



CaMLA Working Papers

2017-01

Validating the MET speaking test through phraseological analysis: A corpus approach to language assessment

Ute Römer, Georgia State University, Department of Applied Linguistics and ESL

Jayanti Banerjee, Trinity College London





Validating the MET speaking test through phraseological analysis: A corpus approach to language assessment

Authors

Ute Römer

*Georgia State University,
Department of
Applied Linguistics and ESL*

Jayanti Banerjee

Trinity College London

About the Authors

Ute Römer

Is currently an Associate Professor in the Department of Applied Linguistics and ESL at Georgia State University. Her research interests include corpus linguistics, phraseology, second language acquisition, and the use of corpora in language learning and teaching. She serves on a range of editorial and advisory boards of professional journals and organizations, and is general editor of the Studies in Corpus Linguistics book series (John Benjamins).

Jayanti Banerjee

Is currently the Director of Language at Trinity College London. Her research interests include: the use of corpus-based methodology to validate speaking and writing tests, the operationalization of interactional competence in speaking tests, the effect of the number of options provided in multiple-choice items on test-taking processes, and the provision of learning-oriented feedback on tests. Her work has been published in *Assessing Writing*, the *Journal of English for Academic Purposes*, the *International Journal of Listening*, and *Language Teaching*.

Table of Contents

Abstract.....	1
Background and Goals	1
Data and Methods	2
The MET Speaking Test	2
Compiling the MET Speaking Corpus.....	3
Extracting Phraseological Items from the MET Speaking Corpus	4
Results	5
Frequent N-grams across Proficiency Levels	5
Frequent Phrase-Frames across Proficiency Levels	7
Conclusion.....	11
Acknowledgements.....	12
References	13
Appendix A: Sample transcript of an MET recording (test-taker speech only) at B1 level (file B1_157).....	15
Appendix B: The top 100 3-grams across proficiency levels together with frequency and range information for each item	16
Appendix C: The top 100 4-grams across proficiency levels together with frequency and range information for each item	19
Appendix D: The top 100 5-grams across proficiency levels together with frequency and range information for each item	22
Appendix E: Frequency-sorted 3-p-frame lists with token and variant numbers.....	25
Appendix F: Frequency-sorted 4-p-frame lists with token and variant numbers.....	26

Abstract

This study uses corpus tools and techniques to analyze the language produced by test takers at four different levels on the MET reporting scale (CEFR bands A2 – C1) with the aim of providing validity evidence for the MET (Michigan English Test) speaking test. Beginning with a corpus of transcribed MET speaking tests, we performed a phraseological analysis of each level. The study investigates whether, and if so in what ways, test-taker phraseological competence differs across proficiency levels. An observed increase in the phraseological competence of test takers placed at each of the CEFR levels tested by the exam would make a positive contribution to the validity argument for the MET. Our research serves to complement work being done elsewhere to identify the language features typical at different CEFR levels. It confirms what learners can do in terms of production/use of phraseological items at each of the examined proficiency levels. The study also contributes to research on L2 learner language development by identifying particular lexicogrammatical features that emerge at each level.

Background and Goals

Kane (2011, p. 8) defines validation as the evaluation of claims made about test scores. These claims can range from the relatively simple claim that the score awarded is accurate, to the more complex claim that the score gives an indication of how the test takers will perform on real-life tasks. Evidence must be gathered for each claim that is made. The more complex the claim, the more complex the evidence required. For instance, when tests are linked to external frameworks, cut scores are defined that delimit different levels of proficiency. The claim, based on these cut scores, is that the test takers at these different levels will produce language that is different in character in terms of a range of features including lexis, sentence structures, and fluency. An evaluation of this claim is an important part of the validity argument for such tests.

Recent research in linguistics, especially in corpus linguistics, has provided evidence for a strong connection between a speaker's fluency and her/his phraseological competence. According to Nattinger and DeCarrico (1992, p. 32), “[i]t is our ability to use lexical phrases that helps us speak with fluency” (see also Pawley & Syder, 1983). Hunston and Francis (2000, p. 271) consider the pedagogical importance of this observation when they state that “patterns are essential to fluency as well as to accuracy. It is an unfortunate learner who has to think of every next word separately when uttering a sentence” (see also Schmitt & Carter, 2004). These and other

studies have demonstrated that lexis and grammar come together in forming phrases (or phraseological items) that carry meaning and hence constitute the building blocks of communication. As corpus research has shown, fluent and proficient language is highly patterned, and vocabulary and syntax are inextricably linked in forming lexicogrammatical or phraseological sequences (e.g., Erman & Warren, 2000; Hoey, 2005; Römer 2005, 2009; Sinclair, 1991, 2004; Stubbs, 2001). Other studies have described developments in the productive use of phraseological items in learner language across proficiency levels, with more proficient learners using higher numbers and a larger set of items (covering a wider range of communicative functions) than less proficient learners (Chen & Baker, 2014; Cushing Weigle & Goodwin, 2016; Leńko-Szymańska, 2014; Staples, Egbert, Biber & McClair, 2013; Vidakovic & Barker, 2010).

Phraseological analysis can hence serve as one way of assessing a speaker's language proficiency, and insights from such analysis are useful in providing empirical evidence of characteristic features of test-taker language at different score levels. In our study, we carry out a phraseological analysis of language data produced by test takers at different levels in the MET speaking test as part of the process of developing a validity argument for this test. The phraseological analysis allows us to address the following research questions:

- RQ1: In the spoken component of the MET, does test-taker phraseological competence differ across proficiency levels?
- RQ2: If there are observable differences across levels, do they point to an increase in frequency, range, and/or complexity of phraseological items from lower to higher proficiency levels?

Data and Methods

Our study is based on a corpus of transcripts of MET speaking test recordings. We used data from 720 tests produced by test takers at four different proficiency levels.

The MET Speaking Test

The Michigan English Test (MET) is a test of general English language proficiency in primarily social and workplace contexts. It has been empirically linked to the Common European Framework of Reference for Languages (CEFR; Council of Europe, 2001) (CaMLA, 2010, 2012; CaMLA, 2014a) and reports both a scaled score from 0 – 80 and a CEFR level, A2 – C1. In 2015 the test-taking population represented 33 different language backgrounds (CaMLA, 2015), the majority of whom (75.53%) were teenagers and young adults in the 13 – 25 age group. The majority of the test takers are intermediate-level English language learners and received either a B1 or B2 classification on the CEFR. The MET is a modular assessment with three modules. The core module assesses listening, reading, and grammar. Vocabulary is assessed within the reading and listening sections. This module is machine scored by CaMLA and test takers receive two scores, one for listening and one for reading. Two additional optional modules assess speaking and writing. The writing module comprises two tasks. Each task is evaluated separately by a CaMLA-certified rater, using the MET writing rating scale. The evaluations for each task are then summed and reported as a scaled score.

Introduced in 2013, the MET speaking test is available for test takers and score users who would like a measure of speaking proficiency in addition to the Listening and Reading scores that are core to the MET. It is a structured interaction between one test taker and one examiner. During the course of the 10-minute test, the test taker completes five distinct tasks: describing a picture; talking about a personal

experience; giving a personal opinion; explaining the advantages and disadvantages of an option; and persuading someone on a topic. Each task is evaluated by a CaMLA-certified rater (who is also the speaking test examiner), using the MET speaking rating scale. The evaluations are summed and reported as a scaled score.

The MET Technical Review 2009 – 2013 (CaMLA, 2014b) proposes an interpretation of MET speaking test scores that relate to the language functions the test taker can perform at each score band. Table 1 summarizes the claims.

Table 1: Proposed interpretation of MET speaking test scores

CEFR Level	Speaking: Language functions
C1	Can speak fluently and appear to make no effort to formulate their thoughts.
B2	Can speak on a variety of topics, elaborating on their ideas and providing examples.
B1	Can speak on a variety of topics related to their hobbies and their job.
A2	Can speak on routine tasks that require a direct exchange of information.

The rating scale for the MET speaking test (<http://www.cambridgemichigan.org/wp-content/uploads/2014/11/MET-RatingScale-Speaking.pdf>) provides further insights into the language expected at different proficiency levels. Among other things it describes the quantity of language produced, the range and appropriacy of the vocabulary used, and features such as pronunciation and fluency. Somewhat unsurprisingly, there is no explicit reference to phraseological competence. This is partly because this concept crosses the boundaries between language resources (vocabulary, grammatical accuracy, and grammatical complexity) and delivery (fluency as well as, possibly, rhythm). It might also be partly because the concept encompasses a complex variety of language elements (such as noticing the use of words that commonly co-occur), and examiners may find it hard to pay attention to a test taker’s phraseological competence while also paying attention to and evaluating the content of their response (see also Xu, 2015, p. 58).

This does not mean, nevertheless, that an investigation of phraseological competence is irrelevant for the validation of the MET speaking test. On the contrary, this study will offer evidence for how the different criteria on the MET speaking test rating scale work together to offer a holistic evaluation of the test taker's language use.

Compiling the MET Speaking Corpus

We used MET speaking test data to compile the MET Speaking Corpus (MET-SC). For this corpus we orthographically transcribed the oral performances of 180 test takers at each of the four score levels assessed by the MET: CEFR levels A2, B1, B2, and C1. This means that, overall, the MET-SC captures performance data from 720 English language learners. Table 2 provides an overview of the dataset with respect to gender, first language (L1), age, and CEFR level achieved. As is the case with the MET test-taking population, half of the dataset is female and the majority of the dataset (67.2%) comprises test takers who are 25 years old or younger. By far and away the majority of the performances are from L1 speakers of Spanish (96.9%). While this is a

higher percentage of L1 Spanish speakers than normally expected, the dataset is still sufficiently representative of the MET test-taking population for inferences to be drawn from the findings of this study to the population as a whole.

The recordings for our study were selected from the MET speaking test pool by stratified random sampling, resulting in a set of data files from an equal number of test takers at each of the four selected CEFR levels. It is important to note that MET speaking tests are single-rated. Ideally, all speaking tests should be double- or multiple-rated in order to ensure score reliability. CaMLA safeguards score reliability through its rigorous examiner training protocols and “each speaking test examiner is monitored annually” (CaMLA, 2014a, p. 4).

Additionally, the dataset represents performances from 27 speaking test prompts. Though the distribution of prompts is not equal, most prompts occur 10 – 20 times. This ensures that the data is unlikely to have a strong prompt effect.

A team of six Applied Linguistics graduate students performed the transcriptions. Each person was given access to a batch of audio files at each of the four score levels. The audio files contained the full test recording, including speech produced by the examiner. Only the test taker's speech was transcribed. Turns were marked with line breaks. The speakers were identified using the convention <INT> (interviewer/examiner) and <TT> (test taker). All text enclosed in angle brackets < > would subsequently be ignored by the extraction tools. A sample transcript of an oral performance of a test taker at proficiency level B1 is provided in Appendix A. Table 3 lists the transcription conventions applied in the creation of the MET-SC.

Short forms of “because” (e.g., “cause,” “cuz,” “cos”) were transcribed as “cos” in order to ensure that they could not be mistaken for another word. For example “cause” is also a noun referring to something one is willing to defend or advocate for. When test takers used their L1 in a response, the non-English words were transcribed as <FOREIGN>. This is because we were not interested in analyzing non-English language production. If test takers used non-standard words (e.g., “preparate” instead of “prepare”), these were transcribed as heard. Mispronounced words (e.g., if “table” sounds more like “teeble”) were transcribed as the intended word (i.e., “table”). This is because though phonological variation was irrelevant to this study, morphological variation was eminently relevant. With respect to numbers, only the number one was spelled out in full. All other numbers

Table 2: Demographic composition of the dataset

		CEFR Level				Total
		A2	B1	B2	C1	
Gender	Male	89	194	92	75	360
	Female	91	76	88	105	360
L1	Spanish	180	180	177	161	698
	Portuguese	0	0	3	18	21
	Malinke	0	0	0	1	1
Age	≤ 12	0	0	0	0	0
	13 – 16	38	26	34	20	118
	17 – 19	40	35	62	50	187
	20 – 22	21	39	21	14	95
	23 – 25	23	30	14	17	84
	26 – 29	23	19	18	23	83
	30 – 39	27	22	23	31	103
	≥ 40	7	8	5	22	42
	blanks	1	1	3	3	8
Total						720

were rendered using number symbols—for example, “23.” This is because “one” may occur in phrases such as “on the one hand” or “the other one.” For standardization, times were rendered as the hour on the 12-hour clock—for example, “2 pm.” Stutters or single-sound repetitions (e.g., “c-can’t”) were transcribed as the intended word (i.e., “can’t”). This is because stutters were not relevant to the focus of this study. However, self-corrections and rephrasings were relevant and were transcribed in full (e.g., “they can’t can”).

Table 3: Transcription conventions

Phenomenon	Transcribe as
hesitation	erm
back channeling	mhm
	hmm
	uhuh
okay, OK	okay
laughter	IGNORE
cause/cuz/cos (short for because)	cos
non-English words/phrases	<FOREIGN>
gonna	gonna
wanna	wanna
numbers	one, 2, 3, 4...
times	2 pm, 2:30 pm

If the transcribers encountered any speech that they were unable to decode, they were instructed to indicate this in their transcripts using square brackets and to provide the approximate point in the recording where the speech occurred (e.g., [unclear, 2:49]). These portions were reviewed by a second and sometimes a third person. Most instances of unclear speech were eventually decoded. Those that could not be decoded were marked with <unclear> in the transcript. The resulting transcripts were then saved in plain text format (character encoding UTF-8) to ensure that they would be compatible with standard corpus analysis programs.

The 720 transcript files were divided into four sets to form four sub-corpora that could be analyzed separately: MET-SC-A2, MET-SC-B1, MET-SC-B2, and MET-SC-C1. Table 4 provides an overview of the composition of the MET-SC and its sub-corpora. While each sub-corpus contains the same number of files, word counts

for the four components range from around 82,000 (MET-SC-A2) to around 135,000 words (MET-SC-C1). The average number of words produced by the test takers increases steadily from the lowest to the highest MET proficiency level, suggesting that speakers at higher levels of proficiency are able to produce speech at a higher rate than speakers at lower proficiency levels.

Table 4: The composition of the MET-SC

(Sub-) corpus	Number of files	Number of words*	Average number of words per file
MET-SC-A2	180	82,516	458
MET-SC-B1	180	107,973	600
MET-SC-B2	180	121,605	676
MET-SC-C1	180	135,789	754
MET-SC overall	720	447,883	622

*Note: These frequencies were generated using the Word List tool in *AntConc* (version 3.4.3w).

Extracting Phraseological Items from the MET Speaking Corpus

In order to measure test takers’ phraseological competence at each MET score level, we extracted frequency-sorted lists of phraseological items from the different sub-corpora of the MET-SC. The types of phraseological items included in our study are n-grams and phrase-frames. N-grams are contiguous sequences of words that are repeatedly used in a corpus, with n being a number that indicates the length of the word sequence. In our analysis of the MET-SC subsets we included sequences of three, four, and five words (i.e., 3-, 4-, and 5-grams). Examples of n-grams frequent in spoken English are *I think so* (n=3), *a lot of people* (n=4), and *you know what I mean* (n=5). Phrase-frames (p-frames) are groupings of n-grams that are identical except for one word (Römer, 2010; Stubbs, 2007)—for example, *I don’t know if*, *I don’t care if*, and *I don’t remember if* can be summarized under the p-frame *I don’t * if*. The framing elements (*I don’t* and *if*) surround a variable slot (*) that

can be filled by a range of different words. The words that fill the * slot (the so-called ‘variants’ of the p-frame) do not occur randomly but are usually syntactically and semantically related. We retrieved p-frames that correspond in length to the extracted n-grams: 3-, 4-, and 5-p-frames.

The software tools used to extract n-grams and p-frames from our sub-corpora are *AntConc* (Anthony, 2014) and *kfNgram* (Fletcher, 2002 – 2007). For the 3-, 4-, and 5-gram extraction in *AntConc*, we set both the frequency and range thresholds to 5. This meant that n-grams were only retrieved and counted if they occurred at least five times in a sub-corpus and in at least five different transcript files. In *kfNgram*, we used a frequency threshold (‘floor’) of 1 for the initial n-gram (‘wordgram’) extraction that is required for the generation of p-frame lists. This was to ensure that all possible 3- to 5-grams in each sub-corpus were retrieved and could serve as the basis for the p-frame extract. In a second step, 3-, 4-, and 5-p-frames were retrieved from these n-gram lists together with their variants, using a frequency threshold of 3 (see also O’Donnell, Römer & Ellis, 2013). This meant that a p-frame variant (e.g., *I don’t know if*) was only included in the results lists if it occurred three or more times in a sub-corpus.

Results

Our analysis at each CEFR level included 3-, 4-, and 5-grams, as well as corresponding phrase-frames of the same lengths, their type and token frequencies above a frequency threshold, and variability of p-frames (How many different items populate the * slot?). Given their low frequencies in the subcorpora, 5-p-frames are not discussed here. While type-token overviews provide insights into productivity and range of phraseological items, the length of repeatedly occurring n-grams and the variability of phrase-frames provide indications of complexity. Especially looking at p-frames and the type frequencies of their variants can help us see to what extent Sinclair’s Idiom Principle (Sinclair, 1991, 1996) is at work and how fixed or variable language units are.

Frequent N-grams across Proficiency Levels

Table 5 provides an overview of the type and token frequencies of n-grams of different spans across the level-specific MET-SC subsets above a minimum frequency threshold of five instances, and with a minimum range specification of five texts. Given the close connection

between a speaker’s fluency and her/his phraseological repertoire, we expect the number of repeatedly used n-gram types to increase with growing proficiency, hence helping higher level learners be more fluent in their second language. Our findings confirm this expectation for n-gram spans three and four. Tri-grams increase consistently from 1,283 types at the lowest MET-SC level, A2, to 2,011 types at the most advanced level, C1. The same pattern can be observed for repeatedly occurring 4-grams, of which there are only 314 different types that are used five times or more at level A2 but 526 at level C1. Five-grams are comparatively rare at all levels of proficiency but type numbers for levels B1, B2, and C1 are considerably higher than at the A2 level. Our MET-SC sub-corpora are arguably too small to carry out a reliable 5-gram analysis.

Table 5: Numbers of n-grams above a frequency threshold of 5 across MET-SC subsets

	MET-SC-A2	MET-SC-B1	MET-SC-B2	MET-SC-C1
3-gram types (tokens)	1,283 (15,064)	1,718 (21,595)	1,977 (23,353)	2,011 (23,458)
4-gram types (tokens)	314 (3,008)	498 (5,035)	513 (4,896)	526 (4,673)
5-gram types (tokens)	68 (453)	103 (826)	84 (610)	85 (580)

To better understand how n-grams contribute to a learner’s developing phraseological competence and which n-gram types learners predominantly use at each proficiency level, we examined frequency-sorted 3-, 4-, and 5-gram lists derived from our four MET-SC sub-corpora. We will discuss core results for each n-gram span in turn. Table 6 displays the twenty most frequent 3-grams in each MET-SC subset. A more detailed list of the top 100 3-grams across proficiency levels together with frequency and range information for each item is included in Appendix B. We see that of the top 20 3-grams, nine are shared across all four lists. These 3-grams (highlighted in italics in Table 6) are *I think that*, *a lot of*, *I don’t know*, *there is a*, *erm I think*, *I like to*, *my name is*, *to go to*, and *and I think*. These are all clusters that are well entrenched in learners’ minds and are core elements of their phraseological repertoire.

Table 6: Top 20 3-grams across MET-SC proficiency levels

Rank	MET-SC-A2	MET-SC-B1	MET-SC-B2	MET-SC-C1
1	<i>I think that</i>	<i>a lot of</i>	<i>a lot of</i>	<i>a lot of</i>
2	<i>a lot of</i>	<i>I think that</i>	<i>I think that</i>	<i>I think that</i>
3	<i>I don't know</i>	<i>there is a</i>	<i>there is a</i>	<i>there is a</i>
4	<i>there is a</i>	<i>I don't know</i>	<i>I don't know</i>	<i>I don't know</i>
5	<i>erm I think</i>	<i>I like to</i>	<i>I like to</i>	you have to
6	<i>I like to</i>	<i>my name is</i>	you have to	going to be
7	<i>my name is</i>	there are a	<i>erm I think</i>	<i>I like to</i>
8	erm there are	<i>erm I think</i>	so I think	I like to
9	go to the	<i>and I think</i>	going to be	my name is
10	<i>to go to</i>	to to to	<i>my name is</i>	are going to
11	erm there is	and you can	<i>to go to</i>	so I think
12	erm I don't	<i>to go to</i>	I think it's	<i>erm I think</i>
13	a good idea	you have to	are going to	one of the
14	erm in the	are going to	I can see	<i>to go to</i>
15	in the in	go to the	I prefer to	I can see
16	the the the	going to be	erm there is	it would be
17	I have to	I prefer to	<i>and I think</i>	erm there is
18	the in the	something like that	and you can	there are some
19	with my family	are a lot	a good idea	I went to
20	<i>and I think</i>	is going to	can see a	<i>and I think</i>

Among the 3-grams that are frequently used by the lowest level (A2) learners but less so by learners at higher proficiency levels are several that start with the hesitation marker *erm* (*erm there are*, *erm there is*, *erm I don't*, *erm in the*). In this context it is interesting to note that the frequency of *erm* decreases across sub-corpora as learners become more proficient and as sub-corpora get larger. At level A2, the normalized frequency of *erm* is 6,389 per 100k words, while this number goes down to 4,289 at B1, 3,612 at B2, and 3,434 at C1. Higher frequencies of *erm* at lower proficiency levels can be interpreted as a sign of lower levels of fluency. Our A2 learners (and B1 learners, though to a lesser extent) use *erm* and phrases that contain *erm* to gain planning time as they work on producing their responses to the interviewers' questions. Other 3-grams that are more frequent at lower than higher levels of proficiency include *the the the*, *to to to*, *in the in*, *the in the*, and *with my family*. These are mostly examples of repetition that may serve as another

strategy to gain the speaker planning time. If we now look at 3-grams that only appear in the top 20 list at the higher proficiency levels, B2 and C1, we notice several items that are part of evaluative expressions and can be used in argumentative contexts or to express a speaker's stance. This includes the 3-grams *so I think*, *I think it's*, *I can see*, and *it would be* (most likely followed by an evaluative adjective). As the table in Appendix B shows, the frequencies for these stance and evaluative markers are much lower in the A2 and B1 lists (or the items do not occur at all), which indicates that these are phrases that learners do not use very productively until they reach high-intermediate or advanced proficiency

Of the top 20 most frequent 4-grams in the MET-SC subsets, eight are shared across all lists and used productively by learners at all four proficiency levels (see Table 7, shared items highlighted in italics). The eight items are *erm I think that*, *in the in the*, *erm there is a*, *I think that the*, *a lot of people*, *are a lot of*, or *something*

like that, and *there are a lot*. Learners at all MET levels are able to express their opinion with *I think that* and use simple discourse structuring devices such as *there is a*, and *there are a lot (of)*. In addition to the eight shared 4-grams, learners at the lower proficiency levels (A2 and B1) frequently use extensions of the *I think that* pattern (*and I think that*, *I think that is*, *I think that it's*, *I think that erm*). Lower proficiency learners' speech also appears to be marked by frequent repetition (*you can you can*, *to the to the*) and hesitation (4-grams containing *erm*), confirming our observations on 3-grams above. Also in line with our findings for 3-grams, learners at the two higher proficiency levels, especially at level C1, frequently use evaluative or stance expressions, including *it seems to be*, *on the other hand*, and *seems to be a*. As the more comprehensive 4-gram overview in Appendix C indicates, these three items either do not appear among the 100 most frequent 4-grams in the A2 and B1 lists,

or they appear toward the bottom of the lists (with frequencies of 13 or less). More evidence on stance/evaluative expressions in advanced test takers' repertoires is provided by the 5-gram lists in Appendix D, which include *it seems to be a*, *I think it would be*, and *seems to be a very* among the most frequent 5-word sequences in the MET-SC-C1. Expressing stance by using phrases other than *I think* (and extensions) seems to be a skill that develops only at advanced proficiency levels.

At this point it is worth noting evidence of a potential training effect. A few substantial chunks occurred at all CEFR levels. For instance, the 4-gram *or something like that* was used 152 times altogether by 101 test takers. In another example, the 5-gram *think it's a good idea* was used 25 times by 23 test takers. This suggests the effect of test preparation or specific language practice, perhaps even the memorization of chunks that might be helpful during the test.

Table 7: Top 20 4-grams across MET-SC proficiency levels

Rank	MET-SC-A2	MET-SC-B1	MET-SC-B2	MET-SC-C1
1	<i>erm I think that</i>	<i>there are a lot</i>	<i>erm there is a</i>	<i>a lot of people</i>
2	<i>in the in the</i>	<i>are a lot of</i>	<i>in the in the</i>	you are going to
3	erm I don't know	I think that it's	so I think that	<i>erm there is a</i>
4	<i>erm there is a</i>	<i>a lot of people</i>	<i>erm I think that</i>	and there is a
5	and I think that	a lot of things	<i>a lot of people</i>	so I think that
6	<i>I think that the</i>	<i>or something like that</i>	<i>are a lot of</i>	have a lot of
7	you can you can	and I think that	<i>or something like that</i>	I went to a
8	<i>a lot of people</i>	<i>erm I think that</i>	<i>there are a lot</i>	<i>there are a lot</i>
9	I don't know erm	<i>in the in the</i>	erm I don't know	<i>are a lot of</i>
10	I think that is	have a lot of	you are going to	<i>in the in the</i>
11	I think that it's	<i>erm there is a</i>	have a lot of	<i>erm I think that</i>
12	a lot of money	you can you can	<i>I think that the</i>	I think that it's
13	I think that erm	<i>I think that the</i>	a lot of things	it seems to be
14	to go to the	and there is a	and I think that	<i>or something like that</i>
15	<i>are a lot of</i>	erm I don't know	I can see a	are going to be
16	<i>or something like that</i>	I think that erm	of the of the	I think it's a
17	a good idea because	to the to the	and there is a	on the other hand
18	a lot of things	a lot of money	it's a good idea	there is also a
19	have a lot of	I think that is	a lot of money	<i>I think that the</i>
20	<i>there are a lot</i>	you are going to	you don't have to	seems to be a

Frequent Phrase-Frames across Proficiency Levels

We were also interested in gauging to what extent MET test takers at different proficiency levels are aware of the variability of phraseological items of different lengths. For each of the four CEFR levels included in the MET-SC, Tables 8 and 9 list the ten most frequent phrase-frames of spans 3 and 4, together with their (up to) three most common variants. The first row of Table 8 for instance indicates that the most frequent 3-p-frame in MET-SC-A2 is *I * to*, most commonly realized as *I like to* (100 instances), *I have to* (56 instances), and *I prefer to* (45 instances). Longer frequency-sorted p-frame lists with token and variant numbers are provided in Appendices E and F. The numbers of variants per p-frame (above a certain frequency threshold) give an indication of how productive a p-frame is in a learner group's phraseological repertoire. A low variant number indicates a rather fixed p-frame with only a few realizations, whereas a higher variant number indicates that the p-frame is more variable and productive.

Of the 10 most frequent 3-p-frames in each list in Table 8, five are shared across all MET proficiency levels. These 3-p-frames are *I * to*, *I * that*, *a * of*, *there * a*, and *to * to* (highlighted in italics in Table 8). While these shared p-frames may point to similarities in learners' phraseological repertoires, the variants and frequencies of variants indicate otherwise. To give just one example, the most frequent realization of *I * that* at all levels is *I think that*, but normed frequencies for this phrase are considerably higher at low than at high proficiency levels (335 per 100k words at A2 and 369 at B1, vs. 263 at B2 and 185 at C1). For some of the shared p-frames, variants preferred by learners at beginner and intermediate levels result in phrases that do not sound entirely idiomatic—for example, *I prefer that* (A2) and *I consider that* (B2). Learners at the advanced proficiency level (C1) appear to make more idiomatic choices. Among the 3-p-frames that are not shared across levels are items that contain the hesitation marker *erm* at A2 level (*erm * the*, *erm * I*, *erm * erm*, *the * erm*), confirming findings discussed in the previous section. At the higher proficiency levels we find a well-entrenched p-frame that is used to express speaker preferences: *I * like*, most commonly realized as *I don't like*, *I really like*, and *I also like*. An interesting item that C1 level learners use frequently but lower level learners do not is the noun phrase frame *the * of* in which the blank slot can be filled with a wide range of nouns, making this a very

productive and hence communicatively useful p-frame. Our test takers at level C1 use 38 different noun variants with this p-frame at least three times, accounting for 227 tokens altogether. Frequent realizations of the p-frame are *the advantages of*, *the kind of*, *the idea of*, *the back of*, *the end of*, and *the number of*. In comparison, using the same frequency threshold, B2 level learners use this p-frame 195 times with 30 variants, B1 level learners use it 138 times with 22 variants, and A2 level learners only 83 times with 16 variants. Learners at lower proficiency levels do not seem to have picked up on the usefulness (and variability) of this frame but stick to fewer realizations that they use repeatedly (30 of the 83 tokens at level A2 are *the advantages of* or *the advantage of*, most likely prompted by one of the speaking test tasks which asks about advantages and disadvantages of an option).

Of the top 10 4-p-frames displayed for each MET level in Table 9 only three are shared across all four learner lists (highlighted in italics in Table 9): *erm there * a*, *erm * is a*, *I * like to*. The most common realization of the last item in this list at levels A2, B1, and B2 is *I don't like to*, while *I also like to* and *I would like to* are most frequent at C1 level. C1 level learners only use *I don't like to* five times in the entire dataset, compared to 14 instances in the A2 and 23 in the B1 sub-corpus. Several of the 4-p-frames (and variants) that are common at A2 level but absent from the top 10 lists at more proficient levels indicate hesitation and repetition (e.g., *in * in the*, variants: *the*, *in*; *to * to the*, variants: *go*, *the*), confirming some of our observations on n-grams discussed earlier.

In addition to these general p-frame results, the more detailed p-frame lists in Appendices E and F show some interesting trends in variant development across proficiency levels. The 3-p-frame *a * of* only has five different realizations at level A2 that occur at least three times (*a lot of*, *a part of*, *a couple of*, *a picture of*, *a pair of*), whereas learners at level C1 use 17 different variants with this frame three or more times (*lot*, *kind*, *part*, *couple*, *picture*, *group*, *way*, *friend*, *painting*, *sense*, *pair*, *cup*, *bunch*, *matter*, *variety*, *discount*, *point*), making *a * of* a much more productive phraseological item in their repertoire. For the 3-p-frame *I * that*, we find eight variants above the same frequency threshold in the A2 results (*think*, *know*, *prefer*, *like*, *believe*, *say*, *love*, *remember*). Except for *think*, which takes the lion's share of the 306 p-frame tokens, all variants only occur between three and six times. In the results lists based on the C1 sub-corpus, there are twelve variants (*think*, *know*, *believe*, *remember*, *consider*, *see*, *like*, *feel*, *can*, *understand*, *love*, *say*), several of them occurring with double-digit

frequencies. This again implies that test takers at more advanced proficiency levels use frames more productively than their lower level peers. Similar trends of increasing variant numbers from lowest to highest proficiency levels can be observed for the p-frames *to * to* (A2: 11 variants; C1: 19 variants), *the * is* (A2: 26 variants; C1: 34

variants), and *the * of* (A2: 16 variants; C1: 38 variants). *The * of* and its extension *the * of the* become increasingly productive as test takers' proficiency goes up. As they become more proficient, learners discover additional nouns that fit the variable slot in the frame and use those realizations productively.

Table 8: Top 10 3-phrase-frames (with most frequent variants) across MET-SC proficiency levels

Rank	MET-SC-A2		MET-SC-B1		MET-SC-B2		MET-SC-C1	
1	<i>I * to</i>	like 100, have 56, prefer 45	<i>a * of</i>	lot 470, couple 14, picture 9	<i>a * of</i>	lot 374, kind 19, picture 9	<i>a * of</i>	lot 323, kind 14, part 12
2	<i>I * that</i>	think 277, know 6, prefer 5	<i>I * that</i>	think 399, remember 20, know 18	<i>I * that</i>	think 320, remember 20, consider 15	<i>I * to</i>	like 110, went 65, prefer 54
3	<i>the * the</i>	the 58, in 54, of 18	<i>I * to</i>	like 121, prefer 78, went 51	<i>I * to</i>	like 125, prefer 82, went 58	<i>I * that</i>	think 251, know 19, believe 15
4	<i>a * of</i>	lot 257, part 5, couple 5	<i>to * to</i>	to 88, go 87, the 42	<i>there * a</i>	is 248, are 62, was 6	<i>there * a</i>	is 230, are 39, was 11
5	<i>erm * the</i>	in 64, erm 32, the 28	<i>there * a</i>	is 213, are 101, was 6	<i>the * the</i>	in 62, of 36, on 26	<i>you * to</i>	have 123, need 41, want 34
6	<i>erm * I</i>	erm 47, I 44, and 24	<i>the * the</i>	in 57, the 52, to 34	<i>I * know</i>	don't 232, didn't 6	<i>the * of</i>	advantages 24, kind 13, idea 13
7	<i>there * a</i>	is 147, are 47, was 10	<i>the * and</i>	city 16, school 11, people 10	<i>you * to</i>	have 124, want 40, need 31	<i>to * to</i>	go 80, to 21, get 18
8	<i>erm * erm</i>	erm 32, okay 20, because 15	<i>to * a</i>	have 46, be 22, get 18	<i>to * to</i>	go 98, to 34, erm 17	<i>I * know</i>	don't 193, didn't 7
9	<i>the * erm</i>	people 15, city 13, music 8	<i>the * is</i>	people 15, dog 14, scene 12	<i>the * of</i>	advantages 27, idea 22, rest 13	<i>it * be</i>	would 76, will 38, could 27
10	<i>to * to</i>	go 73, to 51, the 21	<i>you * to</i>	have 83, need 35, want 35	<i>I * like</i>	don't 48, really 29, I 23	<i>I * like</i>	don't 43, really 36, also 28

Table 9: Top 10 4-phrase-frames (with most frequent variants) across MET-SC proficiency levels

Rank	MET-SC-A2		MET-SC-B1		MET-SC-B2		MET-SC-C1	
1	<i>erm there</i> <i>* a</i>	is 41, are 18, was 6	there * a lot	are 70, is 19, was 4	<i>erm</i> * <i>is a</i>	there 58, it 7, this 4	the * of the	end 11, back 10, middle 7
2	in * in the	the 47, in 4	are a * of	lot 67, couple 4	there * a lot	are 43, is 17, was 3	<i>erm</i> * <i>is a</i>	there 46, this 5, it 5
3	<i>erm</i> * <i>is a</i>	there 41, this 10	I *that it's	think 65, know 4	<i>erm</i> <i>there</i> * <i>a</i>	is 58, are 5	there is * a	also 26, a 14, like 6
4	I think * the	that 37, <i>erm</i> 5, is 4	<i>erm</i> I * that	think 53, remember 4, know 3	the * of the	middle 10, rest 9, center 8	<i>erm there</i> * <i>a</i>	is 46, are 5
5	I think * is	that 28, it 8, this 7	to *to the	the 34, go 26	in the * the	in 55, on 3	<i>I * like to</i>	also 17, would 11, really 8
6	to * to the	go 27, the 15	or * like that	something 56, things 4	<i>erm</i> I * that	think 49, remember 5, guess 3	it * be a	would 20, could 11, will 7
7	a * idea because	good 24, bad 17	<i>I *like to</i>	don't 23, really 13, would 9	so I * that	think 51, guess 4	there * a lot	are 30, is 9, was 4
8	<i>erm</i> * I think	<i>erm</i> 10, well 8, and 6	and I * that	think 54, know 5	I think * is	that 23, it 16, this 5	and there * a	is 37, was 3, are 3
9	I don't * <i>erm</i>	know 33, like 4	<i>erm there</i> * <i>a</i>	is 45, are 14	<i>I * like</i> <i>to</i>	I 12, don't 11, would 10	I * to a	went 30, go 6
10	<i>I * like to</i>	don't 14, would 11, I 8	<i>erm</i> * <i>is a</i>	there 45, this 10	I * to go	prefer 21, like 16, have 4	I think * the	that 25, that's 5, it's 4

Conclusion

The aim of this project was to provide validity evidence for the MET speaking test by using corpus tools to analyze the language produced by test takers at different levels on the MET reporting scale (CEFR bands A2 – C1) with a focus on speakers' lexicogrammatical competence. Before this study, there had been no published research that analyzed the language produced by test takers placed at each of the CEFR levels tested by the MET. We prepared a corpus of about 450,000 words, comprising just over 80,000 words at level A2 and over 100,000 words at levels B1 – C1. We began our analysis of the corpus with the expectation that, as test takers progress along the CEFR continuum, their lexicogrammatical repertoire will develop. Since phraseological competence is an important indicator of the state of learners' lexicogrammar and also closely related to speaker fluency, we performed a phraseological analysis of each level captured by the MET speaking test to establish whether test-taker phraseological competence differs across proficiency levels. We also explored whether the differences observed point to an increase in frequency, range, and/or complexity of phraseological items from lower to higher proficiency levels.

In response to our first research question (RQ1), we found that, in the spoken component of the MET, test-taker phraseological competence does indeed differ across proficiency levels. More specifically, addressing RQ2, Table 5 showed a steady increase in n-gram type numbers from A2 to C1, pointing to an increase in speaker fluency with growing proficiency. Hesitation markers and repetitions were more prevalent at the lower levels (particularly the use of *erm*), indicating that learners at those levels require more time as they plan what to say next, and stance markers were present only at the higher levels. These findings point to an increase in the complexity of phraseological items from lower to higher proficiency levels (again confirming RQ2), especially a development of crucial discourse functions such as expressing evaluation. We also noticed a few unidiomatic sequences among phrases that were repeatedly used by beginning and intermediate learners (e.g., *is a lot of people, to make exercise, it's more easy*), whereas learners at C1 level appear to make more idiomatic choices. Additionally, we observed an increase in the productivity of p-frames (higher variant type numbers) at the higher levels, particularly at level C1 where learners use a much wider range of realizations with higher token numbers than learners at lower

levels do. Lower level learners tend to rely mostly on one particular realization of each p-frame (e.g., *I think that, a lot of, the advantages of*), a phenomenon that has been referred to as the “phrasal teddy bear” in learner corpus research (Ellis, 2012, p. 17).

These findings provide positive, albeit preliminary, evidence that the MET test design enables test takers to demonstrate their language proficiency insofar as this is evidenced by increasing phraseological competence. The findings also provide positive evidence that the MET rating scale, though it does not focus explicitly on the evaluation of phraseological competence, successfully distinguishes between test takers at different levels of phraseological competence. That said, the findings must be viewed with caution. The corpus prepared for this study is a very good starting point but is insufficiently large (particularly at the A2 level) for strong conclusions to be drawn. Indeed, the modest size of the corpus did not support phraseological analysis by task type. Given the length of most of the tasks (30 – 90 seconds of talk), the dataset was too small (i.e. contained too few words) for conclusions to be drawn. In order to achieve more definitive results and perform more detailed analyses, the corpus should be increased in size.

Importantly, also, we should remind readers that MET speaking tests are single-rated. This means that the corpus comprises performances that have been placed at a particular CEFR level using the judgment of a single trained examiner. We could have confirmed the official ratings by asking independently trained raters to double-rate the performances. However, the aim of this study was to provide validity evidence for the scores awarded under operational conditions. Therefore, we elected not to have the tests re-rated. Indeed, this approach ensures that any validity claims are based on the scores actually issued to test takers and interpreted by score users.

Additionally, the n-gram findings have not yet been compared to a reference corpus. In a next step, we intend to identify phraseological items that are used particularly frequently by test takers at a certain CEFR level by deriving key n-grams from each sub-corpus. Key n-grams are the result of comparing frequency n-grams derived from a target corpus (here our MET-SC sub-corpora) against frequency n-grams derived from a reference corpus. We plan to use the 10-million word spoken section of the British National Corpus (BNC-spoken) as our reference corpus. This type of analysis will allow us to not just compare n-gram lists from the different CEFR levels with each other but also compare them against a constant reference dataset. Items that

are unexpectedly frequent and hence highlighted as 'key' in any of the MET-SC subsets (compared to the BNC-spoken) may be items that are typical of learner speech and less typical of native speaker oral production. Conversely, if items are listed as negative key n-grams in the MET-SC sub-corpora, those are likely items that are characteristic of idiomatic spoken language. Those items (should the key n-gram analysis highlight any) could then be focused on in pedagogical interventions.

Despite these cautions, our research complements the work being done elsewhere to identify the language features typical at different CEFR levels (see also, the English Profile project (<http://www.englishprofile.org/>); Hawkins & Buttery, 2010; Hawkins & Filipović, 2012). It can confirm what learners can do in terms of production/use of phraseological items at each of the examined proficiency levels. Additionally, due to the scarcity of available longitudinal or cross-sectional corpora that are available to researchers, there is currently a lack of empirical developmental studies especially of spoken learner language. This research, therefore, has the potential to make an important contribution to research on L2 learner language development (Ortega & Byrnes, 2008). This is particularly the case because the MET uses a standardized elicitation protocol, thus minimizing (if not eliminating) the effect of the task and the interviewer on the language produced by the test.

Acknowledgements

We would like to thank the following people who contributed to this project: a team of MA students who transcribed the MET audio files, including Robin Cathey, Mary Davis, Romondo Davis, Ashley Ekers, Jonathan McNair, and Robert Nelson; James Garner, who supported us in checking all the transcriptions against the audio and helped us prepare them for the *AntConc* analysis; and Adam Gesicki, who prepared the files for the *kfNgram* analysis.

References

- Anthony, L. (2014). *AntConc* (Version 3.4.3w). Tokyo, Japan: Waseda University. Available from <http://www.laurenceanthony.net/>.
- CaMLA (2010, 2012). *Setting cut scores on the Common European Framework of Reference for the Michigan English Test*. Ann Arbor, MI: Cambridge Michigan Language Assessment.
- CaMLA (2014a). *Linking the Common European Framework of Reference and the MET Writing Test*. Ann Arbor, MI: Cambridge Michigan Language Assessment.
- CaMLA (2014b). *MET Technical Review 2009–2013*. Ann Arbor, MI: Cambridge Michigan Language Assessment.
- CaMLA (2015). *MET 2015 Report*. Ann Arbor, MI: Cambridge Michigan Language Assessment.
- Chen, Y-H., & Baker, P. (2014). Investigating criterial discourse features across second language development: Lexical bundles in rated learner essays, CEFR B1, B2 and C1. *Applied Linguistics*, 37(6) 849-880.
- Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.
- Cushing Weigle, S., & Goodwin, S. (2016). Applications of corpus linguistics in language assessment. In J. Banerjee & D. Tsagari (Eds.), *Contemporary Second Language Assessment* (pp. 209–223). London, UK: Bloomsbury Academic.
- Ellis, N. C. (2012). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics*, 32, 17–44.
- Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text*, 20(1), 29–62.
- Fletcher, W. H. (2002–2007). *KfNgram*. Annapolis, MD: USNA. Available from <http://www.kwicfinder.com/kfNgram/kfNgramHelp.html>.
- Hawkins, J. A., & Buttery, P. (2010). Criterial features in learner corpora: Theory and illustrations. *English Profile Journal*, 1(1), e5 (23 pages).
- Hawkins, J. A., & Filipović, L. (2012). *Criterial Features in L2 English*. Cambridge: Cambridge University Press.
- Hoey, M. (2005). *Lexical Priming. A New Theory of Words and Language*. London: Routledge.
- Hunston, S., & Francis, G. (2000). *Pattern Grammar. A Corpus-Driven Approach to the Lexical Grammar of English*. Amsterdam: Benjamins.
- Kane, M. (2011). Validating score interpretations and uses: Messick Lecture, Language Testing Research Colloquium, Cambridge, April 2010, *Language Testing*, 29(1), 3-17.
- Leńko-Szymańska, A. (2014). The acquisition of formulaic language by EFL learners: A cross-sectional and cross-linguistic perspective. *International Journal of Corpus Linguistics*, 19(2), 225-251.
- Nattinger, J. R., & DeCarrico, J.S. (1992). *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.
- O'Donnell, M. B., Römer, U., & Ellis, N. C. (2013). The development of formulaic sequences in first and second language writing: Investigating effects of frequency, association, and native norm. *International Journal of Corpus Linguistics*, 18(1), 83-108.
- Ortega, L., & Byrnes, H. (Eds.). (2008). *The Longitudinal Study of Advanced L2 Capacities*. New York: Routledge.
- Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and Communication* (pp. 19–227). London: Longman.
- Römer, U. (2005). *Progressives, Patterns, Pedagogy. A Corpus-Driven Approach to English Progressive Forms, Functions, Contexts and Didactics*. Amsterdam: Benjamins.
- Römer, U. (2009). The inseparability of lexis and grammar: Corpus linguistic perspectives. *Annual Review of Cognitive Linguistics*, 7, 140–162.
- Römer, U. (2010). Establishing the phraseological profile of a text type: The construction of meaning in academic book reviews. *English Text Construction*, 3(1), 95–119.
- Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.), *Formulaic Sequences: Acquisition, Processing and Use* (pp. 1–22). Amsterdam: John Benjamins.
- Sinclair, J. McH. (1991). *Corpus Concordance Collocation*. Oxford: Oxford University Press.



- Sinclair, J. McH. (1996). The search for units of meaning. *Textus*, IX (1), 75–106.
- Sinclair, J. McH. (2004). *Trust the Text: Language, Corpus and Discourse*. London: Routledge.
- Staples, S., Egbert, J., Biber, D., & McClair, A. (2013). Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for Academic Purposes*, 12, 214–225.
- Stubbs, M. (2001). *Words and Phrases. Corpus Studies of Lexical Semantics*. Oxford: Blackwell.
- Stubbs, M. (2007). Quantitative data on multi-word sequences in English: The case of the word world. In M. Hoey, M. Mahlberg, M. Stubbs, & W. Teubert, *Text, Discourse and Corpora* (pp. 163–189). London: Continuum.
- Vidakovic, I., & Barker, F. (2010). Use of words and multi-word units in Skills for Life Writing examinations. *Cambridge ESOL: Research Notes*, 41, 7–14.
- Xu, J. (2015). *Predicting ESL learners' oral proficiency by measuring the collocations in their spontaneous speech*. Unpublished doctoral dissertation. Iowa State University, Ames, IA.

Appendix A: Sample transcript of an MET recording (test-taker speech only) at B1 level (file B1_157)

<INT>

<TT> Erm it's [TT NAME]

<INT>

<TT> [TT NAME]

<INT>

<TT> No it's good.

<INT>

<TT> Erm okay erm maybe there are it's a clothes shop? Where there are a lot of people maybe buying or or look for erm gift or erm clothes that they need. There are erm two men erm showed erm the man erm elegant suit. There are a woman maybe erm buy anything for her husband or for her <false start> family in the cash register. And also there are two women, women, sorry. There are two women erm watching looking for a dress for erm maybe a party or

<INT>

<TT> Erm I like to buy in a shop that's in erm Curridabat. It's a shop that have erm various erm clothes. Maybe shirts, t-shirts, and erm shoes. And erm the clothes that are here, that are there, are very cheap and well also there are expensive clothes there. The cheaper, the cheapest, it's about five hundred no five thousand Colones <currency in Costa Rica>. Yeah. So I like to buy here and another shop that is erm next to erm next to the shop. The names are Pull and Bear and Bershka. They are the same company that has these two shops and Bershka

<INT>

<TT> Okay.

<INT>

<TT> erm I like to dress casual. Yes? Casual because <unclear> I like I don't like the clothes very white or something like that. And yeah also I like to dress formally if I have a meeting, if I have a explanation in the university. In university I have to wear these clothes. Like a good erm belt. Yeah but actually I like to wear clothes casual, pants, jeans. I don't like long erm sleeves. I don't want to wear erm long erm shirts long sleeves. I prefer short sleeves. Yeah? Erm that's

<INT>

<TT> That's

<INT>

<TT> Erm I like or I think that one of disadvantages maybe it's that he can be a famous, become a famous artist. He can make erm he can makes a lot of erm paints, a lot of retracts I don't know if erm one disadvantages is that maybe he can go to another country and stay long of her of his family. They can't talk to talked to them by phone by internet by I don't know but Skype to see face to face. But yes this is the most disadvantages that he has, that he has. Erm but it's good because he becomes erm famous, he can erm win a lot of money with these paints. Erm also he can showed his paints for a another famous artist and <false start> erm describe their paints much more erm

<INT>

<TT> Well erm that I understand it that you have to work in teams right?

<INT>

<TT> Work in teams I think that it's better because you can makes a lot of opinions in your group. erm you can mix the persons better so you can do the work as much you want or you can prepare more the the topic that you that you have to to up to yes increase. But maybe you erm you maybe have problems but with a group you they you and your partners maybe can help in the other help the other partners that is disagree with you or an opinion that you show them. And I don't know maybe it's erm it's better work in groups. Yeah? Because alone you have to do all and in groups you have to do this the other do that and

<INT>

Appendix B: The top 100 3-grams across proficiency levels together with frequency and range information for each item

Rank	MET-SC-A2			MET-SC-B1			MET-SC-B2			MET-SC-C1		
	N-gram	Freq	Range	N-gram	Freq	Range	N-gram	Freq	Range	N-gram	Freq	Range
1	I think that	277	100	a lot of	470	133	a lot of	374	127	a lot of	323	115
2	a lot of	257	89	I think that	399	120	I think that	330	113	I think that	254	104
3	I don't know	200	76	there is a	213	96	there is a	247	105	there is a	230	101
4	there is a	148	65	I don't know	203	75	I don't know	214	79	I don't know	154	81
5	erm I think	107	70	I like to	121	63	I like to	125	65	you have to	123	61
6	I like to	101	52	my name is	118	118	you have to	124	59	going to be	110	45
7	my name is	100	99	there are a	101	56	erm I think	107	71	I like to	110	57
8	erm there are	76	49	erm I think	97	66	so I think	104	60	I think it's	109	67
9	go to the	75	51	and I think	92	56	going to be	102	51	my name is	108	106
10	to go to	73	50	to to to	88	46	my name is	102	99	are going to	104	44
11	erm there is	71	45	and you can	87	52	to go to	98	61	so I think	97	63
12	erm I don't	69	47	to go to	87	58	I think it's	96	56	erm I think	96	69
13	a good idea	67	44	you have to	83	44	are going to	88	44	one of the	83	52
14	erm in the	63	53	are going to	82	43	I can see	83	44	to go to	80	50
15	in the in	58	45	go to the	80	48	I prefer to	82	53	I can see	78	41
16	the the the	58	36	going to be	79	41	erm there is	76	49	it would be	76	43
17	I have to	56	36	I prefer to	78	53	and I think	71	48	erm there is	75	46
18	the in the	54	43	something like that	77	46	and you can	69	42	there are some	74	50
19	with my family	54	40	are a lot	74	49	a good idea	68	43	I went to	65	49
20	and I think	52	36	is going to	69	38	can see a	66	43	and I think	59	47
21	to to to	51	34	think that it's	65	38	in the in	66	42	I don t	59	14
22	I can see	49	27	I think it's	63	40	we can see	65	27	seems to be	59	37
23	erm erm I	47	39	so I think	63	48	you can see	65	30	and there is	58	44
24	there are a	47	37	and there is	62	43	there are a	63	38	you are going	56	22
25	erm I	46	38	erm there are	61	47	the in the	62	37	you can see	55	31
26	I prefer to	45	36	lot of people	61	46	erm I don't	60	40	I prefer to	54	42
27	erm you can	43	25	erm there is	60	46	you don't have	60	44	lot of people	54	39
28	think that the	42	27	in the in	60	41	go to the	59	45	this is a	54	45
29	and I like	41	29	a good idea	57	40	I went to	58	43	erm there are	53	43
30	and you can	41	27	lot of things	57	38	it is a	58	34	to be a	53	40
31	are going to	41	18	or something like	57	36	something like that	58	32	and it was	52	40
32	this is a	41	30	with my family	57	40	erm there are	55	40	a good idea	51	38
33	okay erm I	40	29	the in the	56	37	is going to	55	37	I have to	51	34
34	erm I like	39	32	I can see	55	31	with my family	55	35	I think it	50	38
35	I don't have	39	29	erm I don't	54	32	it would be	51	30	it was a	50	37
36	I think it's	39	31	I don't like	53	32	it's a good	51	34	I would say	49	34
37	you can you	39	17	you can you	53	27	and there is	49	39	is going to	49	30
38	I think erm	37	30	the the the	52	31	are a lot	47	31	but I think	48	42
39	I don't like	36	26	erm you can	51	33	I don't think	47	39	not going to	48	26
40	I I I	36	26	have a lot	51	33	if you are	46	27	is that you	47	32

Rank	MET-SC-A2			MET-SC-B1			MET-SC-B2			MET-SC-C1		
41	I like the	36	25	I went to	51	37	lot of people	46	35	that you can	46	35
42	something like that	36	24	I have to	49	33	I have to	43	29	well I think	46	37
43	you have to	36	22	we can see	49	22	I think it	43	33	part of the	44	36
44	don't know erm	35	26	to have a	46	37	is that you	43	28	you don't have	44	28
45	is a good	35	25	because you can	45	31	not going to	43	29	and you can	43	33
46	and I don't	34	28	can you can	45	26	or something like	43	33	don't know	43	13
47	can you can	34	15	and there are	44	35	well I think	43	35	and erm I	42	33
48	I went to	33	25	think that the	44	39	have a lot	42	29	to have a	41	31
49	lot of people	33	27	can see a	43	24	don't have to	41	32	you need to	41	23
50	erm erm erm	32	20	erm in the	43	36	erm you can	41	30	be able to	40	28
51	erm erm the	32	23	you can see	43	29	with my friends	41	25	I think the	40	33
52	so I think	32	21	that you can	42	30	you are going	41	19	a little bit	39	26
53	you don't have	32	22	this is a	42	28	and it was	40	27	and there are	39	30
54	and there are	31	24	to the to	42	31	that you can	40	31	there are a	39	26
55	you can see	31	18	it was a	41	29	you want to	40	30	you're going to	39	23
56	a bad idea	30	21	we have to	39	22	think that the	39	31	go to the	38	30
57	and there is	30	25	and in the	38	30	because I think	38	30	in the in	38	23
58	it will be	30	13	have to to	38	29	because you can	38	30	it will be	38	27
59	think that it's	30	21	well I think	38	30	of the city	38	21	would be a	38	28
60	think that is	29	21	with my friends	38	29	go to a	37	27	you have a	38	32
61	with my friends	29	23	and they are	37	28	I think the	37	29	I don't think	37	26
62	erm the the	28	22	but I think	37	27	of the of	37	30	I used to	37	21
63	I go to	28	19	is that you	37	30	that kind of	37	28	in front of	37	27
64	okay I think	28	26	think that is	37	28	erm in the	36	26	okay I think	37	33
65	there are erm	28	20	you can erm	37	25	it could be	36	28	something like that	37	25
66	think that erm	28	19	think that erm	36	25	lot of things	36	26	that you are	37	21
67	erm I prefer	27	26	and I like	35	24	one of the	36	28	erm you can	36	27
68	go to a	27	19	erm I like	35	28	part of the	36	28	I really like	36	29
69	going to be	27	14	I'm going to	35	19	the of the	36	28	that would be	36	21
70	good idea because	27	24	you are going	35	23	it was a	35	27	there is also	36	24
71	is going to	27	16	you need to	35	23	we have to	35	21	we can see	36	18
72	like to to	27	21	you want to	35	27	but I think	34	30	with my family	36	27
73	lot of money	27	23	and I don't	34	27	is a good	34	24	because of the	35	30
74	there is erm	27	22	I don't have	34	24	to have a	34	27	erm I don't	35	26
75	are a lot	26	22	it would be	34	19	to to to	34	22	erm in the	35	29
76	can see a	26	18	okay I think	34	30	erm I can	33	26	go to a	35	28
77	erm if you	26	19	that kind of	34	20	I think I	33	24	I don't like	35	25
78	my family and	26	20	the to the	34	27	in the city	33	26	it was very	35	23
79	or something like	26	18	because I think	33	27	and I don't	32	25	on the other	35	26
80	you can erm	26	17	I go to	33	23	I don't have	32	24	think it's a	35	29
81	but I think	25	19	lot of money	33	28	in order to	32	16	all the time	34	24
82	have a lot	25	20	there are some	33	24	a little bit	31	22	there are two	34	28
83	I I like	25	22	a very good	32	19	and I like	31	24	you want to	34	25
84	in in the	25	22	is a good	32	28	I see a	31	19	have a lot	33	25
85	this kind of	25	19	kind of music	32	17	the things that	31	21	most of the	33	24
86	bad idea because	24	19	one of the	32	25	this is a	31	20	the other hand	33	24

Rank	MET-SC-A2			MET-SC-B1			MET-SC-B2			MET-SC-C1		
87	erm and I	24	21	to do it	32	22	you need to	31	24	the the the	33	25
88	have to to	24	19	it's better to	31	23	and there are	30	26	with my friends	33	23
89	I don't understand	24	20	the things that	31	22	I like it	30	22	the in the	32	21
90	I I don't	24	22	you can go	31	25	it will be	30	20	there are many	32	23
91	if you don't	24	12	and we can	30	20	think that it's	30	24	went to a	32	29
92	it is a	24	21	going to to	30	23	you can you	30	23	and I like	31	24
93	it's a good	24	19	I I I	30	22	can see the	29	23	are a lot	31	22
94	the things that	24	16	I need to	30	19	I really like	29	23	have to pay	31	21
95	advantage is that	23	19	in in the	30	28	if you have	29	25	if you want	31	19
96	and and I	23	19	not going to	30	24	okay I think	29	25	it has a	31	23
97	because I think	23	20	that is a	30	26	that you have	29	26	it is a	31	23
98	I like erm	23	22	you can do	30	23	advantage is that	28	25	it's better to	31	23
99	lot of erm	23	16	you're going to	30	16	going to have	28	21	the most important	31	22
100	lot of things	23	19	and it was	29	27	seems to be	28	20	think that it's	31	24



Appendix C: The top 100 4-grams across proficiency levels together with frequency and range information for each item

Rank	MET-SC-A2			MET-SC-B1			MET-SC-B2			MET-SC-C1		
	N-gram	Freq	Range	N-gram	Freq	Range	N-gram	Freq	Range	N-gram	Freq	Range
1	erm I think that	51	40	there are a lot	70	47	erm there is a	58	45	a lot of people	54	39
2	in the in the	47	37	are a lot of	67	46	in the in the	55	35	you are going to	53	20
3	erm I don't know	45	32	I think that it's	61	34	so I think that	51	34	erm there is a	46	31
4	erm there is a	41	30	a lot of people	60	45	erm I think that	49	38	I don t know	40	12
5	and I think that	38	25	a lot of things	57	38	a lot of people	46	35	and there is a	37	32
6	I think that the	37	25	or something like that	56	35	are a lot of	43	27	so I think that	33	24
7	you can you can	34	15	and I think that	54	33	or something like that	43	28	have a lot of	31	23
8	a lot of people	33	27	erm I think that	53	39	there are a lot	43	28	I went to a	30	27
9	I don't know erm	33	24	in the in the	51	34	erm I don't know	40	26	there are a lot	30	22
10	I think that is	28	20	have a lot of	48	41	you are going to	39	17	are a lot of	29	21
11	I think that it's	28	20	erm there is a	45	37	have a lot of	38	27	in the in the	28	18
12	a lot of money	27	23	you can you can	45	26	I think that the	37	30	erm I think that	27	24
13	I think that erm	27	19	I think that the	42	37	a lot of things	36	26	I think that it's	27	21
14	to go to the	27	24	and there is a	40	29	and I think that	34	26	it seems to be	27	18
15	are a lot of	26	22	erm I don't know	38	22	I can see a	33	22	or something like that	27	20
16	or something like that	26	18	I think that erm	35	24	of the of the	32	25	are going to be	26	17
17	a good idea because	24	22	to the to the	34	27	and there is a	30	25	I think it's a	26	23
18	a lot of things	23	19	a lot of money	33	28	it's a good idea	27	22	on the other hand	26	18
19	have a lot of	23	19	I think that is	33	25	a lot of money	26	21	there is also a	26	21
20	there are a lot	23	20	you are going to	32	21	you don't have to	26	22	I think that the	25	21
21	a lot of erm	22	15	a lot of time	28	23	are going to be	25	18	seems to be a	25	20
22	I think that it	21	13	I don't know i	27	19	I think that it's	25	20	and I think that	23	17
23	okay I think that	21	19	they are going to	27	17	is going to be	25	20	but I think that	22	20
24	I don't know I	20	18	but I think that	26	21	to go to the	25	21	erm I don't know	22	17
25	in a in a	20	12	So I think that	26	22	I think it's a	24	18	so I think it's	22	17
26	so I think that	20	13	to go to the	26	22	I think that is	23	17	a lot of money	21	16
27	it's a good idea	19	16	I think that it	25	16	well I think that	23	19	is going to be	21	15
28	erm I like to	18	16	well I think that	25	19	you can you can	23	17	at the same time	20	16
29	erm there are a	18	17	in a in a	24	19	I don't know I	22	14	I would say that	20	16
30	of the of the	18	13	a lot of erm	23	17	so I think it's	22	17	it would be a	20	15
31	a bad idea because	17	12	is going to be	23	16	I prefer to go	21	16	most of the time	20	17
32	I like to go	17	14	of the of the	23	20	I think that it	21	18	well I think that	20	17
33	I went to a	17	12	we can see a	23	11	if you want to	21	16	are not going to	19	12
34	is a good idea	17	13	do a lot of	22	19	a lot of erm	20	20	if you want to	19	12
35	and I don't know	16	16	I don't like to	22	14	and I don't know	20	15	in front of the	19	17
36	and there is a	16	14	okay I think that	22	20	erm I can see	20	16	a lot of erm	18	16
37	erm a lot of	16	14	I think that I	21	16	I think that you	20	16	a lot of things	18	15
38	well I think that	16	13	I think that that	21	12	they are going to	20	16	a lot of time	18	16
39	a good idea to	15	11	because I think that	20	18	I think that erm	19	17	I don't know if	18	15
40	and I like to	15	13	I think that if	20	19	okay I think that	19	16	to eat at home	18	15
41	I can see a	15	10	if you want to	20	18	on the on the	19	17	you don't have to	18	14

Rank	MET-SC-A2			MET-SC-B1			MET-SC-B2			MET-SC-C1		
42	I don't like to	15	9	is a lot of	20	17	erm a lot of	18	15	I also like to	17	10
43	spend a lot of	15	15	I think it's a	19	16	to go to a	18	14	okay I think that	17	16
44	to the to the	15	12	spend a lot of	19	17	a good idea because	17	16	there's a lot of	17	13
45	but I think that	14	12	there is a lot	19	15	are not going to	17	14	I can see a	16	13
46	I don't I don't	14	13	think that it's better	19	12	I think that if	17	14	of the of the	16	13
47	I go to the	14	12	a good idea because	18	16	m a r i	17	17	okay well I think	16	14
48	I like to to	14	13	there is a a	18	17	there is a lot	17	12	they are going to	16	13
49	I think that that	14	13	to to to to	18	10	a lot of time	16	14	erm I think it's	15	10
50	m a r i	14	14	a lot of clothes	17	13	a r i a	16	16	erm one of the	15	15
51	you are going to	14	6	and I don't know	17	14	because I think that	16	13	I don't know what	15	12
52	a n d r	13	13	of a lot of	17	15	I like to go	16	11	in a in a	15	14
53	a r i a	13	13	with a lot of	17	13	I prefer to eat	16	12	it's a good idea	15	14
54	don't know how to	13	12	and I like to	16	13	I think it is	16	11	spend a lot of	15	12
55	erm I think it's	13	12	erm a lot of	16	15	I think it's better	16	13	think it would be	15	13
56	the advantages are that	13	12	I don't know erm	16	15	is a good idea	16	14	to go to a	15	15
57	there are many people	13	13	I go to the	16	12	is a lot of	16	12	to go to the	15	14
58	with my family and	13	13	to go to a	16	14	there is a a	16	13	a good idea because	14	14
59	with the with the	13	8	you don't have to	16	12	we can see a	16	10	erm I like to	14	11
60	erm I think erm	12	9	a lot of a	15	14	with a lot of	16	16	I think that it	14	13
61	is a bad idea	12	11	think that it's a	15	12	a good idea to	15	13	I think that you	14	13
62	one two three four	12	9	erm I like to	14	13	go to a restaurant	15	11	not going to be	14	11
63	this is a good	12	10	erm there are a	14	12	like to go to	15	11	there is a a	14	10
64	with my with my	12	9	I can see a	14	9	that you have to	15	14	there is a woman	14	12
65	a lot of time	11	9	I don't I don't	14	13	there is a woman	15	14	went to a restaurant	14	14
66	and a lot of	11	10	I like to to	14	13	to the to the	15	12	you have to be	14	12
67	are going to be	11	6	is a good idea	14	13	and I like to	14	11	I don't know like	13	10
68	erm erm I don't	11	11	lot of a lot	14	13	I went with my	14	12	I think that erm	13	11
69	for the for the	11	11	that you have to	14	11	if you if you	14	12	I think that is	13	10
70	I don't know how	11	9	there is a man	14	12	it is a good	14	12	m a r i	13	13
71	I was a child	11	7	there's a lot of	14	9	not going to be	14	10	so it would be	13	10
72	I would like to	11	6	a good idea to	13	9	prefer to go to	14	12	you're going to be	13	10
73	okay erm I think	11	9	a n d r	13	13	see a lot of	14	13	a well known university	12	6
74	there are there are	11	10	I really like to	13	8	but I think that	13	11	another disadvantage is that	12	12
75	think that it's better	11	9	is taking place in	13	13	I don't know erm	13	12	erm a lot of	12	11
76	to go to a	11	8	on the other hand	13	12	I don't know what	13	11	erm there are some	12	11
77	with my family because	11	10	taking place in a	13	13	I think that I	13	12	has a lot of	12	9
78	you don't have to	11	8	there is a dog	13	13	I would say that	13	11	have to pay for	12	9
79	do a lot of	10	10	think that is a	13	10	is that you can	13	11	I like to do	12	11
80	erm erm I think	10	9	you can go to	13	10	the advantages is that	13	13	I think it would	12	11

Rank	MET-SC-A2			MET-SC-B1			MET-SC-B2			MET-SC-C1		
81	erm I can see	10	8	a bad idea because	12	12	there is a man	13	12	if you have a	12	10
82	erm in the picture	10	9	are going to be	12	8	you can see a	13	10	it's going to be	12	10
83	erm this is a	10	10	can do a lot	12	10	all the things that	12	9	one of the advantages	12	12
84	go to the beach	10	7	can go to the	12	10	also there is a	12	8	that you are going	12	9
85	I think it's a	10	8	don't know how to	12	10	going to be like	12	8	there are some people	12	11
86	I think it's better	10	8	have the opportunity to	12	7	have the opportunity to	12	9	there is a couple	12	11
87	I think that if	10	10	I can see that	12	7	I like to	12	10	and r	11	11
88	something like that and	10	10	I don't know if	12	10	I like to eat	12	8	a r i a	11	11
89	when I was a	10	6	it could be a	12	12	I think that that	12	11	and I don't know	11	11
90	and you can erm	9	7	it seems to be	12	10	if you don't have	12	11	erm I can see	11	11
91	because I think that	9	9	it's a good idea	12	11	in a in a	12	10	I prefer to eat	11	11
92	erm there are erm	9	6	or I don't know	12	10	in the middle of	12	9	I would like to	11	10
93	erm there is erm	9	9	so I think it's	12	10	it would be a	12	10	is a good idea	11	9
94	go to the university	9	8	that it's better to	12	8	like I don't know	12	10	it could be a	11	9
95	has a lot of	9	7	they have a lot	12	12	the things that you	12	8	j u a n	11	11
96	I can I can	9	7	to wake up early	12	9	there is also a	12	9	not a good idea	11	10
97	I like to	9	9	you have to to	12	9	there's a lot of	12	9	so that would be	11	7
98	I think that	9	9	all the things that	11	9	to wake up early	12	10	that you have to	11	11
99	I prefer to go	9	5	because I like to	11	10	with my family and	12	10	the end of the	11	10
100	I think that I	9	9	erm I think it's	11	10	with the with the	12	12	to be able to	11	9

Appendix D: The top 100 5-grams across proficiency levels together with frequency and range information for each item

Rank	MET-SC-A2			MET-SC-B1			MET-SC-B2			MET-SC-C1		
	N-gram	Freq	Range	N-gram	Freq	Range	N-gram	Freq	Range	N-gram	Freq	Range
1	there are a lot of	23	20	there are a lot of	67	46	there are a lot of	41	26	there are a lot of	28	21
2	m a r i a	12	12	are a lot of people	19	18	m a r i a	16	16	it seems to be a	14	10
3	are a lot of people	10	10	I think that it's better	18	11	there is a lot of	15	11	I went to a restaurant	13	13
4	erm I think that the	10	7	there is a lot of	18	15	are a lot of people	13	9	I think it would be	12	11
5	I think that it's better	10	8	a lot of a lot	14	13	I prefer to go to	12	10	you are going to be	12	9
6	when I was a child	10	6	lot of a lot of	14	13	you are going to be	11	6	m a r i a	11	11
7	erm there are a lot	9	9	I think that it's a	13	10	in the middle of the	10	7	that you are going to	11	8
8	is a good idea because	9	9	do a lot of things	12	10	can see a lot of	9	8	are a lot of people	10	9
9	erm erm I don't know	8	8	I think that is a	12	9	I like to go to	9	6	erm it seems to be	10	8
10	I don't know how to	8	8	a lot of people in	11	11	I think it's a good	9	6	prefer to eat at home	10	8
11	I don't know I don't	8	7	can do a lot of	11	9	I went with my family	9	8	a lot of people and	9	8
12	it's a good idea because	8	8	erm I think that it's	11	8	in the in the in	9	9	at the end of the	9	9
13	or something like that and	8	8	is taking place in a	11	11	scene is taking place in	9	9	erm there is also a	8	6
14	spend a lot of money	8	8	spend a lot of time	11	10	the scene is taking place	9	9	I don't know what else	8	5
15	a bad idea because erm	7	6	the scene is taking place	11	11	to go to a restaurant	9	8	I think that it's a	8	7
16	can see a lot of	7	7	they have a lot of	11	11	because there are a lot	8	7	there is a lot of	8	7
17	don't know how to say	7	7	think that it's better to	11	7	earn a lot of money	8	6	to eat at home because	8	7
18	earn a lot of money	7	6	m a r i a	10	10	I think that it is	8	8	what else can I say	8	8
19	I don't know erm erm	7	6	to spend a lot of	10	8	is taking place in a	8	8	a n d r e	7	7
20	it's a bad idea because	7	6	scene is taking place in	9	9	prefer to go to a	8	7	a n i e l	7	7
21	okay erm I think that	7	6	you can do a lot	9	7	so I think it's better	8	6	because you are going to	7	7
22	one two three four five	7	7	a lot of kind of	8	6	the in the in the	8	8	d a n i e	7	7
23	the scene is taking place	7	7	a lot of money and	8	7	there is a there is	8	6	I think that is a	7	5
24	this is a good idea	7	6	a lot of people and	8	8	they are not going to	8	7	in the middle of the	7	7
25	to spend a lot of	7	7	a lot of things that	8	7	a n i e l	7	7	in this part of the	7	7
26	a in a in a	6	5	a n d r e	8	8	all the things that you	7	5	is a good idea because	7	7
27	a lot of people and	6	6	are a lot of erm	8	8	and also there is a	7	6	it has a lot of	7	6
28	a lot of things and	6	6	because you can you can	8	5	d a n i e	7	7	seems to be a very	7	5

29	a lot of things that	6	6	erm I don't know i	8	6	in the picture we can	7	6	spend a lot of money	7	6
30	and I think that erm	6	6	have to go to the	8	6	is a good idea because	7	7	spend a lot of time	7	6
31	erm and I think that	6	5	I think that it is	8	7	is a lot of people	7	6	think it's a good idea	7	7
32	erm I don't know erm	6	6	in the middle of the	8	8	is a there is a	7	5	to spend a lot of	7	6
33	erm well I think that	6	6	so erm I think that	8	7	it has a lot of	7	5	a good idea because erm	6	6
34	I like to go to	6	6	the in the in the	8	6	it is a good idea	7	6	a lot of people that	6	6
35	I think it's a good	6	5	you have a lot of	8	8	it's a good idea because	7	7	are not going to be	6	5
36	I think that is a	6	6	a lot of things to	7	6	or something like that so	7	6	but on the other hand	6	5
37	I think that it's a	6	5	and I think that it's	7	6	think it is a good	7	6	erm on the other hand	6	5
38	I went to a party	6	5	don't have a lot of	7	6	to eat at home because	7	6	going to be able to	6	5
39	in a in a in	6	5	erm I think that it	7	5	a good idea to to	6	6	I don't think it's a	6	6
40	in the middle of the	6	6	erm there are a lot	7	6	a lot of money and	6	6	I prefer to eat at	6	6
41	ja n d r	6	6	erm there is a a	7	6	a lot of people there	6	6	I think that it is	6	6
42	scene is taking place in	6	6	erm you can you can	7	7	a lot of things and	6	6	it's not a good idea	6	6
43	spend a lot of time	6	6	I don't know how to	7	7	but at the same time	6	6	okay okay I think that	6	6
44	think it's a good idea	6	5	I don't know I think	7	7	do a lot of things	6	6	so I think it's a	6	5
45	you can you can do	6	5	I think that erm the	7	7	erm another advantage is that	6	5	so it would be a	6	6
46	a lot of money because	5	5	I think that if you	7	7	erm I think that it's	6	6	the advantages and disadvantages of	6	6
47	a lot of people erm	5	5	I think that the advantages	7	7	erm I think that the	6	6	there's a lot of people	6	6
48	a n d r o	5	5	in the in the in	7	6	have to pay a lot	6	6	they are going to be	6	6
49	a r i a n o	5	5	is a good idea because	7	7	I don't know how to	6	6	think it would be a	6	5
50	c o l a s	5	5	n d r e s	7	7	I think it is a	6	5	this part of the test	6	6
51	don't know I don't know	5	5	or something like that erm	7	7	I think that is a	6	6	to have a lot of	6	5
52	erm I don't know I	5	5	spend a lot of money	7	7	learn a lot of things	6	6	you are going to have	6	5
53	erm I think that erm	5	5	there is a man who	7	6	one of the most important	6	5	a lot of time in	5	5
54	erm I think that is	5	5	think that it's a good	7	7	or something like that erm	6	5	and erm there is a	5	5
55	erm in the in the	5	5	up early in the morning	7	6	or something like that i	6	5	because there are a lot	5	5
56	i c o l a	5	5	wake up early in the	7	6	the picture we can see	6	5	can see a lot of	5	5
57	I like to I like	5	5	a lot of people that	6	6	there's a lot of people	6	5	erm I think it's a	5	5
58	I think it is a	5	5	a lot of things and	6	5	think it's a good idea	6	5	erm there are a lot	5	5
59	I think that this is	5	5	and I think that is	6	6	don't know how to say	5	5	erm there is a woman	5	5



60	I think there are more	5	5	and there are a lot	6	5	erm I think that erm	5	5	I don t know it	5	5
61	I was a child i	5	5	because there are a lot	6	6	erm there is a a	5	5	i e l a no	5	5
62	like to go to the	5	5	erm erm I don't know	6	5	erm there is a lot	5	5	I think it's a good	5	5
63	like to I like to	5	5	erm well I think that	6	6	erm you don't have to	5	5	I think that it's important	5	5
64	n i c o l	5	5	go to the to the	6	6	I can see in the	5	5	I think that there are	5	5
65	okay okay I think that	5	5	have a lot of money	6	6	I prefer eating at home	5	5	I think that would be	5	5
66	that this is a good	5	5	I prefer to go to	6	6	I prefer to travel with	5	5	I think this is a	5	5
67	there are there is a	5	5	is a bad idea because	6	6	I think it would be	5	5	I went to a party	5	5
68	you can see a lot	5	5	is a lot of people	6	6	I think it's better to	5	5	in front of the house	5	5
69				is a man who is	6	5	I think that it's better	5	5	is not a good idea	5	5
70				or something like that but	6	6	I think the advantages of	5	5	is taking place in a	5	5
71				scene takes place in a	6	6	in my free time I	5	5	is talking on the phone	5	5
72				that it's a good idea	6	6	in the back of the	5	5	j u a n n o	5	5
73				the scene takes place in	6	6	it seems to be a	5	5	of the picture there is	5	5
74				this part of the test	6	6	okay erm I think that	5	5	on the other hand the	5	5
75				to wake up early in	6	5	or something like that and	5	5	part of the test you	5	5
76				a lot of people erm	5	5	so I I think that	5	5	talking on the phone and	5	5
77				a lot of things for	5	5	spend a lot of money	5	5	that I went to a	5	5
78				and I think that that	5	5	spend a lot of time	5	5	that there are more disadvantages	5	5
79				and I think that the	5	5	that they are going to	5	5	the disadvantages would be that	5	5
80				and in the in the	5	5	that you are going to	5	5	the last time I went	5	5
81				because they have a lot	5	5	to go to restaurants because	5	5	the last time that i	5	5
82				but I think that the	5	5	to spend a lot of	5	5	the most important thing is	5	5
83				but on the other hand	5	5	with my family with my	5	5	the scene is taking place	5	5
84				c a r l o	5	5	you have a lot of	5	5	think that there are more	5	5
85				don't know I think that	5	5				we had a great time	5	5
86				erm but I think that	5	5						
87				I can see in the	5	5						
88				I think that I I	5	5						
89				I think that the government	5	5						
90				I think that this is	5	5						
91				in this part of the	5	5						

92				it seems to be a	5	5						
93				j a n d r	5	5						
94				no okay the scene is	5	5						
95				of becoming a professional artist	5	5						
96				okay erm I think that	5	5						
97				okay the scene is taking	5	5						
98				one of the advantages of	5	5						
99				or something like that I	5	5						
100				so I think that it's	5	5						

Appendix E: Frequency-sorted 3-p-frame lists with token and variant numbers

Rank	MET-SC-A2			MET-SC-B1			MET-SC-B2			MET-SC-C1		
	P-frame type	Tokens (includes variants with at least 3 instances)	No. of variants with at least 3 instances	P-frame type	Tokens (includes variants with at least 3 instances)	No. of variants with at least 3 instances	P-frame type	Tokens (includes variants with at least 3 instances)	No. of variants with at least 3 instances	P-frame type	Tokens (includes variants with at least 3 instances)	No. of variants with at least 3 instances
1	I * to	320	14	a * of	525	9	a * of	430	10	a * of	425	17
2	I * that	306	8	I * that	487	12	I * that	425	12	I * to	423	15
3	the * the	282	27	I * to	428	11	I * to	423	14	I * that	335	12
4	a * of	273	5	to * to	333	21	there * a	321	4	there * a	283	4
5	erm * the	221	20	there * a	323	4	the * the	250	20	you * to	270	13
6	erm * I	213	16	the * the	254	19	I * know	238	2	the * of	227	38
7	there * a	209	4	the * and	217	44	you * to	237	9	to * to	214	19
8	erm * erm	201	29	to * a	201	21	to * to	218	16	I * know	200	2
9	the * erm	200	42	the * is	188	36	the * of	195	30	it * be	186	10
10	to * to	192	11	you * to	188	8	I * like	183	15	I * like	175	14
11	the * and	181	41	erm * I	178	16	I * I	182	23	to * a	170	16
12	I * I	180	18	you * you	162	20	the * and	179	36	the * and	169	35
13	the * is	130	26	I * I	159	18	to * a	177	19	the * is	165	34
14	erm * is	129	11	and * can	159	6	the * is	176	33	the * the	154	16
15	I * erm	123	12	erm * the	153	16	erm * I	156	16	erm * the	136	18

16	in * in	114	8	the * of	138	22	it * be	138	7	erm * I	131	17
17	and * the	111	12	to * erm	134	23	to * the	135	28	I * it's	129	3
18	to * a	111	15	I * like	134	8	erm * the	133	15	the * that	128	15
19	I * the	109	9	and * the	131	14	it * a	126	6	and * I	120	14
20	my * is	109	3	a * a	128	15	the * erm	124	30	my * is	118	4

Appendix F: Frequency-sorted 4-p-frame lists with token and variant numbers

Rank	MET-SC-A2			MET-SC-B1			MET-SC-B2			MET-SC-C1		
	P-frame type	Tokens (includes variants with at least 3 instances)	No. of variants with at least 3 instances	P-frame type	Tokens (includes variants with at least 3 instances)	No. of variants with at least 3 instances	P-frame type	Tokens (includes variants with at least 3 instances)	No. of variants with at least 3 instances	P-frame type	Tokens (includes variants with at least 3 instances)	No. of variants with at least 3 instances
1	erm there * a	65	3	there * a lot	93	3	erm * is a	69	3	the * of the	68	14
2	in * in the	51	2	are a * of	71	2	there * a lot	63	3	erm * is a	56	3
3	erm * is a	51	2	I * that it's	69	2	erm there * a	63	2	there is * a	51	4sw
4	I think * the	49	4	erm I * that	60	3	the * of the	58	11	erm there * a	51	2
5	I think * is	48	4	to * to the	60	2	in the * the	58	2	I * like to	50	6
6	to * to the	42	2	or * like that	60	2	erm I * that	57	3	it * be a	45	5
7	a * idea because	41	2	I * like to	59	6	so I * that	55	2	there * a lot	43	3
8	erm * I think	37	6	and I * that	59	2	I think * is	49	4	and there * a	43	3
9	I don't * erm	37	2	erm there * a	59	2	I * like to	48	5	I * to a	36	2
10	I * like to	37	4	erm * is a	55	2	I * to go	48	5	I think * the	34	3
11	you * you can	36	2	I think * is	51	4	you * have to	44	3	I think * it's	34	2
12	I * to to	36	6	and there * a	49	2	you can * a	44	8	I * say that	34	3
13	erm I * to	35	5	you * you can	49	2	to * to the	43	3	erm I * that	34	3
14	is a * idea	32	3	I think * the	48	2	you don't * to	41	2	I think * a	34	2
15	I think * erm	31	2	I * to to	47	7	I * that the	40	2	okay * I think	33	4
16	I think * it's	31	2	I * that the	45	2	I think * the	40	2	or * like that	32	2
17	there * a lot	30	3	and * is a	44	2	and * is a	35	2	there are * people	31	4
18	in the * erm	30	8	I think * erm	41	3	it's a * idea	35	3	you * have to	31	3
19	it's a * idea	28	2	I * that erm	40	2	of * of the	35	2	I * that the	29	2
20	and I * to	27	4	to to * to	40	6	and there * a	35	2	I * to go	29	5