Agata Križan
Filozofska fakulteta Univerze v Mariboru
agata.krizan@uni-mb.si

DEVELOPING ACADEMIC VOCABULARY WITH AWL GAPMAKER AND CORPUS DATA

1. Introduction

For students studying EFL at university level it is necessary to be acquainted with a certain amount of academic English to be successful not only within the university walls, but to function as competent users of English in contemporary society. Academic English namely entails "multiple complex features of English required for long-term success in public schools, completion of higher education, and employment with opportunity for professional advancement and financial rewards" (Rumberger and Scarcella 2001,1). As Scarcella further explains, knowledge of academic English prepares students to become a part of educated society and to be able to '… challenge the tenets of those in power who use it … without knowledge of academic English, individuals may be excluded from participation in educated society and prevented from transforming it" (2003,1). Marzano and Pickering (2005) point out that academic background knowledge is important to understand the content students encounter while studying scientific and professional texts, and which they will encounter later as they progress professionally. Since most of these texts include a certain number of academic words, it is necessary that some academic vocabulary is also systematically highlighted through the use of various tasks and exercises.

2. Academic vocabulary and context

Learning vocabulary in a foreign language is important because '*mistakes in lexical selection may be less generously tolerated outside classrooms than mistakes in syntax*' (Carter 1998, 185). The same is true for academic vocabulary. Mistakes or inappropriate use of academic words attaches the label of 'academically inefficient' to the user. Since students of EFL will be engaged in different occupational areas appropriate to the degree they have completed (in teaching and non-teaching areas), two abilities are foregrounded and defined by Cummins for successful communication in a professional target setting (1979). The first ability required in order to successfully communicate in an occupational setting is the ability to use the particular jargon characteristic of that specific occupational context. The second is the ability to use a more generalized set of academic skills, such as conducting research, responding to memoranda, and reading and writing academically. Since the professional vocabulary of a particular domain is usually rapidly learnt, learning academic vocabulary is often a much more demanding and complex process. Although professional and academic vocabulary are usually mentioned together in literature, the focus in this paper is solely on academic vocabulary. Before moving to an illustration of one possibility for the learning and testing of academic vocabulary, it is necessary to provide some backround explanation about the acquisition of vocabulary in general since this refers to academic vocabulary, too. When one learns new vocabulary it is not enough to simply know the meaning of a word and its translated equivalent. Knowing a word in a language also means to know both its syntagmatic and paradigmatic relations. This means to know a word in context – how a word is used and how it behaves syntactically in a particular context as well as in other contexts (Richards 1976). A more detailed description of what it means to know a word is provided by Carter (1998, 239). Some of these are given here:

- It means knowing the syntactic frames into which the word can be slotted and the underlying forms and derivations which can be made from it.
- It means knowing the relations it constructs with other words in the language and with related words in the L1 as well.

- It means knowing the different meanings asociated with it and, often in a connected way, the range of its collocational patterns.
- It means knowing words as part of or wholly as fixed expressions .

Since language and vocabulary are not used in isolation but in context, on the basis of which we make and decode meanings, academic and professional vocabulary must also be developed and practised in a context, which is also made clear from the description above. One of the widely used techniques available for enriching vocabulary is gap-fill exercises in context (later in the text referred to as a gap-fill), sometimes also referred to as a cloze. The usefulness of such gap-fill is based on the idea that it also demands inference from the context. Context has a whole variety of potential effects: on the particular meaning of the word, its connotations, the appropriateness of its use etc. (Raed 1997, 319). This means that the filling-in of blanks does not depend on individual items of linguistic knowledge only, but requires the use of language skills in general, which fosters the language acquisition. This is also emphasized by Deyes in his extention of a cloze to 'discourse cloze', where the aim is to 'reflect the reader's ability to follow information through the text and use contextual clues as well as cotextual ones' (1984, 128).

Context thus plays an important role in inferring correct academic entries in a gap-fill by providing certain linguistic clues (often referred to as co-text) as well as other clues, which may be very effective in the development of academic vocabulary (Chern 1993). Besides linguistic clues, 'world knowledge' and knowledge of the domain may also be helpful in decision making about the appropriate and correct entries (Nagy 1997, 76). All this is covered within the term 'context' used in this paper. Lingustic knowledge, for instance, includes the knowledge of grammatical structures and word classes of the gap's co-words, which further signals the required word class of the word being sought (for example, a preposition might signal a noun in the preceding position). It is thus important for students to be aware of the forms that language takes in academic contexts, whether spoken or written. Linguistic proficiency influences how successfully a student can use context; we expect those who undertake English studies to already posssess a higher level of linguistic knowledge of English. However, context cannot be

restricted to the textual neighbourhood of a word only, but also includes the students' world knowledge. By world knowledge we mean the knowledge of particular domains as well as general knowledge. This could be an advantage for students, who have become familiar with a large number of different domains before becoming university students. Training students in how to search for contextual clues and how to help themselves with the kinds of knowledge they have is important since this enhances their ability to infer correct entries from the context.

C-R and noticing are two of the other referential strategies efficiently used in practicing inferring from the context. Context-based frames which can be used for the purposes of developing inferring skills can range from cloze (gap-fill) exercises, and words-in-context exercises to context-enrichment exercises.

2.1. AWL and concordances

The AWL (Academic Word List) is a list of words which appear with high frequency in English-language academic texts. The list was compiled by Averil Coxhead (2000) at the Victoria University of Wellington, New Zealand. The list contains 570 word families of academic words and is divided into 10 sublists. The list is based on the analysis of over 3.500.000 academic journals, textbooks, course workbooks, lab manuals and course notes. Sublist 1 consists of the 60 most common words in the AWL. Sublist 2 contains the next most frequently used words and so on. Each sublist contains 60 word families, except for sublist 10, which contains 30. This list of academic words can help students to expand their general academic vocabulary, to improve comprehension of academic texts, and to write texts academically within their discipline, despite technical words being excluded from the list. The technical words of a discipline can be very rapidly acquired.

In the paper problematic entries are also commented on from the perspective of corpus data.

This means that certain incorrect entries are checked for their collocatibility, relations with other contextual words, and frequency of use with concordances. Corpus can provide useful and valuable insights into the behaviour of a lexical/grammatical item, because it appears in an authentic – real context. The collocatibility and frequency of certain

academic words are checked with a t-score and an Mi-score. The T-score is the statistical measure of frequency which compares the actual cooccurrences of words with the way they would co-occur randomly, and assigns a score to them accordingly. MI-score, as a statistical measure of frequency, emphasises relatively infrequent collocate pairs that are good predictors of each other (Potter 1999, 52). For example, any score above 2 is usually signifies that a pair is worth looking at it. For the purposes of this paper, the British Collins Wordbanks is the employed corpus, which comprises about 56 million words of contemporary UK, US and Australian spoken and written English. The sub-corpora selected are British and American news, books and magazines.

2.2. Gap-fill

According to Oller (1979, 33) a gap-fill exercise can be used effectively not just for testing students' internalised expectancy grammar/ vocabulary but also, in my opinion, to teach new vocabulary – in our case academic vocabulary, which is illustrated in the paper. Cloze procedure was introduced over 40 years ago and its applications have been numerous. It has been widely used and is apparently a simple technique. It has been used for measuring the readability of textbooks, the reader's degree of comprehension of a text, both aural and written, and as a means of teaching and testing linguistic proficiency in a foreign language – professional linguistic proficiency inclusive.

The more advanced the learners become (at the university stage they definitely should be), the more 'inferential' or 'implicit' and learner-centred vocabulary learning strategies will have to become (Carter 1998, 209). Vocabulary should not be treated and learnt as discrete items but learnt in context, as already mentioned above. The aim of such exercises is to develop *'the strategy* [*being*] *a means of acquiring the unconscious skill that an efficient reader already has*.' (Nation 1990, 162).

In its classic form, cloze consists of the regular deletion of words from a text, and their replacement by even-length blank lines. A gap-fill, on the other hand, uses deleted items subjectively chosen by the maker of the exercise (a teacher, for example). The student's task is then to guess the word that has been deleted or chosen from a range of options given. To make the task a little easier, deleted words can be given in

mixed form. There are different types of gap-fill depending on what one wishes to focus on and test: lexical, structural or discourse. They can have 'answers' given in a mixed form from which students choose or without available 'answers'. However, they should demand inferrence from the context, which plays an important part in the compilation of a gap-fill. Everything that is known about grammar, meaning, and the word can be activated to exploit redundancy to the full, and help to restore a word. This means that a greater degree of knowledge is involved in the inference process. With gap-fill we do not simply see the extent of knowledge a student has about individual lexical/grammatical items, but also the actual use of total language skills in a meaningful context. Sometimes a problem can occur in selecting the appropriate words to delete, therefore AWL gap maker can be of great help since the program itself deletes academic words from a text according to the level of difficulty chosen.

3. The procedure

I selected a text of 250 words and entered it into the AWL gapmaker program developed by the University of Nottingham (www. http://www.nottingham.ac.uk/). The text is a short review of a professional book for language teachers and includes several core academic words related to the teaching of languages. The program then deleted all the academic words it recognised in the text. The difficulty levels chosen were 3 and 7. The higher the level the higher the number of deleted words. With level 3, 13 words were deleted, while with level 7, the number of deleted words was 19. The following academic words were deleted:

Level 7: aids, approach, approach, appropriate, areas, comprise, culture, designed, evolved, innovations, major, principles, relevant, structures, target, techniques, thechnological, theory, utilization.

Level 3: approach, approach, appropriate, areas, culture, designed, major, principles, relevant, structures, techniques, technological, theory.

The gap-fill was used by two goups of 1st year students of English at the Faculty of Arts in Maribor. One group worked with level 3, the

other group with level 7. There were 20 students in each group solving a gap-fill. Each gap in the text was numbered (1,2,3,…).

In the text the following abbreviations are used: for Level 3 (L3), for Level 7 (L7).

4. Results and discussion

Correct/incorrect entries (answers/words) at each level were counted for each student. In Table (1) the numbers of gaps are given in succession next to the correct entry and the number of students' correct entries. For example, in gap (4), 16 students entered the correct word (approach). In table (2) correct entries ranked from the highest to the lowest number of correct entries next to the correct word and the gap number. Some of the most problematic incorrect entries were then isolated and discussed in terms of the possible reasons for incorrectness. Also, suggestions for correct answers based on contextual clues, lingustic or other, are given. What is also apparent is how knowledge of the domain (knowledge about English grammar) can be helpful in filling in the gaps in a professional text. Concordances were used as a reliable and useful source for supporting or rejecting claims made about some problematic entries, as well as for emphasizing their important role in the students' decision–making about the entries. By way of illustrating how concordances can be explored by students and training them to use them with a gap-fill, students become engaged in 'discovery learning' – which means coming to conclusions about the use and patterns of collocation of (academic) words (Barbieri and Eckhardt 2007, 320).

Level 3

Table 1:

Gap number in consequtive order	Correct entry	Number of correct entries
1	designed	8
2	Major	0
3	Areas	5
4	approach	16
5	Culture	13
6	Theory	5
7	approach	8
8	principles	2
9	structures	4
10	techniques	6
11	appropriate	5
12	relevant	8
13	technological	14

Table 2:

Rank of correct entries		Gap number
16	approach	4
14	technological	1
13	Culture	5
8	relevant approach designed	1, 12, 7
6	techniques	10
5	areas, theory, appropriate	3, 11, 6
4	structures	9
2	principles	8
0	Major	2

As both tables for level 3 show, most correct entries (16) were given in gap 4 (approach), followed by gap 1 (technological with 14 entries), gap 5 (culture – with 13 entries), and gaps 1, 12 and 7 (designed; relevant and approach with 8 correct entries). The least correct entries were given in gap 8 (principles with 2 entries) followed by gap 9 (structures with 4 entries), gaps 3,11 and 6 (areas; appropriate and theory with 5 correct entries), and gap 10 (techniques with 6 correct entries). In gap 2 (major) either no answer or an incorrect one was given. This gap was also the most problematic for the students. Gap (2) appeared in a pair with gap 3 (areas), which was also problematic for the students.

Level 7

Table 1:

Gap number in consequtive order	Correct entry	Number of correct entries
1	designed	8
2	innovations	5
3	major	6
4	areas	0
5	approach	15
6	culture	6
7	theory	4
8	approach	9
9	evolved	16
10	principles	3
11	utilization	3
12	structures	5
13	target	1
14	techniques	0
15	appropriate	1
16	relevant	3
17	technological	4
18	aids	5
19	comprise	0

Table 2:

Rank of correct entries	Correct entry	Gap number
16	evolved	9
15	approach	5
9	approach	8
8	designed	1
6	major, culture	3, 6
5	innovations, structures, aids	12, 18, 2
4	theory, technological	7, 17
3	principles, utilization, relevant	10, 11, 16
1	target, appropriate	13, 15
0	areas, techniques, comprise	4, 14, 19

The most problematic entries were with gaps (4 – areas; 14 – techniques and 19 – comprise with o correct entries) followed by gaps (13 – target and 15 – appropriate with 1 entry). Problematic were also gaps (10 – principles; 11 – utilization and 16 – relevant with 3 entries) followed by

gaps (7 – theory and 17 – technological with 4 entries). The least problematic entries was gap (19 – evolved with 16 correct entries) followed by gaps (5 – approach with 15 correct entries), (8 – approach with 9 entries) and gap (1 – designed with 8 entries).

Since there were some most problematic common entries in both groups, I excluded them and commented on them first, then I proceeded to the most problematic entries of each group.

Most problematic common entries with both levels were: principles (L3=2, L7= 3); structures (L3=4; L7=5) and theory (L3=5; L7=4). Most problematic entry with L3 was: major (L3 = 0; L7 = 6). Most problematic entries with L7 were: 'techniques' (L3 = 6; L7 = 0); areas (L3=5; L7=1) and appropriate (L3=5; L7=1) + words deleted only at L7 (comprise, target, relevant, utilization).

The results show that the number of correct entries at level 3 was in general slightly higher than at level 7, which might be due to more co-text (less deleted words). This would be true for, for example, following entries: approach: L3 = 16, L7 = 15; relevant: L3 = 8, L7 = 3; culture: L3 = 13, L7 = 6); technological: L3 =14, L7 = 4). A close scrutiny, however, reveals that the higher number of deleted words in the vicinity of gaps is not always responsible for the incorrect entries, and does not always affect correct entries, as is the case with, for example, 'appropriate' and 'techniques'. Relevant (L3 = 8; L7 = 3) and technological (L3 = 14; L7 = 4) are used at both levels without additional gaps, but still did group (L3) better than group (L7). This is also true for the entry 'structures', since there is only a slight difference in number of correct entries at both levels despite more deleted words at L7 (L3 = 5, L7 = 4). The reason for this could be found in the use of other contextual clues, as is illustrated in the paper – general knowledge and/or knowledge of the domain (related to the topic of the review and grammar of English).

In the following part of the paper I focus first on the most problematic common entries (problematic in both groups) in the sense of trying to provide possible reasons for wrong entries, and to suggest/illustrate how clues in the context with the support of concordances can be helpful in inferring the right answer. Then I proceed to entries which show major discrepancies, finishing off with some of the most problematic entries from each group. For the sake of

easier understanding, sentences with gaps under focus are excluded from the whole text for the comment.

It is necessary to mention that, although quite a high number of correct entries appeared with gaps 1, 12 and 7 in both groups, the number of correct entries is still not high enough to see the knowledge students showed with entries as sufficient.

4.1. Most problematic common entries

Major and *areas*

Although 'major' represented more difficulties for L3 compared with L7 (L3 = 0, L7 = 6), I discuss it in pair with 'areas' (despite being more problematic for L7 (L3 = 5, L7 = 0) since they apear in both texts in succession in a sentence. The same goes for other gaps in succession.

Designed to update the contemporary language teacher's knowledge of the innovations in his field, this book introduces some of the essential 1_____ 2 _____ of which he should have an understanding to…

There are two important clues for the students given in the vicinity of the two gaps which could be helpful in inferring correct entries. The first piece of relevant information for the student is the knowledge of word order in a sentence, which explains that most adjectives can occupy an attributive (before a noun) or predicative (as a complement of a linking verb) position. They are called modifiers because they do something to change or modify a noun. In our case 'essential', as the given adjective, should be recognised as a grammatical clue implying that gaps (1) and (2) should be filled with another adjective followed by a noun. The use of the preposition 'of' after gap (2) further signals a relative clause modifying the preceding antecedent, which needs to be a noun.

The knowledge of word order at sentence level also suggests that the antecedent is possibly premodified by one or more adjectives. Since 'essential' as an adjective already occupies second place before a noun in the adjectival group (first being reserved for the determiner 'the'), the

next adjective gaps can be occupied by adjectives in the following order: general description, physical state and proper adjectives. Concordances reveal that 'essential', as an adjective, is always placed first when in combination with other adjectives preceding nouns. There are four adjectives available as options for both groups: appropriate, major, relevant and technological. The most frequent collocates with 'essential' found in concordances are adjectives, such as daily/fatty/legal/ long/new/small/large/main, which points at the exclusion of 'technological', as well as of 'relevant' and 'appropriate' since the latter two do not appear in combination with 'essential' in concordances. The closest in meaning compared to 'large' is thus 'major'. Once this word is used, one can test it's collocatibility with some of the nouns given as options. For example, by looking at concordances, 'areas' collocates strongly with 'major', while at the same time, concordances reveal a great number of the combination 'major+areas+of' (t-score = 7.699; Mi = 2.559), which implies 'areas' is the correct answer in gap (1). Focusing on sentence/clause structure might also provide useful clues for entering correct words. The gaps in focus appear in a clause complex that consists of one main clause and one subordinate clause. Since 'the most important new information goes at the end of the rheme in a clause', one can establish that 'knowledge of the innovations', with 'knowledge' as a head, (or aspects related to 'knowledge') is the main focus in the rest of the text (Goalty 2000, 16). Concordances show that 'knowledge' collocates, although not frequently, with 'areas', which suggests some semantic interconnection of both words and can be seen as an indication of the possible use of both words in the vicinity of each other. Besides, there is another example found within concordances ('major+areas+of+modern+ knowledge'), which might also be a useful clue for the choice of 'major' and 'areas' in the sentence in focus.

Appropriate

…and the use of testing 1_____ 2 _____ to the linguistic problems being tested.

This complex noun phrase with the nominalized head (the use) is one among other complex noun phrases used as a grammatical structure encoding four of the modern principles of language teaching (the utilization, the association, the postponement and the use of testing techniques) enumerated. Since the text is a professional book review, the use of complex noun phrases with nominalized forms is a logical choice for packing in more information and sounding more scientific. Since complex clauses using nominalizations in head positions often cause comprehension problems, we can perhaps assume that this was one of the main reasons for the wrong entries, especially with gaps at the point of contact, as is the case here. By having more knowledge of ellipsis students could have thought of this possibility. Ellipsis and its possible uses is often neglected in EFL classrooms, despite its frequent use by native speakers. Therefore, it may be reasonable to place more emphasis on the teaching of this grammatical structure.

It can be observed that there are two postmodifying relative/adjective[1] clauses in this complex phrase: 'which is/are ….; which/that are being tested. The adjective 'testing' is already given to signal the following gap to be occupied with a noun. On the collocation list for 'testing', entered as an adjective in corpus, 'techniques' do not appear, however, there is one collocate (methods), which could be, perhaps, seen as a near-synonym of 'techniques'. Shifting this information to options available, 'techniques' would have been chosen. It is also relevant that one can test techniques, but not, for example, culture. Once the noun is found, one can proceed to the next gap, previously recognized as part of a postmodifying relative clause with the omitted pronoun 'which' and the verb 'are'. Knowledge of the zero-marker would be useful here, as well (Blaganje and Konte 1987, 429).

[1] The adjective clause is also called a relative clause (Frank 1972, 276).

The use of the preposition 'to' in this omitted clause further signals that the preceding word should be an adjective.

Concordances reveal that from the options available only 'appropriate' and 'relevant' could be used in this gap since they both collocate with prepositions - 'relevant' collocates with 'in' and 'to', and 'appropriate' with 'to' and 'for'.

'Relevant' collocates far more often with 'to' (1866 occurrences out of 9244) than with 'in' (187 out of 9244). 'Appropriate', on the other hand, collocates slightly more often with 'for' (t-score = 37.188, Mi = 1.626) than with 'to' (t-score = 31.220, Mi= 0.842).

From this one could conclude that 'relevant' is the correct entry. By checking the concordances of how 'appropriate' and 'relevant' behave as collocates, these reveal 1 occurrence of 'appropriate+ techniques', while there are no occurrences of 'relevant+ techniques'. Even this single example of use found in concordances can be seen as a sufficient indicator for 'appropriate' as the right choice.

Principles

Part two systematically treats such modern _____ of language teaching as…

The MacMillan English dictionary for advanced learners describes 'principle(s) as a basic belief, theory, or rule that has a major influence on the way in which something is done, and as a scientific theory or basic natural law that explains the way in which something works (2007, 1177). If this useful information is joined with the contextual knowledge of the domain (teaching and learning languages), one can see how this can be applied in inferring the right entries. For example, there are some titles of scientific and professional books on teaching and learning languages which use 'principles' in their titles (Brown *Principles of language learning and teaching*; Krashen *Principles and practice in second language acquisition*) or even use 'techniques and principles' in the title (Larsen-Freeman *Techniques and principles in language teaching*). This knowledge may make the choice of entry much easier. Although 1st year students are not yet familiar with the domain of didactics, examples of the titles above are used as an illustration of the role of contextual

knowledge of the domain in inferring entries in such gap fills. However, general (world) knowledge might also be of help here.

For example, it is still possible that some students do have a general idea of what 'principle' means, similar to the definition in the dictionary. When the knowledge of the domain is of no help, a look at the concordances reveal that 'learning' is a frequent collocate of 'principles+of' occupying the 24th place on the collocation list among 50 collocates (t-score = 4.780; Mi = 4.511). This information can be used for assuming a close connection with 'teaching'. Teaching and learning nevertheless belong to the same semantic field.

Another important piece of information revealed by concordances is that 'modern' also appears on the collocation list for 'principles+of' occupying the 34th place (t-score = 4.257; Mi = 3.439). Although only one example of 'principles+of +teaching' is found in the following context - …learning the principles of teaching and working in a classroom - this information could be seen as sufficient for choosing 'principles' for the gap here.

Structures

The utilization of linguistic knowledge in the control of the sounds and _____ of target languages.

The immediate grammatical item after the gap is the preposition 'of' (of-prepositional phrase), an equivalent of the genitive noun phrase expressing the genitive relation between nouns, and thus requiring the use of a noun in the gap. The of-construction is mostly used with nouns denoting things, as is the case here. The joining 'and' following a noun (sounds) also requires the use of a noun in the gap. This knowledge of clause system and structure could be effectively supported with the general knowledge about language being made of sounds and structures, which would imply 'strucutres' to be the answer. Since 'structures' as a noun is given as one of the options, the decision for correct entry should not have been too difficult for the students. When such knowledge is not available, a look into the concordances might be more helpful. Although these reveal a non-frequent collocatibility of

'sounds' and 'structures', one should try further by searching for 'sounds+and', which shows 2 occurrences of this combination. This is enough to use this combination in the gap. One can also check the collocates for 'structures', where 'sentence' and 'language' can be found as frequent collocates. All three imply some associative interconnection. Since 'language' also appears in the vicinity of a gap, this supports the choice of 'structures' as the correct answer.

Theory

…sets the tone of the book with discussions of the relationship of language to lingustics and culture and the learning _____ from which the new scientific approach to …

The determiner 'the' coupled with the adjective 'learning' signals the use of a noun in the gap.

Since there are a few options available with which 'learning' could appear in combination (techniques, principles, areas, theory), it is necessary to look for additional clues in the proceeding relative clause, where one can find the wording 'scientific approach'. For a scientist and a professional it is logical that approaches evolve from theory. By checking further in concordances, one can see that 'learning' strongly collocates with 'theory' (t-score = 10.807 and Mi - 3.644). The greatest difference in the number of correct entries was found with 'relevant' and 'technological' (relevant: L3=8, L7=3; technological: L3=14, L7= 4) within the same given context. This shows that the context (lack of it) could not be responsible for the small number of correct entries at L7 compared with L3. One of the possible reasons could perhaps lie in filling other gaps with incorrect words due to the lower level of context in L7. However, the following comments focus on linguistic and other contextual clues in the vicinity of the slots to explain their role in determining correct entries.

Relevant and *technological*

1_____ to 2_____ teaching devices, part three discusses the language laboratory, visual aids, and teaching machines,…

This sentence has two clauses, seperated by a comma, which signals the dependency relation of one clause on the other. A relevant piece of grammatical information for students is that main clauses mostly consist of finites, seen here in 'discusses'.

Also important for students is that the content expressed in the main clause and in subordinated ones is usually related, which means in this case that the dependent clause somehow provides extra information about the main clause. Students should also know that clauses of cause or reason may occupy an initial, medial, or final position. Extra information in dependent clauses can be expressed through the use of finite clauses, non-finite clauses, or verbless clauses. If one knows that English clauses normally consist of a subject and a finite, a logical conclusion would be to try to enter one of the finites (verbs: evolved, designed, comprise) given as options into the gap (2). Of course, they make no semantic sense. Another possibility then is to try to use a noun acting as a subject. For example, 'approach', may fit in since it can be followed by the preposition 'to'. Again, it does not make sense. Here, wider grammatical knowledge indicates that clauses can also be verbless (with an omitted verb, and often with no subject.). This means the use of another grammatical element for gap (1) which is not a verb and, as seen before, not a noun. Since the main clause expresses present action (discusses), lingustic knowledge of clausal time reference requires in the use of a finite in the dependent clause, which is in a kind of simultaneous relation with the main one. This means that the finite should refer to the present time. Since adjectives can be positioned predicatively, one can assume that the omitted verb would be 'be – being', as an expression of a 'simultaneous' relation).

Also, by looking at the main clause, one can further determine its resultative function implying the causal function of the dependent clause (why part three discusses something) on the basis of the 'to+complex noun' construction as the indirect object. This further

implies the use of an adjective in gap (1). Meaningful adjectives which collocate with the preposition 'to' could be either 'appropriate' or 'relevant'. Since I have already commented on both uses above (see under 'appropriate'), the correct answer here is 'relevant'. Interestingly, by entering 'being+relevant' as a search item in the corpus, 28 such instances occur, which is very useful information for the choice of entry. Also, if students had been aware of some examples of verbless clauses, they could have compared their choices of adjectives with an example provided by Leech and Stvartvik accompanied with important information about the adjective – verbless clause information (1994, 250): An adjective, alone or as head of an adjective phrase, can function as a verbless clause: e.g. Anxious for a quick decision, the chairman called for a vote.

The use of verbless clauses is one among other features formal written professional/scientific texts employ to sound scientific and to pack more information in (ibid., 183). Therefore the knowledge of clause relations and its types is very important for students. Another important contextual clue for gap (2) can be found in the main clause, specifically, in the nouns enumerated in the object position (language laboratory, visual aids, and teaching machines), which signal the logical inference of a common cover adjective for these objects - 'technological'. Concordances reveal that although, for example, 'technological' is not on the collocation list for 'devices', there is one collocate found which is similar in meaning: electronic.

Table (2) for L3 shows that the most problematic entries are those which were also the most problematic in L7, which have already been commented upon above (see ''major', 'principles', 'structures', 'areas', 'theory', and 'appropriate'). Therefore I will now deal with the most problematic L7 entries. The most problematic ones were the following: comprise, target and utilization.

4.2. Most problematic entries with L7.

Comprise

A glossary of useful terms, a bibliography, and a chart of the modern language association qualifications for secondary school language teachers _____ the appendixes.

One of the reasons for the incorrect entry in this sentence could have been unfamiliarity with the word, however no one asked for the definition of the word or the translated equivalent. The grammatical knowledge of the uses of this word shows that it always takes a direct object as the immediate contextual element. Since the gap is followed by the determiner (the) and the noun (appendixes), this knowledge could have been useful with this gap. The problem could have occurred in the recognition of the head word in several postmodifiers in the long complex noun phrase which is the subject, which would have helped the students to connect the associative meanings of charts with appendixes. World knowledge suggests that glossary, bibliography and charts in books usually appear at the end as appendixes. Since the usual indicative structure of an English clause is subject followed by finite, the student should have recognized the missing element as a verb in this clause. The options offered were the following verbs: approach, evolved and designed. All these three options mostly need a preposition either as their immediate contextual element or in their vicinity. Since parts of the book are described in the simple present tense (sets the tone; treats; discusses), the same logic can be applied with this example. This means that the past tense form of 'evolve' (evolved) cannot be used here. The knowledge of the semantic meaning of 'approach' also excludes this word from options. Even if the student did not know the meaning of this word, additional grammatical and common knowledge could have been helpful.

Target

The utilization of linguistic knowledge in the control of the sounds and structures of _____ languages.

The problem with 'target' might have occured because of the students' knowledge of the adjective's most common forms (ant, ous, al, ful, ing, ed, ish,…) to which 'target' does not fit in. Students probably treated this 'target' as a noun. As a noun it appears less frequently in concordances (123 occurrences) than as a verb (406 occurrences), which might be the reason students treated this word as a verb rather than a noun. There is one instance found in which 'target' is combined with 'language' and with 'knowledge' in the vicinity, which can be transferred to the example in focus.

Utilization

The _____ of linguistic knowledge in the control of the sounds and structures of target languages.

The determiner 'the' indicates the use of a noun (head) in the gap of the complex noun phrase. Although concordances reveal only one example of 'utilization' collocating with 'knowledge', this is enough to exclude other options for this gap. Another signal that a noun should be used here is the 'of-phrase', an equivalent of the genitive used to express the relation between nouns (utilization and lingusitic knowledge).

5. Conclusion

The paper attempts to illustrate one of the ways in which we can develop academic vocabulary with AWL gap maker and concordances. The text chosen for a gap-fill exercise is a short book review using professional and formal language, and several core academic words. The gap-fill exercise was designed at two different levels and solved by the 1st year students of English at the Faculty of Arts in Maribor in two

groups, each working with one difficulty level. The higher the level, the higher the number of deleted words in the text. The aim was to test students' existing knowledge of academic vocabulary in context, and to enrich it through the analysis of word in context/co-text relations using a concordancing program. This means that the results of both groups are compared in terms of the number of (in)correct entries with the major focus on the most common problematic entries of both groups. The results show that the number of correct entries at level 3 was higher than at level 7, which is due to more context. However, there are also occasions, where despite there being more context, the number of correct entries at level 3 was lower than at level 7.

Possible reasons for these and other wrong entries are provided on the basis of contextual clues, which include grammatical knowledge of sentences or/and the knowledge of the domain (world knowledge), which could be helpful in choosing correct entries. The use of contextual clues is checked with concordances, which display how the English language behaves in real use (in a context). This way, the paper illustrates how important it is for students to be guided in their search for different grammatical (linguistic) and other contextual clues to make correct inferences, and in how to use concordances to support or reject the claims by looking at collocates. By doing this, students become actively involved in the process of developing academic vocabulary by exploring the behaviour of academic words in context and their collocates. By noticing identical or similar combinations, students can come to conclusions about the use of academic words, which also involves the development of implicit knowledge.

References

BARBIERI, F. and S. ECKHARDT. 2007. "Applying corpus-based findings to form-focused instruction". *Language Teaching Research* 11, no. 3: 319-46.

BLAGANJE, D. and I. KONTE. 1987. *Modern English Grammar*. Ljubljana: Državna založba Slovenije.

CARTER, R. 1998 (2nd edition). *Vocabulary – applied linguistic perspectives*. London and New York: Routledge.

CHERN, C. 1993. "Chinese students' word-solving strategies in reading in English." In Second Language Reading and Vocabulary Learning, eds. T. Huckin, M. Haynes and J. Coady: 67-85.New York: Ablex Publishing Corporation.

COXHEAD, A. 2000. "A New Academic Word List". *TESOL Quarterly*, 34(2), 213-238.

CUMMINS, J. 1979. Cognitive/academic language proficiency, linguistic interdependence, the optimum age question and some other matters. In *Working Papers on Bilingualism*, 19: 121-129.

DEYES, A. 1984. "Towards an authentic discourse cloze." In *Applied Linguistics*, vol.5, no. 2: 128-37.

FRANK, M. 1972. *Modern English: a practical reference guide*. New Jersey: Prentice Hall.

GOATLY, A. 2000. *Critical Reading and Writing*. London and New York: Routledge.

HUCKIN, T.N. and Z. JIN. 1987. "Inferring word-meaning from context: a study in second language acquisition." In: *Proceedings of the third eastern states conference on linguistics*, eds, Marshall, F., A. Miller and Z. Zhang. Columbus: The Ohio State University.

LEECH, G. and J. SVARTVIK, 1994 (2nd edn). *A Communicative Grammar of English*. London and New York: Longman.

MARZANO, R. and PICKERING, D. 2005. *Building academic vocabulary*: *Teacher's manual*. Alexandria, VA: ASCD.

NAGY, W. 1997. "On the role of context in first- an second-language vocabulary learning." In: *Vocabulary: Description, Acquisition and Pedagogy*, eds. Schmitt, N. and Mcarthy, M. 1997, CUP, Cambridge.

NATION, I.S.P. 1990. *Teaching and Learning Vocabulary*. Boston, MA: Heinle and Heinle.

OLLER, J.W. 1979. *Language Tests at School*. Harlow: LongmanPotter, L. 1999. 1999. *A Guide to Using the COBUILDDIRECT Corpus*. The University of Birmingham: The Centre for English Langauge Studies.

RAED, J. 1997. Vocabulary and testing. In: *Vocabulary: Description, Acquisition and Pedagogy*, eds. Schmitt, N. and Mcarthy, M. Cambridge: Cambridge University Press.

RICHARDS, J.C. 1976. "The role of vocabulary teaching." *TESOL Quarterly* 10 (1): 77-89.

RUMBERGER, S., and SCARCELLA, R. 2000. "Academic English." *Linguistic Minority Research Institute Newsletter*, 1, 1-2. University of California, Santa Barbara.

POVZETEK

Razvijanje akademskega besedišča s pomočjo awl gap makerja in korpusnih podatkov

Pri študentih angleščine kot tujega jezika (za pedagoške in nepedagoške namene) narašča potreba po razvijanju akademskega besedišča – razviti ga morajo vsaj do določene mere, če je eden izmed glavnih ciljev študija ustvariti uspešne uporabnike angleščine. Od študentov se pričakuje, da berejo in razumejo zahtevnejša strokovna in znanstvena besedila, ne zgolj z namenom uspešnega zaključka študija, pač pa, da postanejo del izobražene družbe in tako zmožni soočati se z zahtevami v smislu znanja na svojih poklicnih poteh. Branje in razumevanje strokovnih besedil je lahko zelo zahtevna aktivnost zaradi določenega števila besed akademskega besedišča, ki ga takšna besedila vključujejo. Članek podaja primer uporabnosti naloge z zapolnjevanjem vrzeli v strokovnem besedilu kot ene izmed mnogih možnosti za razvijanje in preverjanje akademskega besedišča v kontekstu, sestavljene s pomočjo AWL gap makerja. Nalogo za zapolnjevanje vrzeli v besedilu so reševali študenti istega letnika v dveh skupinah in je le-ta bila nastavljena na različni težavnostni stopnji pri istem besedilu zaradi dveh razlogov: preveriti obstoječe znanje o akademskem besedišču in obogatiti besedišče skozi analizo besed v odnosu s kontekstom z uporabo programa konkordanc. Članek primerja odgovore obeh skupin z namenom ugotoviti najbolj težavna mesta zapolnitev (nepravilne odgovore) in podati možne razloge za le-te. Poleg tega so upoštevani kontekstualni indici kot koristni viri pri ugotavljanju pravilnih odgovorov; preverjeni so s pomočjo programa konkordanc. Poleg poudarka na uporabnosti takšne naloge v kombinaciji s korpusom za razvijanje in obogatitev akademskega besedišča, članek poudarja pomembnost vodenja študentov v smislu, kako iskati različne slovnične in druge ko(n)tekstualne indice, ki jim lahko pomagajo pri ugotavljanju pravilnih zapolnitev, ter v smislu, kako uporabljati konkordance, da le-te podprejo ali zavržejo odgovor.

ABSTRACT

Developing Academic Vocabulary with AWL Gapmaker and Corpus Data

There is a growing need for the development of at least some academic vocabulary in students of EFL (studying English for teaching and non-teaching purposes), if one of the main aims of EFL study is to produce competent users of English. Students are expected to read and understand advanced professional texts, not simply to successfully complete their degrees but to become a part of educated society, and to be able to meet the demands placed upon their knowledge in a prefessional environment. The reading and understanding of professional texts can be a highly demanding activity because of the number of academic and professional terms such texts include. This paper exemplifies the usefulness of a gap-fill exercise, as one among a range of options for developing and testing academic vocabulary in a professional context. The exercise was compiled with AWL gapmaker based on a professional review text, which includes academic items of vocabulary. The gap-fill exercise was used with two groups of EFL students in the same year of study and installed at two difficulty levels for two reasons: to test students' existing knowledge of academic vocabulary, and to enrich it through the analysis of words in context/co-text relation using a concordancing program. The paper compares the answers of both groups to determine the most problematic entries (incorrect answers), and to provide the possible reasons for them. Further, contextual clues are taken into consideration as a helpful source for determining correct entries, and tested with the concordancing program. Besides the emphasis on the usefulness of such gap-fill exercises combined with corpus for the development and enrichment of academic vocabulary, the paper also emphasizes the importance of guiding students in how to search for different grammatical and other contextual clues to make correct inferences, and how to use concordances to support or reject the claims.