

What do learners say? Successful spoken English and linguistic competence: lessons from a learner corpus

Dr Christian Jones

Christian.jones2@liverpool.ac.uk

@ELTResearch



In this talk I will:

1. Briefly discuss the reasons for exploring successful spoken English.
2. Briefly outline the corpus data used and its design.
3. Explore what the data can tell us about word frequency and lexical chunks.
4. Outline some possible implications arising from the data.

Why successful spoken English?

1. The native speaker is not always a realistic model for learners.
2. Important to look at what learners can do, as well as the errors they make.
3. It can help us to implement the Common European Framework of Reference for Languages (CEFR) (Council of Europe, 2001).

Linguistic competence

...the ability to use **lexis**, grammar, lexico-grammar and phonology, effectively.

One important element of communicative competence (Hymes, 1972)

Alongside and in combination with:

Discourse competence

Pragmatic competence

Strategic competence

Jones, C., Byrne, S., & Halenko, N. (2017) based on Hymes(1972), Canale & Swain (1980), Canale (1983), Bachman & Palmer (1996).

Corpus data used

Successful learner exams from B1 – C1 levels plus native speakers

Speaking Test Corpus (USTC)

Interactive, paired interviews, lasting approximately 12 minutes

Part A : Introductory questions (Examiner & candidate)

Part B: Paired discussion task (Candidate & candidate)

Part C: Follow up discussion (Examiner & candidate & candidate)

All exams independently rated 0-5 with only those scoring 3.5 or 4 included

<i>Total word count including examiner</i>	91, 173 tokens
<i>Total number of exams used</i>	57
<i>Total number of speakers</i>	121 (60 males (49.6%), 61 females (50.4%) Native speaker = 14 (4 males, 10 females)
<i>Average age</i>	23 years
<i>Average time spent learning English</i>	8 years
<i>Average time in UK</i>	14 months
<i>Nationalities represented</i>	Chinese = 50; British = 14; Saudi = 13; Japanese = 11; Qatar = 6; Republic of Korea = 4; Nigerian = 4; Unanswered = 3;United Arab Emirates = 3; Iraqi = 3; Libyan = 2; Omani = 2; Egyptian = 2; Columbian = 2;Turkish = 1; Italian = 1

Findings 1

Word frequencies

Findings

Exam Level	Freq. Band	Cumul. Token %
B1	K-1	92.81
	K-2	96.29
B2	K-1	92.86
	K-2	97.25
C1	K-1	91.64
	K-2	96.62

Tokens and types from B1- C1

Level	Freq. band	Types			Tokens		
		Freq.	Mean	SD	Freq.	Mean	SD
B1	K-1	878	197.59	28.75	14762	868.12	191.63
	K-2	219	19.24	6.87	553	32.53	14.35
B2	K-1	1011	226.00	31.69	18706	1100.24	286.69
	K-2	267	26.00	6.13	884	52.00	14.93
C1	K-1	1136	252.12	41.00	20303	1194.18	316.98
	K-2	340	33.94	10.40	1104	64.94	29.30

Example

K1 and K2 words at B1

<\$19M> **Er** actually the most popular sport in my country is football **er** I I like football and **er** I've found national team. **Er** actually national team played yesterday last night and **er er losed** the cup silver cup. I'm sad today but **er** the the you know the football it's a game I think it's help the politics to keep the people of the country happy to keep the people in the country you know **er** fans watch the TV <\$=> it's </\$=> and also the people **er** happy when they when they watch **er** the the **foot=** the football match.

Observations

1. The first two thousand words are crucial for successful speech at B1-C1 levels.
2. An increased use of very frequent vocabulary including an increase in types marks out learner progress across levels.

See Leech, Rayson and Wilson (2001) for BNC wordlists

What are the most frequent words used by successful learners?

	B1	Freq	B2	Freq	C1	Freq
1	Er	1391	Er	1312	The	1147
2	I	1126	I	926	I	807
3	The	688	The	827	Er	770
4	And	523	And	591	And	665
5	To	461	To	585	To	640
6	Erm	320	You	518	Erm	460
7	In	296	Yeah	380	Yeah	442
8	Is	288	In	376	Is	417
9	So	288	A	325	In	414
10	A	269	Is	316	You	376

What are the most frequent words used by successful learners?

	B1	Freq	B2	Freq	C1	Freq
11	You	252	Erm	308	A	350
12	Yeah	250	Like	304	Like	348
13	My	238	Think	276	Think	342
14	Think	237	My	252	Of	312
15	Like	222	So	252	So	295
16	Because	188	But	233	They	285
17	Of	161	Have	229	It's	276
18	It's	152	They	226	It	254
19	Can	150	Of	222	We	221
20	We	149	We	212	Because	196

What are the most frequent words used by native speakers?

1 I	374	11 Of	134
2 And	249	12 That	128
3 The	249	13 Think	127
4 Like	214	14 It's	100
5 To	197	15 In	97
6 You	184	16 So	79
7 A	172	17 Just	75
8 Yeah	171	18 If	74
9 It	166	19 On	71
10 Erm	134	20 But	69

Example

The use of yeah with 'we can' to develop points across speaker turns at C1

<\$35F> <\$076> Mhm we can bring them </\$076> And
normally they have hairdryers so +

<\$34F> Yes.

<\$35F> or laundry er machines **so yeah** it's alright.

<\$34F> **Yeah** if there is amenities i= it's fine +

<\$35F> **Yeah.**

<\$34F> + and it's good for me +

<\$35F> **Yeah.**

<\$34F> + but **yeah we can** bring it.

Observations

1. The functions of frequent words change as the levels progress e.g. use of 'yeah' , use of 'think' use of 'like'.
2. Focus at B1 is on learners' own turn and increasingly on interaction from B2 level onwards.
3. Several functions unlikely to have been taught e.g. use of 'like' as discourse marker.
4. Native speakers hedge and soften more e.g. with 'just' and link ideas with less repetition and more synonyms.

4-word lexical chunks

Recurring formulaic sequences accessed as a single choice (based on Wray, 2002)

	B1		B2		C1	
Rank	Chunk	Freq.	Chunk	Freq.	Chunk	Freq.
1	I THINK IT'S	26	I AGREE WITH YOU	21	I THINK IT'S	41
2	I DON'T LIKE	23	I THINK IT'S	20	I DON'T KNOW	12
3	ER I THINK ER	15	A LOT OF TIME	19	WHAT DO YOU THINK	12
4	I DON'T KNOW	13	WHAT DO YOU THINK	16	YEAH I AGREE WITH	12
5	AND ER AND ER	10	SPEND A LOT OF	15	I AGREE WITH YOU	11
6	I WOULD LIKE TO	10	I DON'T LIKE	12	IT'S IT'S	10
7	GO TO THE CINEMA	9	I DON'T KNOW	10	LOCATION OF THE HOTEL	10
8	TO BE IN A	9	THEY DON'T HAVE	10	THE LOCATION OF THE	8
9	ER A LOT OF	8	DO YOU THINK ABOUT	9	YEAH IT'S VERY	8
10	I DON'T THINK	8	I DON'T THINK	9	DO YOU THINK ABOUT	7

Lexical chunks – 4 word chunks produced by native speakers

1	I DON'T KNOW	15
2.	I THINK IT'S	13
3.	I DON'T THINK	11
4.	AND THINGS LIKE THAT	7
5.	I DON'T REALLY	7
6.	A LOT OF THE	6
7.	I THINK THAT'S	6
8.	LOT OF THE TIME	6
9.	AND STUFF LIKE THAT	5
10.	BUT I DON'T	5

Observations

1. Frequency and uses of chunks often changed as levels developed e.g. 'I think it's ..' rather than the chunks themselves.
2. When chunks do change across levels, it tends to indicate greater interaction e.g. 'What do you think?' from B2 onwards.
3. Lots of similarity to NS data e.g. use of 'I think it's..', 'a lot of'
4. Differences seem to be in terms of vague language e.g. 'and things like that'.

Three possible implications for teaching

1. Higher levels could serve as a model of lower level learners e.g. B2 for B1. Frequent language could inform syllabuses.
2. Strong focus on the top 2000 words needed in and out of class. Meaning, uses, collocations, chunks. Chunks which can fulfil a range of functions need to be learnt e.g. 'I think it's...' Rather than many chunks for one function e.g. giving opinions.
3. Focus on the interactive nature of spoken language and on 'turn grammar' (Clancy and McCarthy, 2014) as indicated by uses of frequent language.

Questions?

References

- Bachman ,L.F., & Palmer, A.S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Canale,M. (1983). From communicative competence to communicative language pedagogy. In J.C Richards and R.W. Schmidt (Eds.) *Language and communication* (pp.2-27).New York: Longman.
- Canale, M., & Swain, M. (1980) Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1-47.
- Council of Europe (2001). Common European framework of reference for languages:: language, teaching, assessment. Cambridge: Cambridge University Press.
- Clancy, B., & McCarthy, M. (2014). Co-constructed turn taking. In K.Aijmer, and C.Ruehleemann (Eds.) *Corpus pragmatics : A handbook* (pp.430-453). Cambridge: Cambridge University Press.

References

- Gilquin, G., DeCock, S., & Granger, S. (2010). *LINDSEI: Louvain International Database of Spoken English Interlanguage*. Louvain: Presses Universitaires de Louvain
- Hymes, D. (1972). On communicative competence. In Pride, J.B.; Holmes, J, (Eds) *Sociolinguistics: Selected Readings* (pp.269–293). Harmondsworth: Penguin.
- Jones, C., Byrne, S., & Halenko, N. (2017). *Successful spoken English: Findings from learner corpora*. London: Routledge.**
- Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English: Based on the British National Corpus*. London: Longman. Companion website with lists: www.ucl.ac.uk/lanes/bncfreq
- Lextutor. www.lex tutor.ca
- Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press

ROUTLEDGE APPLIED CORPUS LINGUISTICS

Successful Spoken English

Findings from Learner Corpora

Christian Jones, Shelley Byrne and
Nicola Halenko

