Computing the Vocabulary Demands of PODCASTS

Ulugbek Nurmukhamedov
u-nurmukhamedov@neiu.edu

What is a **podcast**?

A multimedia file, such as a radio program or video, that can be downloaded or streamed from the Internet onto a computer or mobile device. The method of publishing files that can be used in this way is called **podcasting**. (Macmillan Dictionary)

Over **525,000 active podcast** shows and over *18.5 million episodes* (FastCompany, April 25, 2018)

English podcasts …

L2 learning and podcasts

PODCASTS ...

- are convenient and portable (Thorn & Payne, 2005)

- enable learners to pause, skip, slow down (Giordano, 2016)

- have positive effects on students' study habits (Abdous et al, 2009)

- help to reinforce previously learned vocabulary (O'Bryan & Hegelheimer, 2007)

- improve learners' listening skills (Ashraf et al, 2011; Qasim & Fadda, 2013)

- consist of [mostly] **authentic** [listening] **texts**

Podcast Categories

- "Podcast creators have different interests and motivations" (Hegelheimer & O'Bryan, 2007, p. 332).

- Podcasts into different categories

(1) **Discrete Category**	(3) **General Audience** *
Grammar Girl™ Quick and Dirty Tips FOR BETTER WRITING	npr RADIOLAB
(2) **ESL-focused**	(4) **ESL Super**
RACHEL'S ENGLISH CULips ESL Podcast LEARN ENGLISH & CULTURE	VOA Voice of America BBC LEARNING ENGLISH

(Nurmukhamedov & Sadler, 2011)

Purpose of the Study

examine the vocabulary demands of English podcasts ...

How many words do English language learners need to know to understand ***General Audience*** podcasts?

The Study

1. Lexical coverage
2. Methodology
3. Analysis
4. Preliminary Findings
5. Conclusion

Lexical coverage: why?

5 FINGER RULE

What To Do: Pick a book. Open the book to any page. Put one finger up for each word you don't know.

0-1 Too Easy
1-2 Perfect Choice
4 Give It A Try
5+ Too Hard

Find a "just right" book with the

FIVE FINGER RULE!

Choose a book and read the first page or two.

Put one finger up for every word you don't know.

If five of your fingers go up while reading, choose another book.

If only two or three fingers go up, you've found a "just right" book.

Happy Reading!

Background: lexical coverage I

- # of words necessary to reach **'adequate'** discourse comprehension.

- **95%** to **98%** coverage is needed to establish '*adequate comprehension*'

- **READING comprehension** (Laufer & Ravenhorst-Kalovski 2010)
  - 95% - minimal threshold
  - 98% - optimal threshold

- **LISTENING comprehension** (van Zeeland & Schmitt, 2013)
  - 95% - sufficient for good comprehension
  - 98% - provide a high-level comprehension

A

**T
H
E
M
E**

reasonable comprehension level may be achieved at … above **90%** …

Background: lexical coverage II

- **95%** OR **98%** coverage "provide a useful indication of whether or not a text may be understood" (Webb & Nation, 2013)
  - 95% = **1** word out of **20** are unknown
  - 98% = **1** word out of **50** are unknown

- to examine the number of words necessary to reach the coverage figures associated with adequate comprehension
  - word frequency lists (*e.g., COCA, BNC*)
  - representative corpus of a domain

B
T
H
E
M
E

Background: lexical coverage II

AUTHORS	LISTENING	95%*	98%*
Webb & Rodgers (2009)	English TV Programs	3,000	7,000
Dang & Webb (2012)	British Academic Lectures	4,000	8,000
Nurmukhamedov (2017)	TED Talks Presentations	4,000	8,000

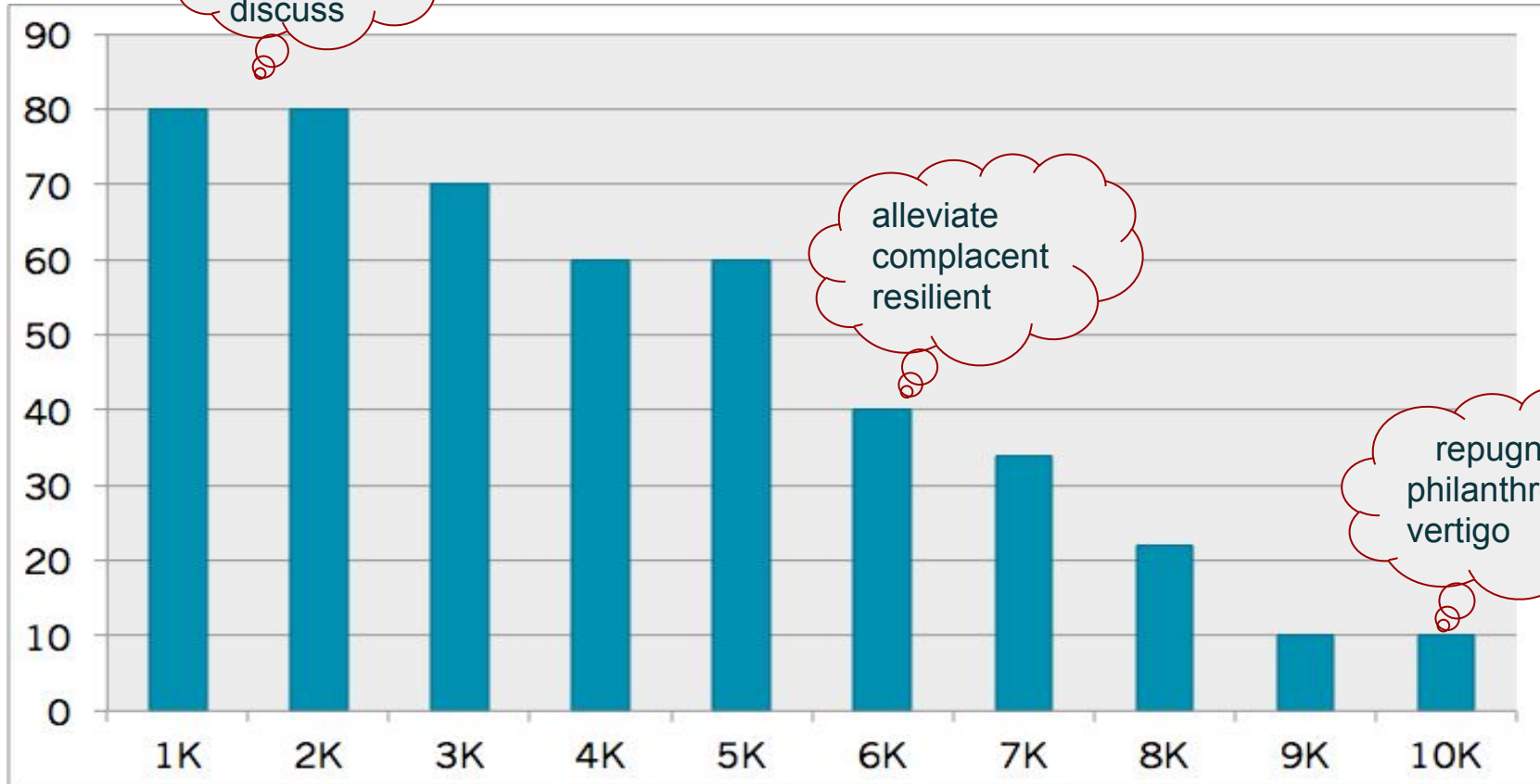(*plus proper nouns and marginal words)

Word Families: "a base word with its inflections and derivatives"
(Schmitt & McCarthy, 1997)

argue, arguing, argues, argued (*inflections*)

arguable, arguably, unarguably, argument, arguments (*derivatives*)

Methodology: **Corpus Compilation**

PODCASTS	# of EPISODES	# of WORDS [tokens]	# of HOURS [hour:min:sec]
FRESH AIR	20	131199	11:53:58
HIDDEN BRAIN	20	98231	09:27:43
HOW I BUILT THIS	20	161162	15:13:28
INVISABILIA	20	136061	13:51:34
TED RADIO HOUR	20	162634	16:28:51
TOTAL	**100**	**689,287**	**66:55:34**

Analysis I: **RANGE software**

The RANGE lists all the words that occur in a **corpus** according to where they are found in **word lists** (Nation & Heatley, 2002)
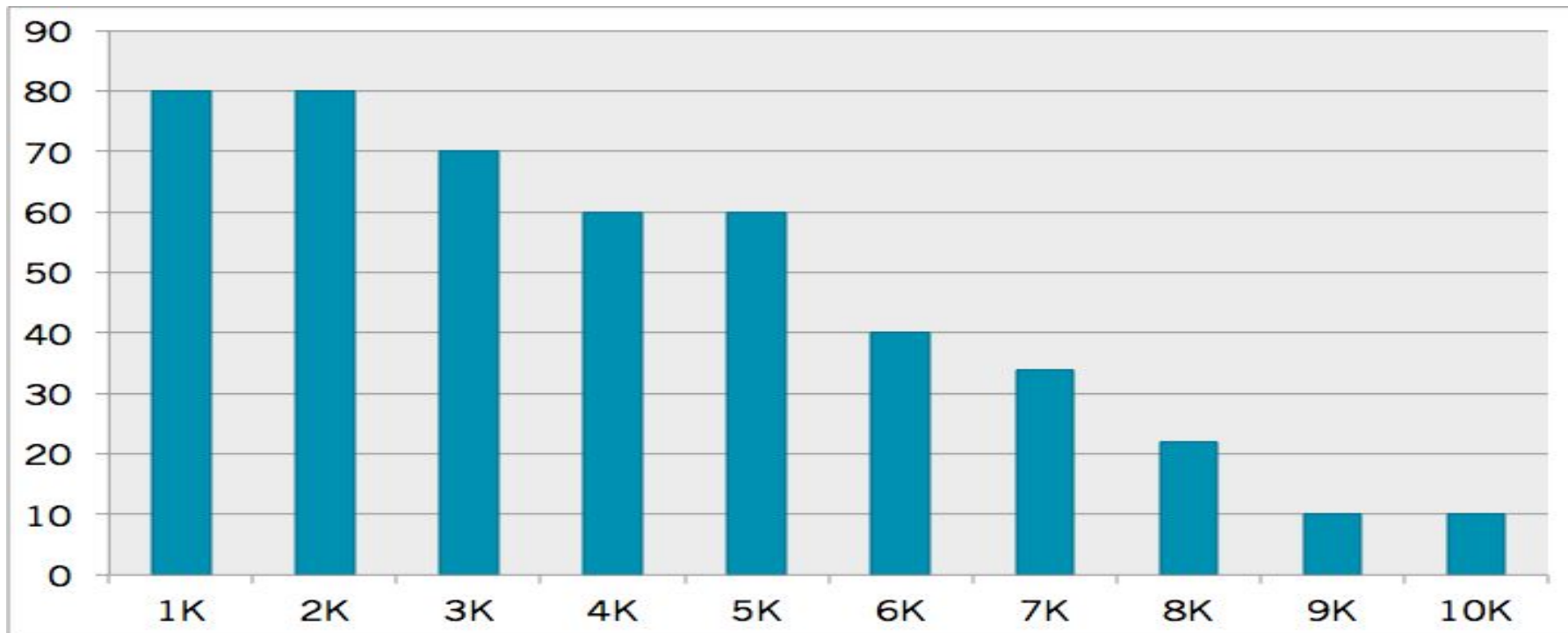
Analysis II: **BNC/COCA** (Nation, 2012)

BNC/COCA Corpus:
- 29 word family [*word frequency*] lists
- The four additional lists are the lists of
  - proper nouns (*e.g., USA, Uzbekistan*)
  - marginal words (e.g., *swear words, exclamations*)
  - transparent compounds (*e.g., countertop, poolroom*)
  - abbreviations (*e.g., UN, UNESCO*)

Findings I: Lexical Coverage of Podcasts

95% coverage ⇒ 3,000 word families, *plus proper nouns, marginal words, abbreviations*

98% coverage ⇒ 19,000 word families, *plus proper nouns, marginal words, abbreviations*

Findings II: Lexical Coverage of Each Podcast

PODCASTS	**95%** coverage* ⇒ *good* comprehension	**98%** coverage* ⇒ *high-level* comprehension
FRESH AIR	3,000 wf	7,000 wf
HIDDEN BRAIN	3,000 - 4,000 wf	20,000 wf
HOW I BUILT THIS	3,000 wf	20,000 wf
INVISABILIA	3,000 wf	17,000 - 18,000 wf
TED RADIO HOUR	3,000 wf	23,000 wf

*(plus, marginal words, proper nouns, abbreviations)

Implications: **RESEARCH**

AUTHORS	LISTENING	95%*	98%*
Webb & Rodgers (2009)	English TV Programs	3,000	7,000
Dang & Webb (2012)	British Academic Lectures	4,000	8,000
Nurmukhamedov (2017)	TED Talks Presentations	4,000	8,000
Current study	Podcasts in English	3,000	19,000

(*plus proper nouns and marginal words)

Vocabulary knowledge of educated native speakers:
  - *a range of 16,000-20,000 word families* (Schmitt, 2010, p. 6)**.**

Implications: **TEACHING**

- ***Discrete Category Podcasts:*** *although authentic - use with caution*

- ***K1, K2, K3 words****: emphasize the importance of the first 1,000 frequent words;*

- ***Teacher-created podcasts***: examine the vocabulary load of your podcast episodes;

- ***Teacher-selected podcasts***: find out the students' vocabulary size
  - Administer the ***Vocab Size Test***

Limitations

- ***Not In the List***: some of these words are proper names (*Facebook, Wikileaks, Kushner, 23andme.com*).

- ***Future research***: vocabulary coverage does not necessary mean comprehension.
  - Connected speech
  - Background knowledge
  - Familiarity with the culture

References

Dang, T. & S. Webb (2014). The lexical profile of academic spoken English. *English for Specific Purposes* 33, 66–76.

Heatley, A., I.S.P. Nation & A. Coxhead (2002). Range: A program for the analysis of vocabulary in texts. Retrieved from
http://www.victoria.ac.nz/lals/about/staff/paul-nation

Hegelheimer, V., & O'Bryan, A. (2009). Mobile technologies and language education: A look at language learning podcasts. In Thomas, M. (Ed.), *Handbook of Research on Web 2.0 and Second Language Learning* (pp. 331-349). Hersley, PA: Information Science Reference.

Nation, I.S.P. (2012). The BNC/COCA word family lists. Retrieved from
http://www.victoria.ac.nz/lals/about/staff/paul-nation

Nurmukhamedov, U. (2017). Lexical coverage of TED Talks: Implications for vocabulary instruction. *TESOL Journal, 8,* 268-290.

Nurmukhamedov, U., & Sadler, R. (2011). Podcasts in four categories: Applications to language learning. In M. Abdous & B. R. Facer (Eds.), *Academic podcasting and mobile assisted language learning: Applications and outcomes* (pp. 176-195). Hershey, PA: IGI Global.

Nurmukhamedov, U., & Webb, S. (in press). Research timeline: Lexical coverage and profiling. *Language Teaching*.

Stæhr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition* 31.4, 577–607.

Thorne, S., & Payne, S. (2005). Evolutionary trajectories, Internet-mediated expression, and language education. *CALICO Journal*, *22*, 371-397.

van Zeeland, H. & N. Schmitt (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics* 34.4, 457–479.

Webb, S. & M.P.H. Rodgers (2009). Vocabulary demands of television programs. *Language Learning* 59.2, 235–366.

Vocabulary & L2 Listening

Stæhr (2009)

- Relationship between vocabulary size and depth on listening comprehension.

- Learners' scores on the Vocabulary Levels Test (Schmitt, Schmitt & Clapham 2001)

- **Conclusion**: 98% coverage provided <u>adequate</u> L2 listening comprehension.

95% <u>good</u> listening comprehension

van Zeeland & Schmitt (2013)

- Relationship between lexical coverage and listening comprehension.

- L1 and L2 learners' listening comprehension of informal narratives.

- **Conclusion**: 95% may be sufficient for good comprehension of informal narratives.

98% <u>high-level listening</u> comprehension

Informative infographics: https://www.podcastinsights.com/podcast-statistics/

PODCASTING STATS & FACTS

2018 US PODCAST STATISTICS

44% OF US POPULATION HAS LISTENED TO A PODCAST

49% OF PODCAST LISTENING IS DONE AT HOME. 22% OF LISTENING IS DONE IN THE CAR

In car/truck 22%

At work 11%

Riding public transportation 4%

Working out 4%

Walking around 3%

Other 7%

At home 49%