = 785, p = 435, d = *0.18*.

Finally, the affective aspect of attitude of the students of the USPT was also more pronounced (M = 3.02, SD = 0.81) than that of the students of the PSPS, a statistically non-significant difference, M = 0.15, 95% CI [-0.19, 0.49], $t(74)$ = 871, p = 387, d = *0.20.*

Table 4. Comparison of the scores on the CORPATT between the two sub-samples

Study programme	N	Mean attitude	SD attitude	Mean cognitive -behavioural	SD cognitive-behavioural	Mean affective	SD affective
USPT	33	3.23	0.70	3.36	0.73	3.02	0.81
PSPT	43	3.09	0.67	3.22	0.73	2.87	0.69

6. *CONCLUSION*

The procedure described above has shown that a psychometrically valid instrument for measuring students' attitudes towards corpus use has been created. It is a multi-item scale questionnaire that does not intend to explore particular benefits and drawbacks of using corpora in teaching but is meant to be an actual measure of attitude. This measure can be included in studies researching the influence of other individual and contextual factors on students' readiness and willingness to engage in corpus use. In doing so, the context in which this instrument was created should be considered and necessary changes introduced accordingly in the wording of individual items or in the decision to keep them or reject them.

The items of the questionnaire showed to have high internal consistency from the very beginning, but its size was reduced for easier administration. Only the items with high item-total correlations were kept (r > 0.60 except for one item). The factor analysis was performed to try and distinguish between different sub-components characteristic of 'attitudes architecture'. Two factors were recognised: cognitive-behavioural and affective. The final version of the questionnaire has kept high internal consistency of the whole scale and of each of the two sub-scales.

The data analysis indicated that this sample showed an average attitude towards corpus use and the two sub-samples were compared. The fact that professional study students usually have lower academic records and are less

ambitious in terms of pursuing further careers while undergraduate study students start their higher education with better previous achievements and are more ambitious in terms of their academic and career goals has not significantly affected their attitude towards direct corpus use.

7. *LIMITATIONS AND FUTURE STUDIES*

The ideas for further studies partly reflect the limitations of the study. As has already been noted the results of this analysis should not be taken as conclusive because the factor analysis was performed on a relatively small sample compared to the number of items in the questionnaire. This was done in order to test whether in the attitude towards corpus use we can reveal the three constituent parts of attitude: cognitive component, affective component, and behavioural component. Rather than these three factors our analysis has produced two. Further studies should include more participants and confirm, reject or modify the current results. Administering a new round of questionnaires requires the teaching intervention to be repeated first which, as shown by the literature review, is not a simple task and requires time, resources, and effort but will certainly be implemented again.

Furthermore, in future studies the measure of attitude towards corpus use can be correlated with other measures indicating either learners' proficiency or their individual differences (e.g. language learning strategies, beliefs about language learning, self-regulation of language learning, and lexical competence, etc.) to find out how they are related.

Finally, a possible alteration in the corpus use intervention itself could be an introduction of different types of texts (e.g. different genre and register) and texts of increased difficulty to provide a greater challenge to the students' language proficiency. The assumption would be that they may show a more positive attitude towards corpus use as a learning tool if a more challenging learning task was provided.

REFERENCES

Anthony, L. (2018) AntConc (Version 3.5.7) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/software

Aşık, A., Şarlanoğlu-Vural, A. and Akpınar, K. D. (2016) Lexical awareness and development through data driven learning: Attitudes and beliefs of EFL learners. *Journal of Education and Training Studies* 4 (3), 87–96.

Aston, G. (1997) Small and large corpora in language learning. In Melia J. and Lewandowska-Tomaszczyk B. (eds) *PALC '97 Proceedings. The First International Conference: Practical Applications in Language Corpora*. Łódź: Łódź University Press, 51-62.

Dörnyei, Z. (2005) *The Psychology of the Language Learner: Individual Differences in Second Language Acquisition*. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Dörnyei, Z. and Taguchi, T. (2010) *Questionnaires in second language research: Construction, administration, and processing*. New York: Routledge.

Eagly, A.H. and Chaiken (2007) The advantages of an inclusive definition of attitude. *Social Cognition* 25 (5), 582-602.

Flowerdew, J. (1993)An educational, or process, approach to the teaching of professional genres. *ELT Journal* 47, 305–16.

Flowerdew, L. (2005)An integration of corpus-based and genre-based approaches to test analysis in EAP/ESP: Countering criticisms against corpus-based methodologies. *English for Specific Purposes* 24, 321–2.

Geluso, J. and Yamaguchi, A. (2014) Discovering formulaic language through data-driven learning: Student attitudes and efficacy. *ReCALL* 26 (2), 225–242.

Hilgard, E. R. (1980) The trilogy of mind: Cognition, affection, and conation. *Journal of the History of Behavioural Sciences* 16 (2), 107-117.

Horwitz, E.K. (1987) Surveying student beliefs about language learning. In Wenden, A. and Rubin, J. (eds) *Learner strategies in language learning*. Englewood Cliffs, NJ: Prentice Hall, 119 – 132.

IBM Corp. Released 2015. IBM SPSS Statistics for Windows, Version 23.0. Armonk, NY: IBM Corp.

Leech, G. (1997) Teaching and language corpora: A convergence. In Wichmann, A., Fligelstone, S., McEnery, T. and Knowles, G. (eds) *Teaching and language corpora*. London: Longman, 1–23.

Liu, D. and Jiang, P. (2009) Using a Corpus-Based Lexicogrammatical Approach to Grammar Instruction in EFL and ESL Contexts. *Modern Language Journal* 93 (1), 61–78.

Luo, Q. (2016) The effects of data-driven learning activities on EFL learners' writing development. *Springer Plus* 5 (1):1255.

Mizumoto, A., Chujo, K. and Yokota, K. (2016) Development of a scale to measure learners' perceived preferences and benefits of data-driven learning. *ReCALL* 28 (2), 227–246.

Römer, U. (2011) Corpus research applications in second language teaching. *Annual Review of Applied Linguistics* 31, 205-225.

Stevick, E. W. (1980) Teaching languages: A way and ways. Rowley, MA: Newbury House Publishers, Inc.

Thurstun, J and Candlin, C. N. (1998) Concordancing and the teaching of the vocabulary of academic English. *English for Specific Purposes* 17 (3), 267-280.

Yoon, H. and Hirvela, A. (2004) ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing* 13, 257–283.

Zvonarević, M. (1981) *Socijalna psihologija*. Zagreb: Školska knjiga.

CORPATT: SKALA ZA MJERENJE STAVOVA PREMA UPOTREBI KORPUSA

Ovaj rad opisuje proces i rezultat izrade instrumenta za mjerenje učeničkog stava prema upotrebi jezičnoga korpusa u nastavi stranoga jezika. Istraživanja su pokazala da proučavanje jezika u jezičnim korpusima može doprinijeti razvoju jezičnih kompetencija jer korpusi predstavljaju važan izvor stvarne jezične upotrebe (Luo, 2016; Mizumoto, Chujo and Yokota, 2016). Osim proučavanja efekta upotrebe korpusa na stvarno jezično napredovanje, istraživanja su se bavila i učeničkim percepcijama takvoga iskustva u učenju jezika u različitim kontekstima, koristeći čitav niz metoda.

Cilj je ovoga rada doprinijeti upravo području istraživanja učeničkih stavova prema radu s korpusom izradom instrumenta za kvantitativno mjerenje stava. Nadamo se da na taj način upotpunjujemo jedan nedostatak u istraživanju jer se u dosadašnjim studijama od učenika najčešće tražilo da se osvrnu na uočene prednosti i probleme rada s korpusom, ali ne i da izraze jasno mjerljivi stav.

Ispitanici u ovome istraživanju bili su studenti turističkih smjerova Ekonomskog fakulteta Sveučilišta u Splitu koji se dijele na dva poduzorka: studente sveučilišnoga studija Turizma i studente stručnoga studija Turističkog poslovanja.

Studentski stav prema iskustvu pretraživanja korpusa ocijenilo se upitnikom koji je sadržavao 41 česticu koje su nakon statističke analize svedene na 24. Analiza glavnih komponenti ukazala je na dva podkonstrukta stava koje je moguće ispitati ovim instrumentom: (i) kognitivno-ponašajni podkonstrukti (ii) afektivni podkonstrukt ukupne mjere ovoga stava. Unutarnja je konzistentnost upitnika kao i svake pojedine subskale visoka pa zaključujemo da je sastavljena psihometrijski pouzdana skala za mjerenje stavova prema upotrebi korpusa. Analiza podataka za trenutačni uzorak pokazala je umjereno pozitivan stav prema upotrebi korpusa, a t-testom nije utvrđena značajna razlika između dva poduzorka. Ograničenja istraživanja kao i prijedlozi za daljnja istraživanja izneseni su na kraju članka.

Ključne riječi: *korpus, učenje iz korpusa, stav, kognitivni stav, afektivni stav, upitnik*

APPENDIX

Please, tell us to what extent you agree with the following statements by circling the number corresponding to your answer.

I strongly disagree	I disagree	I neither agree nor disagree	I agree	I strongly agree
1	2	3	4	5

1	Browsing a corpus is useful.	1	2	3	4	5
2	Browsing a corpus makes me an active participant in learning.	1	2	3	4	5
3	Browsing corpora seems to be more efficient than some traditional types of exercises.	1	2	3	4	5
4	When I look into corpus data I feel helpless.	1	2	3	4	5
5	Using a corpus helps me practice the words/phrases that I recognise but I do not use actively.	1	2	3	4	5
6	I believe that by using a corpus I develop autonomy in language learning.	1	2	3	4	5
7	Browsing a corpus is tiring.	1	2	3	4	5
8	Studying corpus examples is not an efficient method for acquiring a foreign language.	1	2	3	4	5
9	In a corpus I easily find ideas that I need.	1	2	3	4	5
10	Using corpora inspires creativity.	1	2	3	4	5
11	I do not need such a detailed view of the language.	1	2	3	4	5
12	I am delighted by the number of examples I see in a corpus.	1	2	3	4	5
13	I love seeing a large number of useful examples of language use in one place.	1	2	3	4	5
14	I am confused by a large number of data I see when browsing a corpus.	1	2	3	4	5
15	I like browsing a corpus.	1	2	3	4	5
16	Using a corpus frustrates me.	1	2	3	4	5
17	I use a corpus only because I have to.	1	2	3	4	5
18	It is more fun to look at corpus examples rather than at course book exercises.	1	2	3	4	5
19	Browsing a corpus is exhausting.	1	2	3	4	5
20	Browsing a corpus makes me feel insecure.	1	2	3	4	5
21	Browsing a corpus is time-consuming.	1	2	3	4	5
22	Using a corpus is fun.	1	2	3	4	5
23	I would love to use corpora in the future.	1	2	3	4	5
24	Using a corpus enables us to work at our own pace.	1	2	3	4	5