Spoken lexical chunks used by successful learners at B2 level: forms and functions

Christian Jones, Senior Lecturer in TESOL at the University of Liverpool, UK **Daniel Waller,** Senior Lecturer in EFL and Testing at the University of Central Lancashire, UK **Patrycja Golebiewska,** Associate Lecturer in EFL at Jiangnan University in Wuxi, China.

Abstract

This study analyses the lexical chunks used in successful spoken language by a sample of learners at Common European Framework of References for Languages (CEFR) B2 level in the speaking component of a four-skills test. The data forms the initial part of a corpus of speaking tests at this level and the intention is to provide practitioners and researchers with evidence in order to more clearly interpret the broad 'can do' statements of the CEFR. Results suggest that the most frequent chunks used at this level employ the first thousand most frequent words in the British National Corpus (BNC) spoken lists and are often comparable to frequent chunks used in native speaker corpus data. The findings also show that learners make use of a narrow range of multi-functional chunks rather than employing a range of items for one function. Moreover, some chunks we normally would expect to find at this level were underused. The data has implications for how spoken language, in this case, chunks, could be taught to students at this level and for further research in this area.

Introduction

Interest in the written and spoken performance data gained from language learners has grown considerably in recent years with the development of large-scale corpora of learner data. Analysis of such corpora, and of learner spoken corpora in particular, has led to questions about how learner performances vary from those of native speakers as well as how they change at different levels of proficiency. Projects such as English Profile have attempted to identify features of learner performance at different levels of the CEFR with a view to assisting educationalists in course, material and assessment design as well as teachers and learners in identifying the fundamentals required for levels of proficiency. The current study builds on previous work carried out at the University of Central Lancashire (see Jones, Waller and Golebwieska 2013), where a corpus of spoken learner data was used to identify features of successful performance at level B2. This study extends this investigation to the question of which chunks of language successful learners at the B2 level use by exploring the two following research questions:

RQ1. What were the most frequent three- and four-word lexical chunks in the learner test corpus?

RQ2. How did common lexical chunks function in this context?

Literature review

The prevalent nature of lexical chunks in native speakers' linguistic behaviour has been extensively researched (e.g. Firth 1957; Fillmore 1979; Hymes 1962; Nattinger and DeCarrico 1992; Schmitt and Carter 2004; Sinclair 1991, 2004; Wray 1999, 2000, 2005) and is nowadays widely agreed on. Corpora such as the Collins Birmingham University International Language Database (COBUILD), the British National Corpus (BNC) and the Cambridge English Corpus (CEC) have revealed that words recur in multiple units and that language production relies to a great extent on the retrieval of prefabricated chunks which have a semantic (i.e. pig in a poke, to have a party, heavy rain) or a syntactical (i.e. of course, in order to, despite of) meaning (Moon 1997:43). Studies such as Erman and Warren (2000) and Foster (2001) suggested that formulaic sequences constitute 58.6% of the spoken and 52.3% of the written texts examined, further demonstrating the extensive use of multi-word items in native speakers' discourse.

While the pedagogical value of lexical chunks has been emphasised (e.g. Lewis 1993, 1997, 2000), it is also clear that the number of existing multi-word items can pose a problem regarding their selection for teaching. At each level of the CEFR it is essential to consider their 'teachability' and usefulness for learners. One possible solution is to look at corpora as a basis for selection. Recently, a number of publications have used native speaker corpus data to illustrate common features of spoken English including the high frequency chunks (e.g. Biber, Johansson, Leech, Conrad and Finegan 1999; Carter and McCarthy 1997, 2006; McCarthy and Carter 1995, 2001). While such analysis has proved influential within English language teaching (ELT), there has also been some criticism. Some have suggested that in a world where English is increasingly being used as a lingua franca (Kirkpatrick 2007; Seidlhofer 2004, 2011) corpus data based on the speech or writing of native speakers may not always provide the most appropriate model for learners wishing to be functionally successful users of English or, as Prodromou (2003, 2005) terms them, "SUEs"

As a result, a large number of learner corpora have also been developed (see Universite Catholique de Louvain, 2012, for a helpful list), which demonstrate how learners of English in various contexts use English successfully and the errors they make. The

International Corpus of Learner English (Granger, Dagneaux, Meunier and Paquot 2009), for example, has 3.7 million words of written English authored by high-level learners with a wide variety of first languages.

Although learner corpora of this kind are helpful in describing what successful users of English do and the errors they may make, they do not generally directly link into the CEFR levels and competencies, and as such cannot always make the *can do* statements easier to interpret. The

English Profile project is an example of research which has begun to address this issue. The English Profile use the Cambridge English Profile Corpus (CEPC, 2013), a corpus of learner English at A1-C2 levels, to provide valuable insights into language use at the different levels of proficiency.

The data used in this study are not intended to replicate the work of the English Profile project. Rather, the intention is to create a small, spoken corpus focused on the UCLan/ESB examinations and to profile *only* successful language use in these tests. These data will be used to inform learners, teachers, oral assessors and test writers of features which successful learners are using, and the ways in which they are achieving the CEFR *can do* statements at this level.

We recognise that successful English use at B2 level cannot of course be solely defined by how learners perform in a speaking test. However, passing English examinations is a goal for many learners and B2 is also a crucial defining level, as it is at this stage at which students can gain entry into many higher education institutions in the UK and other English speaking countries. Data of this kind can help to inform which chunks we might teach to B1 level learners, whose goal is to reach B2 level.

Methodology

The data used to answer the research questions in this study are based upon a 17, 827 word sample of the UCLan/ESB spoken tests at B2 level. Although the sample size is, at this stage, relatively small, we felt that there were enough data to provide some initial answers to our research questions. The samples were taken from 32 candidates (12 males and 20 females) in Greece, each of whom passed the speaking section of the test. Candidates undertake the tests in pairs and are marked by an examiner and interlocutor, the former who assigns an analytic score and the latter who assigns and impressionistic holistic score. Candidates are scored on pronunciation, grammar, vocabulary, discourse

management and interactive ability on a scale of 0-5 and each criterion is weighted equally. The pass mark is 2.5 and each candidate sampled obtained a mean score of either 3.5 or 4.0, as we wished to profile learners who clearly displayed successful language use, according to the test marking criteria.

The global descriptor for the CEFR suggests that at the B2 level a learner can interact "with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party" (Council of Europe 2001: 24) and each aspect of the test is

aimed at assessing candidates' ability to do this in a range of interactional formats. This is necessary as O'Sullivan, Weir and Saville (2002) suggest that particular formats of oral examination will elicit different functional profiles.

Part A serves in part as a warm-up and features the kind of everyday questions which form part of much regular interaction. The interaction is, as O'Sullivan, Weir and Saville (2002) suggest, largely informational. Part B assesses the ability of candidates to interact without interlocutor support, thereby eliciting what O'Sullivan, Weir and Saville (2002) and Lazaraton & Frantz (1997, cited in Galaczi, and Ffrench 2011:114) have found to be a far wider range of functions. Part C aims to push candidates towards their linguistic ceiling with questions which challenge them to explore the topic of part B in greater depth through the use of questions that are intended to widen out the discussion into more abstract areas.

Each test was transcribed using the conventions of the CANCODE corpus, which have been widely published (e.g. Carter 2004; O'Keeffe, McCarthy and Carter 2007). Interlocutor questions were removed when the data were analysed, alongside all proper nouns, as it was felt that these items would be of particularly high frequency but not generalizable to other contexts and may also, potentially, compromise the anonymity of participants.

The data were analyzed to find the most frequent three- to four-word chunks. The tool used for this initial analysis was the freely available Lextutor (Compleat Lexical Tutor 2014).The software is not able to distinguish between those chunks which are syntactically whole (e.g. *Sorry to bother you)* and those which are not (e.g. *you help)* and simply searches for combinations of two, three and four words which frequently cluster together. This has led some researchers to label them "lexical bundles"(Biber et al 1999) or "clusters" (Handford 2011) but we have opted for the more frequently used term

"chunks" following the definitions used by O'Keeffe, McCarthy and Carter (2007). This means that the samples produced are sometimes recognizable as intact phrases and are sometimes strings of words commonly found together. The chunks were also analysed using the freeware corpus tool AntConc (Anthony 2012) which allowed us to more closely investigate the collocates and chunks which appeared in the learner data. These have been compared throughout to a larger reference corpus, in this case the ten million word spoken section of the British National Corpus, accessed through the BYU-BNC interface (Davies 2013).

Results and discussion RQ1. What were the most frequent three- and four-word chunks in the learner test corpus?

A search in LexTutor revealed that the following chunks were the ten most frequent in the corpus.

These are shown alongside the most frequent three- and four-word chunks from the spoken section of the BNC, as described in Adolphs and Carter (2013 :29-30)

Table 1

The most common three- and four-word chunks in the learner corpus and BNC spoken corpus

4 word chunks (learner corpus)	4 word chunks (BNC spoken)	3 word chunks (learner corpus)	3 word chunks (BNC spoken)

1.[27]HAVE A LOT OF

2. [27] I AGREE WITH YOU

3. [25] I WOULD LIKE TO

4. [17] A LOT OF TIME

5. [16] YES I AGREE WITH

6. [14]WITH I FRIEND ERM

7. [12] SPEND A LOT OF

8. [10] ERM I THINK ERM

9. [9] AGREE WITH YOU AND

10. [9] FRIEND AND

1. [1,163] THE OFTHE END

2. [1,103]AT THE END OF

3. [1,031] THANK YOU VERY MUCH

4. [868]I DON'T WITH KNOW WHAT

5. [667]HAVE A LOOK AT

6. [628] I THINK IT WAS

7. [625]DO YOU WANT TO

8. [601] A LOT OF PEOPLE

9. [600]I DON'T KNOW WHETHER

10. [589]IF YOU WANT TO

1. [69] I

2. [50] I AGREE WITH KNOW

3. [42] WITH I FRIEND 2. [5,728] A LOT OF

4. [33] I

BELIEVE THIS 3.[3,836] I DON'T

5. [29] HAVE A LOT THINK

6. [29] AGREE WITH YOU 4.[3,241]ONE OF

7. [29] I THINK ERM THE

8. [28] I WOULD LIKE 5. [2,462] I MEAN I

9. [25] WOULD LIKE TO 6. [2,441] DO YOU

10. [25] ERM I THINK WANT

7. [2,345] AND I THINK

8. [2,217] BE ABLE TO

9. [2,209] YOU WANT TO

10. [2,172] AT THE

A LOT OF 1.[7,015] I DOMOMENT

What is noticeable about these chunks is that there is clearly a reasonable amount of similarity between the two sets of data. For example, *a lot of* is very frequent in both sets of data, alongside chunks with the word *think.*

Some chunks such as *I would like to* and *I agree with you* occur with more frequency in the test data as a result of the test format since tests of this nature tend to ask students to formulate opinions and express preferences. What was striking about *I would like to* was that there are no samples of *I'd like to* in the learner data, which is more frequent in the BNC spoken corpus than *I would like to* (887 vs. 487 occurrences). This is likely to reflect a lack of teaching of the phonology of such a chunk.

What seems to be absent from the learner data are discourse marking chunks such as *I mean I,* which often serves a textual function (Fung and Carter 2007) to reformulate ideas and *I don't know,* which is often used interpersonally as type of hedge in native speaker data, rather than with its standard propositional meaning (O'Keeffe, McCarthy and Carter 2007). This may partly be due to the nature of the speaking test itself. Despite its interactional nature, it is likely that learners view it very differently to the unplanned conversations in native speaker data, and thus may emphasise answering questions without reformulation, rather than focus on interpersonal interaction. It is also likely to be the case that discourse markers such as *I mean I* are simply ignored in teaching materials so learners do not encounter them as items which could help their interaction. *I mean I,* for example, can be taught as part of a textual strategy of reformulation (McCarthy, McCarten and Sandiford 2006). This lack of focus on such forms has implications for the spoken production of learners even at high levels, such as the Proficient User levels of the CEFR, where learners still tend to underuse discourse markers in informal speech (Muller 2005). With regards to *I don't know,* while there are ten occurrences of this chunk in the learner data, it tends to be used to buy time, as well as truthfully addressing a proposition. In native speaker data it is used in these ways but also as a form of deliberate vagueness. This is shown in figures one and two below.

Figure 1 *I don't know: Learner corpus*

Erm that's I think that's erm it's erm I don't know erm what every

her their lifestyle erm but I don't know. Okay I think it's erm you

So do you erm d= do you do some sports? I don't know any sport I think general

Figure 2 *I don't know: BNC spoken corpus*

Well , unless I never heard I don't know but erm (pause) I went

And (unclear) Well (pause) I don't know but it's still going .

difficult to get one of them (unclear) I don't know cos there 're all together

The next step we took was exploring the frequent chunks *Erm I think/I think erm* in order to find out the most common words which collocate with these chunks and thus the patterns with which they occur. These are displayed in comparison with collocates extracted from the spoken section of the BNC. Some are already clear from the chunks listed above, but this table shows all the collocates from the learner corpus with

 five or more occurrences.

 A search for *I think erm* and *Erm I think,* revealed the following patterns, displayed in table four

below. As the collocates in the learner data did not follow a consistent pattern, we have shown those occurrences which were felt to be comparable with the BNC data with at least two samples in the learner data.

Table 2
Patterns with I think erm and Erm I think in the learner and BNC spoken corpora

Spoken test learner corpus (17, 827 words)	BNC Spoken corpus (9, 963, 993 words)
Total occurrences of 'I think' = 95 (0.53%)	Total occurrences of 'I think' = 25, 825 (0.25 %)
Total occurrences of ' I think erm' = 29 (0.16%)	Total occurrences of a ' I think erm' = 169 (0.001%)
Total occurrences of 'Erm I think (erm)' = 10 (0.05%)	Total occurrences of 'Erm I think' = 379 (0.03%)
I think erm +	I think erm +
I think erm I (4)	I think erm PAUSE(36)
I think erm WE (4)	I think erm THE (10)
I think erm IT (4)	I think erm I (7)
	I think erm IT (6)
+I think erm	+ I think erm
ERM I think erm (10)	AND I think erm (16)
AND I think erm (3)	SO I think erm (12)
BUT I think erm (2)	BUT I think erm (8)

The chunks with *I think* occur with high frequency in part due to the demands of the test, since the tasks on many occasions require students to give opinions and to discuss questions. As we would predict, the chunk is used to give personal views more than views about others.

What is also worth noting is how similar some patterns in the learner data are to other investigations of *I think.* Adolphs and Carter (2013:130-131), for instance, show that *erm I think* occurs 48 times in a corpus of almost 400, 000 words of interviews with British English speakers, while *I think erm* occurs 41 times. These occurrences account for 3.82% and 3.28% of their data,

while *Erm I think erm* accounts 0.05% in our data and *I think erm* accounts for 0.16 %. While these figures are lower for the learner data than Adolphs' and Carter's data, they are higher as a percentage of the total data when compared with the BNC, which suggests that variants of *I think* are useful chunks for learners at this level, particularly within a test format such as this. The final observation is, that as Adolphs and Carter observe in their data, *erm* does not seem to occur in the middle of the chunk e.g. *I erm think* , although filled and unfilled pauses may proceed or follow it. This suggests that *I think* is generally processed and used holistically in the learner test data, as it is in their native speaker data.

RQ2. How did the chunks function in this context?

The chunks which are highly frequent seem to be those which can function on several levels. To take the variations on *I think,* firstly, as we have noted above, it is used to simply state opinions, where it is often synonymous with *I believe.* While the need to express opinions in a test of this nature at this level is predictable, it is interesting to note that students did not generally feel the need to employ the type of stock phrases often taught, such as *In my opinion* or *As far as I'm concerned.* Variants of *I think* were also used as a multi-functional discourse marker. In these uses, it seemed to be employed to buy time, to hold the turn or to hedge an opinion or assertion, uses which have been recently described in grammars of spoken English (e.g. Carter and McCarthy 2006).

1. Hedging

<$I>Okay what about you? Same question.

<$18F>Erm yes I think erm technology isn't for rich people. All people use technology and it is very important because <$=>can erm<\$=>technology can help all people in their lives and all and not rich people can use it.

2. Buying time/holding the floor

<$22F>Because I think that this is erm erm<$E>pause < \$E>erm .

<$I> Uh-huh.

<$22F>Erm I can get together with my family and with my relatives and sometimes we celebrate together erm

I agree with you seemed to operate in a similar way. Firstly, it acted as a way to simply

agree but there was also a suggestion that it was used with a discourse marking function, to buy time whilst composing the next part of the utterance, as the following example shows. *<$31F>How's you?*

<$32F>I agree with you and erm I think that it is important to know erm<$=>where to be<\$=> where you are because erm you should get better like a person so it would be a good thing to be able

to see into the future<$=> if you erm< \$=>in this way you know exactly if you have family <$=>to< \$=> and you work to grow up your children.

Implications for teaching and research

There are several implications for both teaching and research. The first is that chunks are clearly useful for learners at B1 level seeking to reach B2 level, as these successful learners made extensive use of such items. Martinez and Schmitt (2012) detail such expressions in their phrasal expressions list, linked to different level of the BNC and with information about frequencies in spoken and written contexts.

The second implication for teaching is to teach chunks which are multi-functional e.g. *I think* rather than many chunks with a similar function. Range of expression is a source of anxiety for teachers but in fact, when it comes to the functions required for effective speaking at B2 level, what 'works' is the very predictable and repetitious, rather than the unpredictable and lexically infrequent. Teachers could highlight multi-functional chunks with samples from corpus data, such as that used in this article, to illustrate how such functions work in discourse, in order to raise learners' awareness. An extension of this would be to teach high-frequency chunks as functions explicitly connected to speaking strategies or gambits such as 'playing for time' or 'interrupting'. Concordance lines could be used with students, a method suggested by Mukherjee (2009), which would allow the learners to identify the use of the chunk and discuss how effective it might be. The approach could be used not only for exam preparation classes but to allow learners to see that these are legitimate strategies used by fluent users of the language in day-to-day conversation and interaction. In terms of research, it is clear that due to the size of the data sample, these results are only illustrative. More work is needed to expand the size of the corpus and extend across other levels of successful learners.

Conclusion

Clearly, there is more work to be done on successful spoken language at B2 and other CEFR levels. Such work should, we believe, seek to develop the other aspects of spoken language used by successful learners at various levels. This paper has offered and discussed some initial findings which can begin to inform teachers and other researchers about the kind of chunks used in this data, and which can be used alongside native speaker corpus data to analyse successful spoken language.

References

Adolphs, S., & Carter, R.A.. (2013). *Spoken Corpus Linguistics: From Monomodal to Multimodal.*

London: Routledge

Anthony, L (2012) AntConc (Version 3.3.5) Tokyo, Japan: Waseda University http://www.antlab.sci.waseda.ac.jp/software.html (Date of access: 1 November 2013) Biber, D., S. Johansson, G. Leech, S. Conrad & E. Finegan. *(1999). Longman Grammar of Spoken and Written English.* London: Longman Cambridge English Profile Corpus (2013) http://www.englishprofile.org/index.php/corpus (Date of access: 1 October 2013) Carter, R. & M. McCarthy. (1997). *Exploring Spoken English.* Cambridge: Cambridge University Press.

Carter, R. & M. McCarthy. (2006). *Cambridge Grammar of English.* Cambridge: Cambridge University Press

Carter, R. (2004). *Language and Creativity: The Art of Common Talk.* London: Routledge Compleat Lexical Tutor. (2014). http://www.lextutor.ca/. (Date of access: 15 September 2014)

Council of Europe. (2001). *Common European Framework of Reference for Language:*

Learning, Teaching, Assessment. http://www.coe.int/t/dg4/linguistic/cadre_en.asp. (Date of access: 1 January 2013) Davies, M. (2013). BYU-BNC British National Corpus http://corpus.byu.edu/bnc/ (Date of access: 30 October 2013)

English Profile .(2013) .http://www.englishprofile.org/ (Date of access: 25 October 2013) Erman, B., & Warren, B. (2000).The idiom principle and the open-choice principle *Text.*
20 (1), pp. 29-62

Fillmore, C. J. (1979) On fluency. In C.J. Fillmore, D. Kempler, and W.S.Y. Wang. (eds.) *Individual differences in language ability and language behaviour.* New York: Academic Press. pp. 85-101

Firth, J.R. (1957) *Papers in linguistics 1934-1951.* Oxford: Oxford University Press

Foster, P. (2001) Rules and routines: A consideration of their role in the task-based language production of native and non-native speakers. In M. Bygate, P. Skehan, and M. Swain. (eds.) *Researching pedagogic tasks: Second language learning, teaching, and testing.* Harlow: Longman. pp. 75-93

Fung, L., & Carter, R. (2007). Discourse markers and spoken English: Native and learner use in pedagogical settings. *Applied Linguistics,* 28 (3), pp. 410-439 Gairns, R. & S. Redman. (2002). *Natural English Intermediate Student's Book.* Oxford: Oxford University Press

Galaczi, E. & A. Ffrench. (2011). Context Validity. In L. Taylor (Ed.), *Studies in Language Testing 30: Examining Speaking* .Cambridge: Cambridge University Press. pp.112-170

Granger, S., E. Dagneaux, F. Meunier & M. Paquot. (2009). *International Corpus of Learner English v2*. Louvain-la-Neuve: Presses universitaires de Louvain Handford, M.(2011). *The Language of Business Meetings.* Cambridge: Cambridge University Press Hymes, D. (1962). The Ethnography of speaking. In Gladwin, T. and Sturtevant, W.C. (eds.) *Anthropology and human behaviour.* Washington, DC: Anthropological Society of Washington 5.pp.13-53

Jones, C, Waller, D., & Golebiewska, P. (2013). Defining successful spoken language at B2 Level: Findings from a corpus of learner test data. *European Journal of Applied Linguistics and TEFL,* 2(2), pp.29-45

Kirkpatrick, A. (2007). *World Englishes: Implications for International Communication and English Language Teaching*. Cambridge: Cambridge University Press.

Lewis, M. (1993) *The lexical approach: the state of ELT and a way forward.* Hove: Language Teaching Publications

Lewis, M. (1997) *Implementing the lexical approach: putting theory into practice. Hove: Language.* Teaching Publications

Lewis, M. (2000) *Teaching collocation: further developments in the lexical approach.* Hove: Language Teaching Publications

Martinez, R.,& Schmitt, N. (2012). A Phrasal Expressions List. *Applied Linguistics* 33, 3: 299-320

McCarthy, M., & O'Dell (2005). *English Collocations in Use* Cambridge: Cambridge University Press

McCarthy, M. & R. Carter. (1995). Spoken Grammar: What is it and How can we Teach it? *ELTJournal,* 49(3), pp.207-218

McCarthy, M. & R. Carter. (2001). Ten Criteria for a Spoken Grammar. In E. Hinkel and S. Fotos (Eds.), *New Perspectives on Grammar Teaching in Second Language Classrooms* .Mahwah, NJ: Lawrence Erlbaum Associates pp.51-75 McCarthy, M., J. McCarten & H. Sandiford. (2006) Touchstone Level 4 Student's Book. Cambridge: Cambridge University Press

Moon, R. (1997). Vocabulary connections: multiword items in English in N. Schmitt and M. McCarthy (eds.). Vocabulary: Description, Acquisition and Pedagogy. Cambridge: Cambridge University Press pp. 40-63.

Mukherjee, J. (2009). The grammar or conversation in advanced spoken learner English: Learner corpus data and language-pedagogical implications. In Aijmer, K. (ed.) *Corpora and Language Teaching.* Amsterdam: John Benjamins, pp.203-230 Muller, S. *(2005). Discourse Markers in Native and Non-native English Discourse.* Amsterdam/Philadelphia: John Benjamins

Nattinger, J., & DeCarrico, J. (1992) *Lexical phrases and Language Teaching,* Oxford: Oxford University Press

O'Keeffe, A., M. McCarthy & R. Carter. (2007). *From Corpus to Classroom.* Cambridg*e:* Cambridge University Press

O'Sullivan, B., C. Weir & N. Saville. (2002). Using Observation Checklists to Validate Speaking-test Tasks. *Language Testing,* 19(1), pp.33-56 Prodromou, L. (2003). In Search of the Successful User of English. *Modern English Teacher,* 12(2),pp. 5-14 Prodromou, L. (2005). *'You See, it's Sort of Tricky for the L2 User'. The Puzzle of Idio- maticity in English as a Lingua Franca.* Unpublished doctoral dissertation, University of Nottingham, Nottingham

Prodromou, L. (2008). English as a Lingua Franca. A Corpus-based Analysis. London: Continuum

Schmitt, N., & Carter, R. (2000) Lexical Phrases in language Learning.The *Language Teacher,* 24(8) pp.6-10

Schmitt, N. and Carter, R. (2004) Formulaic Sequences in Action. In: N. Schmitt, *Formulaic sequences: acquisition, processing and use.* Amsterdam Philadelphia: John Benjamins Seidlhofer, B. (2004). Research Perspectives on Teaching English as a Lingua Franca. *Annual Review of Applied Linguistics,* 24, 209-39

Seidlhofer, B. (2011). *Understanding English as a Lingua Franca.* Oxford: Oxford University Press

Sinclair, J. (1991). *Corpus, concordance, collocation: Describing English language.* Oxford: Oxford University Press

Sinclair, J. (2004) Intuition and annotation - the discussion continues. In: K.Aijmer and B. Altenberg *Advances in corpus linguistics. Papers from the 23rd International Conference on English Language Research on Computerized corpora,* Amsterdam: Rodopi.pp.39-59 Wray, A. (1999).Formulaic language in learners and native speakers. *Language Teaching.* 32 (4), pp. 213-231

Wray, A. (2000) Formulaic Sequences in Second Language Teaching: Principle and Practice . *Applied Linguistics.* 21 (4), pp. 463-489

Wray, A. (2005) *Formulaic Language and the Lexicon,* Cambridge: Cambridge University Press