

**The New General Service List (NGSL) Project:  
A Modular, Corpus-based Approach for the  
Systematic Development of 2nd Language Vocabulary**

Dr. Charles Browne  
Professor, Department of English  
Meiji Gakuin University, Tokyo, Japan  
[browne@meiji.ac.jp](mailto:browne@meiji.ac.jp)

A few of My Free Websites & Tools:  
[www.newgeneral-service-list.org](http://www.newgeneral-service-list.org)  
[www.newgeneral-service-list.org/2012/02/](http://www.newgeneral-service-list.org/2012/02/)  
[www.newgeneral-service-list.org/2012/02/business-service-list/](http://www.newgeneral-service-list.org/2012/02/business-service-list/)  
[www.et-central.com](http://www.et-central.com)  
[www.et-central.com/etgale/](http://www.et-central.com/etgale/)

1

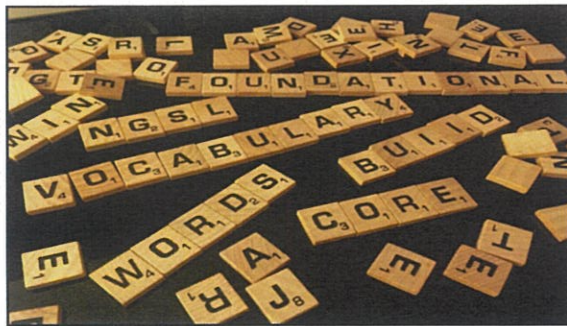
**Why Focus on Vocabulary?**

2

The seminal work in 2<sup>nd</sup> language vocabulary instruction...

(1990, Heinle & Heinle) (2000, 2010, 2022, Cambridge University Press)

3



4

**Quick question:**

How many words do you think there are in the English language? Think about the largest dictionaries you know of. Think about all the special words like scientific and medical words.

Discuss it with a few people next to you and make your guesses.

5

Number of words in the English Language:

**600,000**

Oxford English Dictionary, 2015.

6

**Another quick question:**

There are about 600,000 words in the English language. How many of these do you think a college-educated native speaker of English knows.

Discuss it with a few people next to you and make your guesses.

7

**30,000**

Hayes, Anderson, 1984. Goodales, Nelson, Farnell, 1990

8

**Discussion question:**

From birth until they graduate from college, native speakers learn an average of 1000 words a year. Most plateau at between 22,000-30,000 words. Why do you think so? Discuss several possible reasons and then present to the class

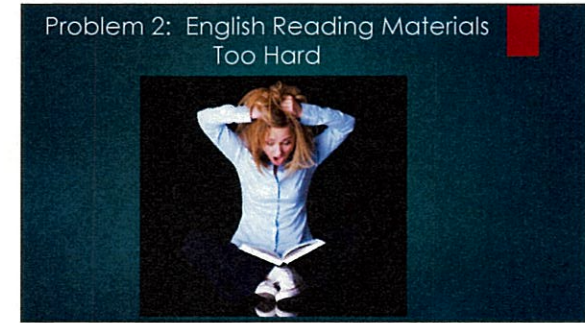
9



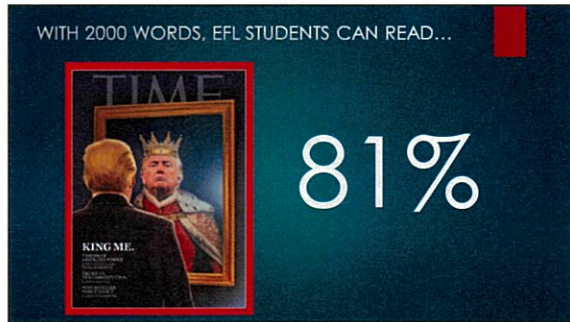
10



11



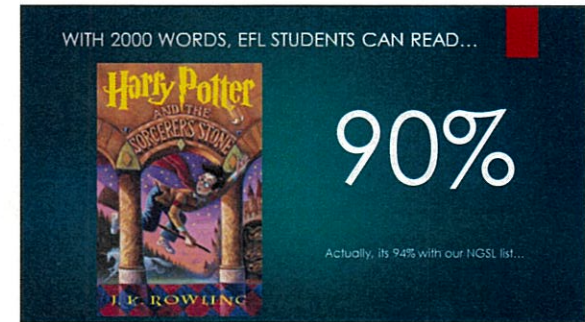
12



13



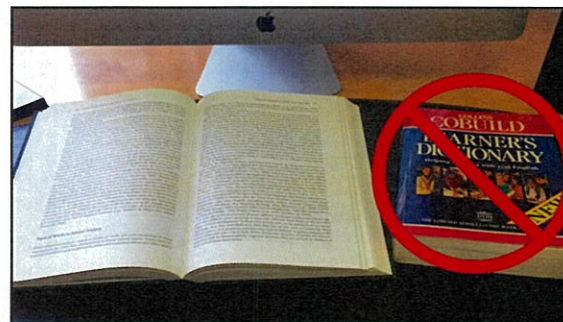
14



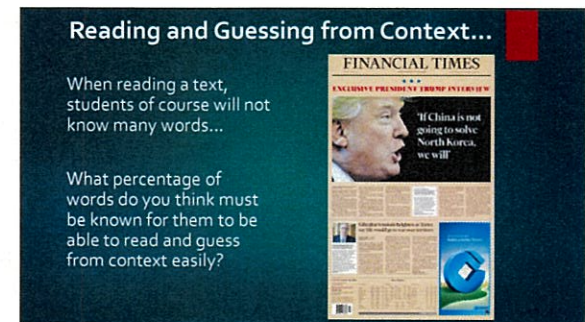
15



16



17



18

### 75% Coverage

1000 high frequency words

**Forestry** [15 missing words]

If \* planting rates are \* with planting \* \* in each \* and the forests \* at the earliest opportunity, the \* wood supplies could further increase to about 36 million \* meters \* in the \* 2001-2015. The additional \* wood supply should greatly \* \* \* , even if much is used for \* production.

19

### 85% Coverage

2000 high frequency words

**Forestry** [9 missing words]

If \* planting rates are **maintained** with planting **targets satisfied** in each \* and the forests **milled** at the earliest opportunity, the \* wood supplies could further increase to about 36 million \* meters \* in the **period** 2001-2015. The additional \* wood supply should greatly **exceed** \* \* \* , even if much is used for \* production.

20

### 95% Coverage


5000 high frequency words

**Forestry** [1 missing word]

If current planting rates are **maintained** with planting **targets satisfied** in each **annually** and the forests **milled** at the earliest opportunity, the **available** wood supplies could further increase to about 36 million \* meters **annually** in the **period** 2001-2015. The additional **available** wood supply should greatly **exceed** **domestic requirements** , even if much is used for **export** production.

21

### Vocabulary Thresholds Research...



- Below 80%, reading comprehension is almost impossible (Hu & Nation, 2001)
- In studies of student performance in ESL settings, knowledge of about 90% was the point at which there were more readers than non-readers (Laufer, 1992)
- 95% coverage is the point at which learners can read without the help of dictionaries (Laufer, 1989)
- 98% coverage is the point at which learners can **read for pleasure** (Hu & Nation, 2001)

22

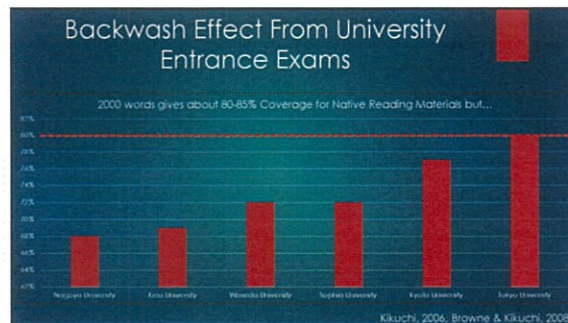
### Are Japanese students **reading** the right vocabulary? (Browne, 1996, 1998)

Text	Coverage from 2000 High Frequency Words
Spectrum	71%
Milestone	78%
Unicorn	79%
Unsimplified Native Texts	<b>80-85%</b>

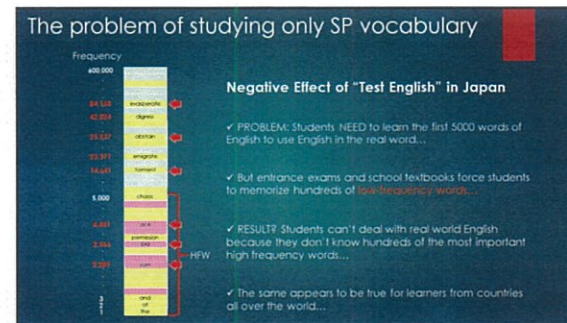
23



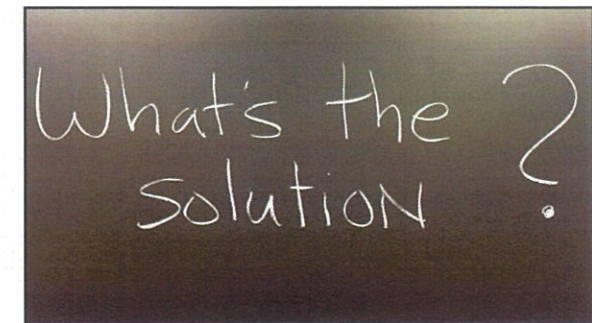
24



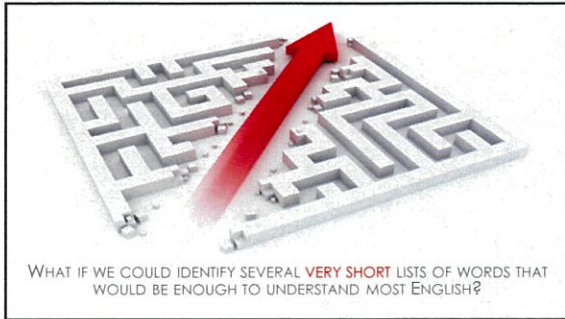
25



26



27



28

### The New General Service List (NGSL)

- ▶ 600,000 words in English
- ▶ 30,000 known by adult native speakers
- ▶ 90% is minimum to read/listen and be able to guess from context
- ▶ **2800** words in NGSL covers **92%** of words students are likely to ever meet in average newspaper, book, magazine, TV show, Movie or daily speech...

Browne, Culligan & Phillips, 2013

29

### A NGSL: Created with Full Access to the Cambridge English Corpus...

30

### A Balanced NGSL Corpus ....

273 million words from the CEC 2 billion word corpus

CEC corpora included in final analysis for NGSL:

Corpus	Tokens
Learner	38,219,480
Fiction	37,792,168
Journals	37,478,577
Magazines	37,329,846
Non-Fiction	35,443,408
Radio	28,882,717
Spoken	27,934,806
Documents	19,017,236
TV	11,515,226
Total	273,613,534*

\*273 million word subsection used is 100x larger than original GSL corpus...

31

### How is this possible? Zipf's law...

George K. Zipf (1902 – 1950) established a mathematical model for describing frequency data:

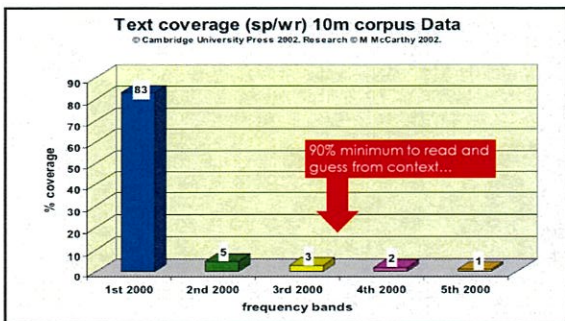
*"Frequency decreases with rank. More precisely, frequency is inversely proportional to rank"*

32

### How Zipf's Law works for English

Words	% known	# known	Researcher
1	7%	97,100	West (53, Nation90)
10	25%	3.4	West (53, Nation90)
100	50%	1.2	West (53, Nation90)
1000	75%	1.4	West (53, Engles(68)
3600 (2260)	80%	1.7	West (53, Nation90)
<b>2800</b>	<b>92%</b>	<b>0.713</b>	<b>Browne et al (2013)</b>
5000	95%	1.20	Hirsch & Nation (92)
8000	98%	1.50	Laufer (92), Coady(93)
600,000	100%	100/100	Oxford English Dictionary

33



34

### A Modular Approach to EFL Success:

STEP 1: Learn high frequency **General English** words

STEP 2: Learn high frequency **Specific Purpose** words

99%  
87-94%

- 30,000 Native Speaker Vocabulary Size
- 1-2000 Specific-Purpose Vocabulary Core Daily Vocabulary
- 2000 Core General English Words - Written/Spoken (NGSL)
- 700 Core Spoken (NGSL-S)
- 875 Kids English (NDL)
- 300 Academic Words (NAWL)
- 700 Top 500 Words (HSU)
- 700 Business Words (BSL)
- 600 Finest Words (FSL)

Browne, Culligan & Phillips, 2013. Browne & Culligan, 2015. Browne & Culligan 2020

35

### Why is the NGSL always step 1?

- ▶ These words occur EVERYWHERE in student's daily life
  - ▶ Books, newspapers, magazines, internet
  - ▶ TV, radio, movies
  - ▶ Conversational English
- ▶ Knowing these words gets you **92%** of the way there for written texts, even more for many spoken genres (**95%** of FRIENDS for example)
- ▶ The 2800 NGSL words represent less than **0.5%** of the English language and only **9%** of a college-educated native speaker


36

### SOME NGSL RESEARCH

#### Coverage of NGSL for the Japanese National Center Exam

MacDonald, E. (2019). An Analysis of Vocabulary Level in Reading Passages of the National Center Test. *Shien*, 23(2), 19-27.

**Abstract:**  
A strong knowledge of high frequency vocabulary is essential for second language (L2) learners of English as it provides a large proportion of text coverage for general English texts. Additionally, a minimum understanding of 78% to 98% of the words of a text is essential to allow for reasonable comprehension and parsing of words in context. The current study examines the text coverage provided by the New General Service List (NGSL), a list of high frequency vocabulary for L2 English learners, for reading comprehension passages of Japan's National Center Test, a standardized high school national university entrance examination. Results showed that 70-90% of the words of the passages were covered by the NGSL. The findings present a compelling argument that supporting students to acquire high frequency vocabulary should be strongly focused on at secondary schools in Japan. MacDonald, E. (2019). An analysis of vocabulary level in reading passages of the National Center Test.



37

### SOME NGSL RESEARCH

#### Coverage of latest MEXT high school textbooks

Hakayama, S. (2022). A close examination of vocabulary in Japanese EFL textbooks. In F. Chen, S. Dornan, and P. Ferguson (Eds), *Reflections and New Perspectives*. JALT, pp. 1-8.

**Abstract:**  
A corpus analysis of the top 3 selling MEXT-approved high school English textbooks in the Tokyo area revealed that **vocabulary of the 2000 word NGSL is provided between 70-90% of the texts in these books.**

Textbook	Number of entries	Number of full text words	Coverage (%)
VISTA I	1,833	110	92.8
NEW MASTER I	3,864	91	97.9
COMET I	1,237	80	96.5
VISTA II	1,795	79	95.8
ESSENTIAL II	4,349	69	98.9
COMET II	3,475	64	98.1
LANDMARK III II	4,823	59	98.8
ESSENTIAL III	7,487	48	99.3
All books	2,946	217	92.6




38

### SOME NGSL RESEARCH

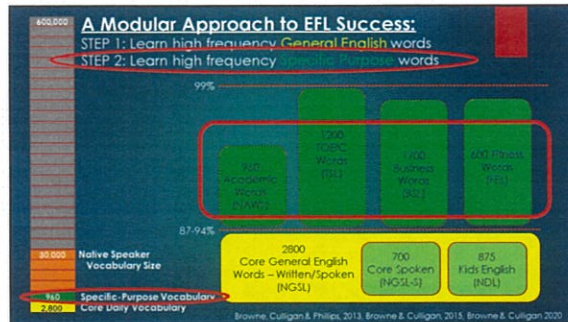
#### Coverage of NGSL for SHS Entrance Exam and National Center Test

Iwamoto, M., Kramer, B. (2020) *The Language Teacher* 44:4

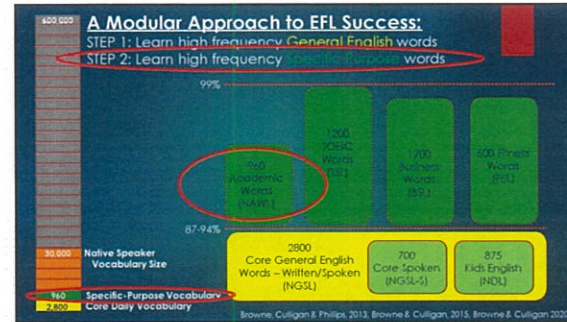
**Abstract:**  
While the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) currently expects students to learn 1,200 English words in junior high school and 1,800 English words in high school (MEXT, 2017), there is little to no guidance on the specific words required. Looking at the reading sections on Japanese public high school entrance examinations and the university National Center Test, this study reports the lexical coverage provided by a well-known and publicly available word list, the New General Service List (NGSL) (Browne, Culligan, & Phillips, 2013). **The NGSL provided a high 98.11% coverage of the vocabulary on junior high school entrance examinations using only 1,000 words but was only able to cover 92.25% of the vocabulary on the National Center Test with all 2,800 words.** The results will be discussed in detail, along with the utility of the NGSL in Japanese junior and senior high school classrooms.



39




40



41

### NAWL: NEW ACADEMIC WORD LIST

(Browne, Culligan & PHILLIPS 2013)



42

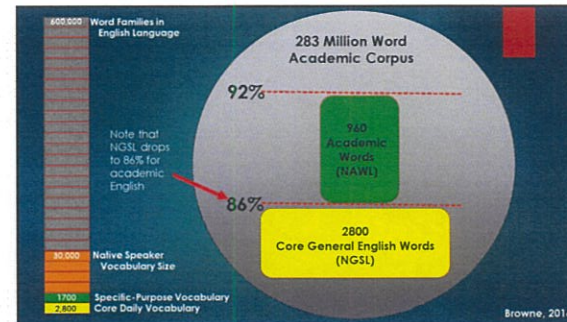
### The New Academic Word List (NAWL)

Corpora	Size	Percent
CEC Academic	248 million	86.30%
Oral	3 million	1.10%
Textbooks	36 million	12.60%
Total	288 million	100%

The **944** words of the NAWL gives you **about 92%** coverage in Academic Texts when combined with NGSL...

CEC: academic journals, non-fiction, student essays, & academic discourse  
Oral: MICASE (Michigan Corpus of Academic Spoken) BASE (British Academic Spoken English)  
Textbooks: Corpus of 100 top-selling academic textbooks

43




44

### SOME NGSL RESEARCH

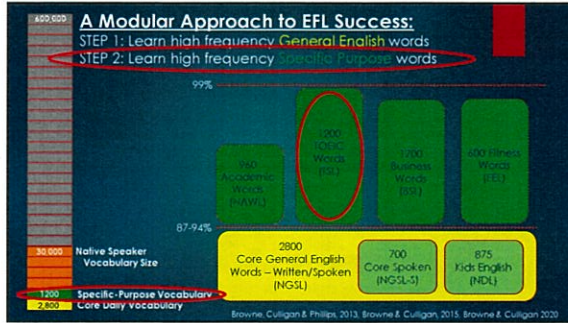
#### Coverage of NGSL & NAWL for Research Publications in Education

Hendry, C., & Sheepy, E. (2018). How much vocabulary is needed for comprehension of research publications in education? In P. Tozlat, J. Jalkanen, L. Bradley & S. Thouny (Eds), *Future-proof CALL: language learning as a exploration and encounters - short papers from EUROCALL 2018* (pp. 94-99).

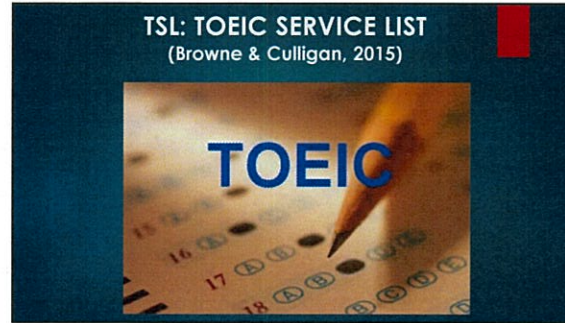
**Abstract:**  
The American Education Research Association (AERA) is one of the largest education conferences in the world. Using the AERA Open Access repository, we created a 3,000,000 word corpus of over 18,000 abstracts. We explored the coverages of the New General Service List (NGSL), the New Academic Word List (NAWL), and the Social Science Word List (SSWL). We found that the NGSL covered 86% of the abstracts, the NAWL covered 92% of the abstracts, and the SSWL covered 92% of the abstracts. The SSWL showed little additional coverage. Our discussion highlights the research and pedagogical implications of our findings and the AERA abstract corpus.



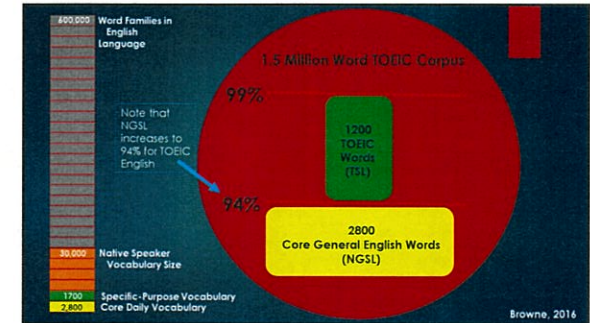
45



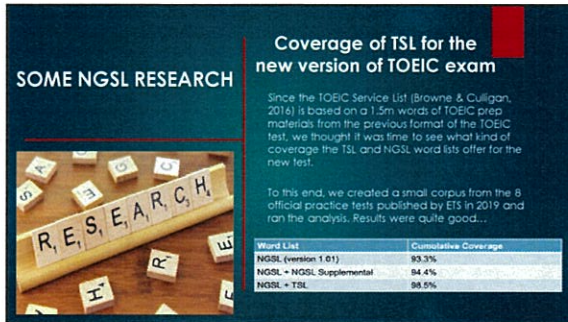
46



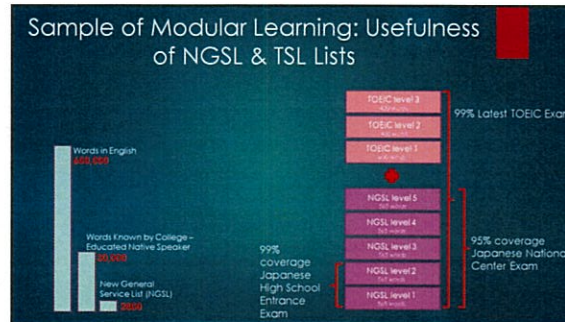
47



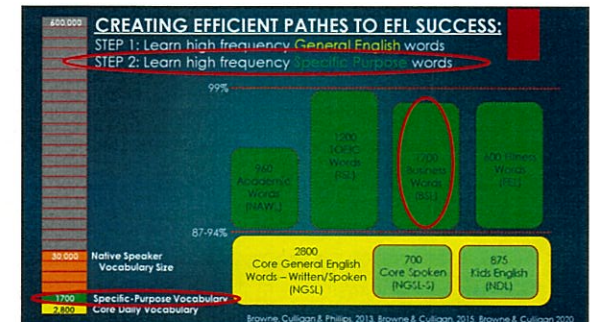
48



49



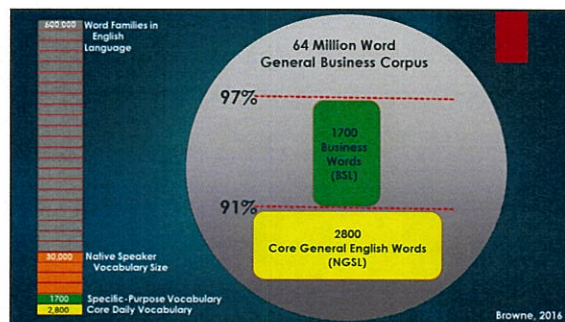
50



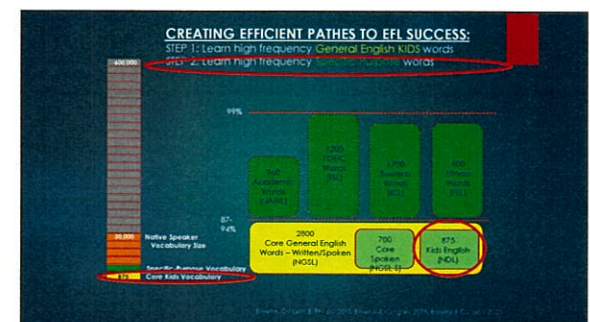
51



52



53



54

### NDL: NEW DOLCH LIST (Browne & Culligan, 2020)

55

### Like the Old General Service List (GSL), the Dolch List is quite dated and in need of an update

56

### The original Dolch and New Dolch Lists have very different focuses:

- ▶ The Original Dolch List was designed to help **NATIVE YOUNG LEARNERS** to build **sight recognition** of words they already knew
- ▶ Our NDL is designed to help **NON-NATIVE YOUNG LEARNERS** build **basic understanding** of the words they are most likely to encounter in **EFL CONTEXTS**...

57

### 2.5 Million-word Children's corpus (with a focus on the needs of EFL Learners)

CATEGORY	TOTAL
Children's Graded Readers	991,104
Elementary EFL Textbooks (MEXT)	4,623
Middle School Textbooks	322,169
Picture Books (for L1)	13,347
EFL Textbooks	575,976
Top Kids YouTube Videos	548,172
New Dolch Corpus Total	2,455,391

58

### Identifying valid video contents to add to our Children's Corpus...

59

### Knowing just 128 of our NDL function words (sight words) provide nearly 50% coverage

Word	Frequency	Word	Frequency	Word	Frequency
The	10000	Of	1000	With	1000
Be	8000	Do	8000	Can	8000
And	6000	Have	6000	At	6000
To	5000	For	5000	but	5000

A few sample "sight" words from the top 25 most frequent words in the NDL

60

### CREATING EFFICIENT PATHS TO EFL SUCCESS:

STEP 1: Learn high frequency **General English** KIDS words  
STEP 2: Learn high frequency **General English** words

Category	Count	Coverage
Native Speakers' Vocabulary Size	~10,000	95%
General English Vocabulary	~10,000	95%
2000 Core General English Words - Written/Spoken (NGSL)	2000	95%
700 Core Spoken (NGSL)	700	95%
825 Kids English (NDL)	825	95%

61

### Quick Discussion:

Work with the person next to you and try to come up with some words that you think might be more important or more frequent for young EFL learners than for adult EFL learner.

Share a few of these words with the group.

62

### Many "kids" words not in the NGSL but frequent in kids EFL materials...

- ▶ Ant
- ▶ Bark
- ▶ Bee
- ▶ Boo
- ▶ Butterfly
- ▶ Cheer
- ▶ Clap
- ▶ Classroom
- ▶ Cookie
- ▶ Crayon
- ▶ Dinosaur
- ▶ Doll
- ▶ Duck
- ▶ Elephant
- ▶ Frog
- ▶ Giggle
- ▶ Hop
- ▶ Kite
- ▶ Ooh
- ▶ Pencil
- ▶ Quack
- ▶ Squeak
- ▶ Swish
- ▶ Teddy
- ▶ Vroom
- ▶ Whale
- ▶ Woah
- ▶ Witch
- ▶ Yap
- ▶ Yum

63

On a large corpus of Children's English:  
 315 words of old Dolch list offers 70% coverage  
 315 words of New Dolch list offers 78% coverage

871	241	company	81724	Lexeme	-2.51	0.0001	0.89	NGSL1017	1	NGSL
872	240	figure	129625	Lexeme	-3.176	0.0001	0.89	NGSL1017	1	NGSL
873	240	bill	41861	Lexeme	-4.191	0.0001	0.89	NGSL0945	1	NGSL
874	240	packet	14564	Lexeme	-2.323	0.0001	0.89	NGSL1383	1	NGSL
875	240	haul	598	Lexeme	-1.207	0.0001	0.89	NGSL1289	1	NGSL
876	239	betwe	6487	Function	-1.134	0.0001	0.89	NGSL1761	1	NGSL
877	239	gamma	1092	Lexeme	-2.751	0.0001	0.89	NGSL1188	1	NGSL
878	239	whale	1568	Lexeme	-1.911	0.0001	0.89			
879	239	handsh	732	Lexeme	-3.095	0.0001	0.89			
880	239	homewo	89064	Lexeme	-4.244	0.0001	0.89			
881	238	capful	19154	Lexeme	-4.497	0.0001	0.89	NGSL1615	1	NGSL
882	238	insh	13740	Lexeme	0.0001	0.90				
883	238	gripho	90	Lexeme	0.0001	0.90				
884	237	inshad	47091	Lexeme	-2.118	0.0001	0.90	NGSL0958	1	NGSL
885	237	journe	15528	Lexeme	-2.403	0.0001	0.90	NGSL1819	1	NGSL
886	237	disappo	18416	Lexeme	-4.029	0.0001	0.90	NGSL1649	1	NGSL
887	237	poet	81521	Lexeme	0.0001	0.90				
888	237	insh	4891	Lexeme	-1.911	0.0001	0.90			
889	236	semibod	17270	Lexeme	-4.405	0.0001	0.90	NGSL0981	1	NGSL

That's 6% better... which is good, but clearly 78% is not enough for second language learner children to be able to interact with EFL materials successfully...

64

At 875 words, the NDL reaches the critical 90% minimal reading threshold...

884	236	semibod	17270	Lexeme	-4.405	0.0001	0.90	NGSL0981	1	NGSL
-----	-----	---------	-------	--------	--------	--------	------	----------	---	------

65

The NGSL Project 2013-2022: Making an impact...

- 7 open-source EFL Wordlists, each offering highest coverage in the world
- 10+ online learning tools and resources
- 1.1 million NGSL Website page views
- 25+ NGSL Plenary/Keynote Presentations
- 40+ academic articles focused on NGSL
- 400+ citations of NGSL wordlists
- Countless textbooks, online learning platforms and apps based on NGSL

66

Constant improvement & innovation of the wordlists over time: Quantitative & Qualitative Data, discussion, debate...

Original GSL 1936-1953 (Michael West, Lawrence Faucett and Harold Palmer)

New GSL 2013-2022 and beyond... (Charles Browne, Brent Culligan, Joe Phillips)

67

www.newgeneralservicelist.org

68

www.newgeneralservicelist.org

FREE NGSL RESOURCES INCLUDE:

- Wordlists (NGSL, NAWL & others)
- Definitions (English & Japanese)
- Published research articles on NGSL
- Online Learning applications
- Online Assessment & Placement Tools
- Online Research Tools
- Online Text Analysis & Grading Tools
- TED Talk on NGSL and NAWL

69

NGSL RESEARCH  
 NGSL TED Talk

70

List downloadable in many forms  
 www.newgeneralservicelist.org

Headword list...

71

List downloadable in many forms...  
 www.newgeneralservicelist.org


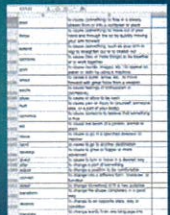
Lemmatized list...

72



List downloadable in many forms...  
www.newgeneralservicelist.org

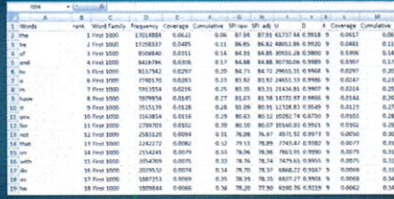
List with definitions in easy English...

73

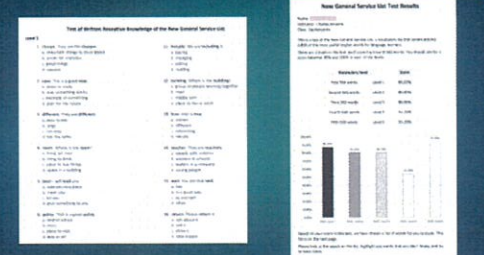
List downloadable in many forms...  
www.newgeneralservicelist.org

List with raw data...




74

NGSLT (New General Service List Test)  
there is also a NAWLT for our New Academic Word List



75

Gamified Learning of the NGSL lists:  
Wordle, Quizlet, Flashcards and more...



76

What is Gamification?

**Definition:**

Gamification is the use of game design elements in non-game contexts"

Defending, Dixon, Khaled and Nacke (2011)

77

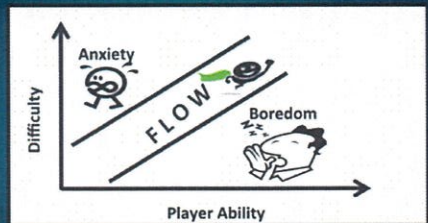
**Discussion question:**

Why do you think gamification of language learning might be helpful to learners? Think about games you have played in the past? Why did you enjoy them? What is different about playing a game than learning English?

Discuss this with the people nearby and then share your ideas.

78

Importance of FLOW



79

Gamified Learning Tools for NGSL

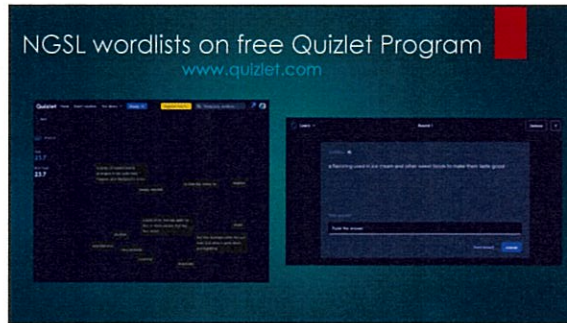
- ▶ NGSL Quizlet Flashcards & Games
- ▶ NGSL Word-Learner Flashcard & Games
- ▶ NGSL Kahoot Game
- ▶ NGSL Wordle Game
- ▶ NGSL crossword puzzles
- ▶ NGSL Word Search
- ▶ NGSL Youghlish Glossary

80

NGSL wordlists on free Quizlet Program...  
www.quizlet.com



81



82



83



84



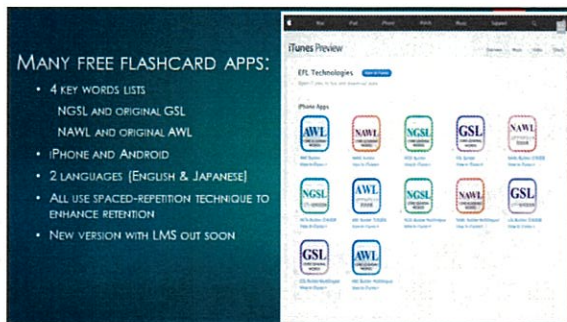
85



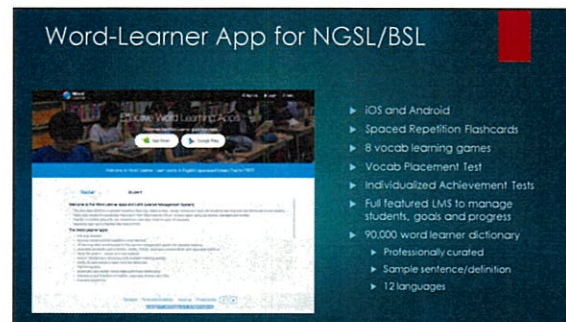
86



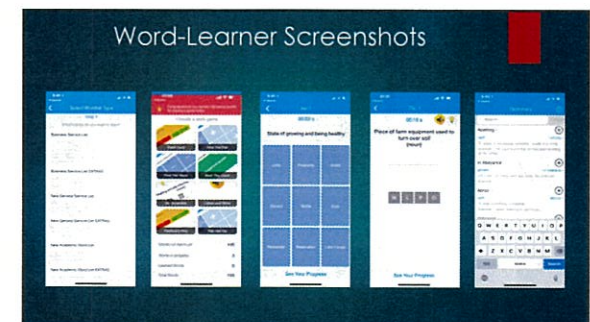
87



88



89



90

### NGSL WORDLE GAME

91

### Use of Wordle for NGSL review & Practice

493 4-letter NGSL words  
466 5-letter NGSL words  
470 6-letter NGSL words  
1429 NGSL words inside game

92

### NGSL Video concordance

93

### NGSL crosswords

94

### NGSL Word searches

95

### NGSL, BSL, NAWL and TSL Tools for Text Analysis and Text Creation

automated data mining survey responses content transcripts qualitative root cause classification insights ad-hoc analysis product reviews sentiment voice of the customer dashboards consumer trends ad-hoc analysis early warning

96

### Online Graded Text Editor (OGTE)

a free text analysis and editing tool developed by Rob Waring and Charles Browne  
www.er-central.com/ogte/

- Analyzes text difficulty by vocabulary level
- NGSL, NAWL, BSL, TSL
- CAMBRIDGE, KET, PET, FCE
- CEFR A1-B2
- Oxford 3000
- Original GSL and AWL
- Huge Proper Noun Database
- Shows words out of level/list and frequency

97

### Online Graded Text Editor (OGTE)

a free text analysis and editing tool developed by Rob Waring and Charles Browne  
www.er-central.com/ogte/

98

### Online Graded Text Editor (OGTE)

NGSL Coverage for Harry Potter (book 1)

99

NGSL + BSL coverage for a typical current events article (From CNN)...

94.9%

100

NGSL + BSL coverage for a typical business article (from Entrepreneur)...

97.43%

101

Coverage by NGSL is about 92% (BSL adds 5.4% coverage)

92.05%

- Shows NGSL's usefulness for business texts
- Shows very nice additional coverage by BSL

102

NGSL, BSL, NAWL AND TSL on VocabProfile  
<http://www.lex tutor.ca/vp/bnc/>

103

NGSL, BSL, NAWL AND TSL on VocabProfile  
<http://www.lex tutor.ca/vp/bnc/>

104

NGSL on VocabProfile  
<http://www.lex tutor.ca/vp/bnc/>

Sample text from chapter 1 of Harry Potter...

105

NGSL on VocabProfile  
<http://www.lex tutor.ca/vp/bnc/>

106

NGSL on VocabProfile

Harry Potter coverage:  
86% old GSL  
93% NGSL

107

Hands-on time! Which of our many free tools do you want to try out?

108

**The New General Service List (NGSL)  
Project: A Modular, Corpus-based  
Approach for the Systematic Development  
of 2nd Language Vocabulary**

**MUCH MORE TO COME... THANK YOU!**

Dr. Charles Browne  
Professor of Applied Linguistics & TESOL  
Director, EFL Teacher Training Program  
Meiji Gakuin University, Tokyo, Japan  
[cbrowne@meiji.ac.jp](mailto:cbrowne@meiji.ac.jp)

A few of My Free Websites & Tools:  
[www.newgeneral-service-list.org](http://www.newgeneral-service-list.org)  
[www.newgeneral-service-list.org/2023/10/](http://www.newgeneral-service-list.org/2023/10/)  
[www.newgeneral-service-list.org/new-general-service-list/](http://www.newgeneral-service-list.org/new-general-service-list/)  
[www.az1.com/](http://www.az1.com/)  
[www.researchomatic.com/](http://www.researchomatic.com/)

