Optimizing EFL Vocabulary Learning with IRT and Online Technology

2008 AILA World Congress

Dr. Charles Browne Professor of Applied Linguistics Chairman, Teacher Training Program Meiji Gakuin University, Tokyo Japan

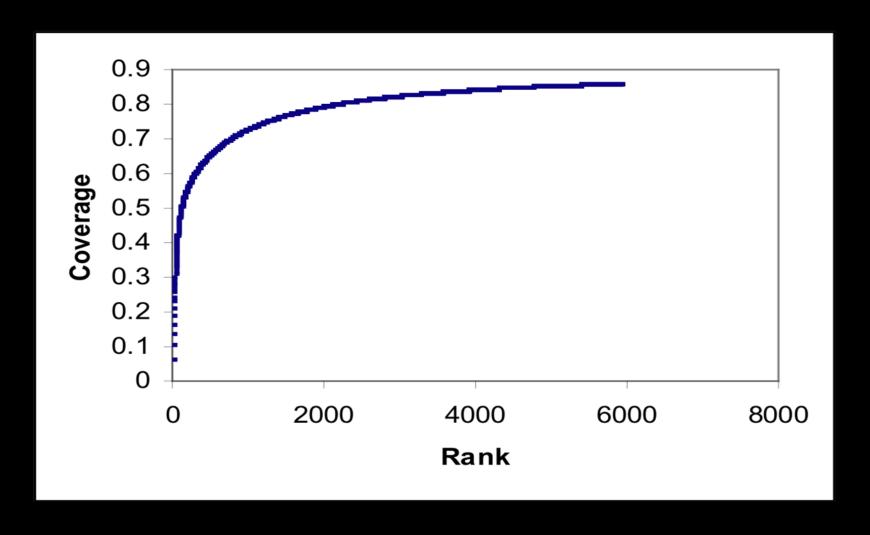
Outline of Presentation

- 1. Basic "facts" regarding vocabulary research
- 2. Introduction to problems faced by Japanese EFL learners related to vocabulary and reading
- Introduction to a variety of free online vocabulary & reading development software

The Importance of "Coverage"

Words	% known	# known	Researcher
1	7%	97/100	West(53), Nation(90)
10	25%	3/4	West(53), Nation(90)
100	50%	1/2	West(53), Nation(90),
1000	75%	1/4	West(53), Engles(68)
2000	85%	1/7	West(53), Nation(90)
4200	95%	1/20	Culligan (2008)
8000	98%	1/50	Laufer (92), Coady(93)
350,000	100%	100/100	Oxford English Dictionary

Coverage within the BNC for high frequency words (Leech, Rayson, & Wilson, 2001)



Problem 1:

EFL learners don't know enough high frequency words...

How many words do L2 learners know?

Minimum 4200 words needed for independent learning

Country	Vocab. Size	Hours of Instruction	Reference
Japan (University)	2000-2300	800-1200	Shillaw (95), Barrow (99)
China (English Majors)	4000	1800-2400	Laufer (99)
Indonesia (University)	1220	900	Nurweni & Read (99)
Oman (University)	2000	1350	Hort et al (98)
Israel (HS graduates)	3500	1500	Laufer (98)
France (HS students)	1000	400	Arnaud et al (85)
Greece (age 15, HS)	1680	660	Milton & Meara (98)
Germany (age 15, HS)	1200	400	Milton & Meara (98)

75% Coverage

is provided by 1000 high frequency words

16 missing words

Forestry

If	*	_ pla	anting	rates	are _	*	wit	h pla	anting	g	*	
	*	_ in 6	each	*		*	and	the f	forest	S_	*	_ at
the	ear	liest	oppo	rtunity	, the	*	WO	od s	suppli	es	coul	d
furt	her	incre	ease	to abo	ut 36	mill	ion	*	_ me	ters	S *	
in t	he _	*	_ 200	1-201	5. Th	e ad	dditior	nal _	*	WO	od	
sup	ply	shou	ıld gr	eatly _	*		*		*	, ev	en if	
mu	ch i	s use	ed for	*	prod	duct	tion.					

85% Coverage

is provided by 2000 high frequency words

9 missing words

Forestry

If* planting rates are maintained with planting	
targets satisfied in each*_ period and the forests	5
milled at the earliest opportunity, the* wood	
supplies could further increase to about 36 million*	
meters* in the period 2001-2015. The additio	nal
* wood supply should greatly exceed*	<i>*</i> '
even if much is used for * production.	

95% Coverage

is provided by 4200 high frequency words

1 missing word

Forestry

If current planting rates are maintained with planting targets satisfied in each annual period 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 energy production.

Vocabulary Thresholds:

• Below 80%, reading comprehension is almost impossible (Hu & Nation, 2001)

• 95% coverage is the point at which learners can read without the help of dictionaries (Laufer, 1989)

Problem 2:

Reading materials in Japan are too difficult...

Vocabulary & Readability: How do Japanese schools measure up?

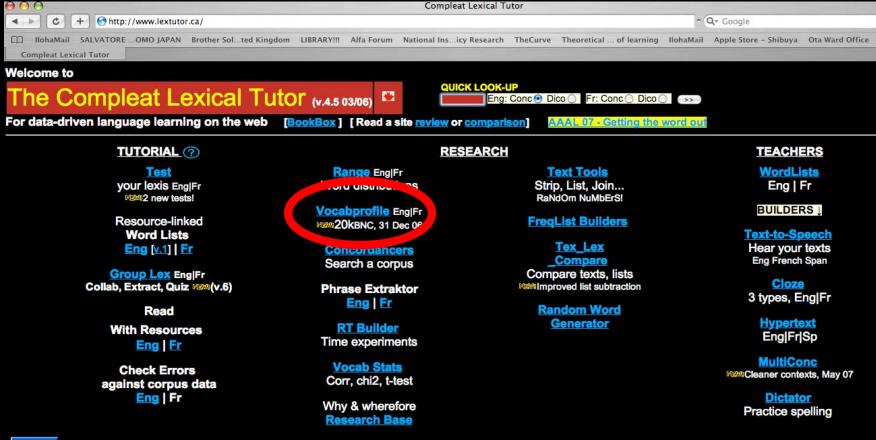
Junior High:

- Teaches first 1000 words fairly well
- Readability of texts seems good short passages, easy vocabulary, lots of pictures to support texts

Senior High:

- Focus changes dramatically to teaching of low frequency words
- Many, many words from 1000-2000 are never taught...
- Readability of texts is actually MORE difficult than unsimplified native speaker texts!

The Compleat Lexical Tutor www.lextutor.ca





UniversitÈ du QuÈbec : Montrèal Tom Cobb (Site likes IE5/6 for Windows.) To the free dissemination of knowledge on the WWW

Vocab Profile: Online Vocabulary Analysis Tool www.lextutor.ca

Home > VocabProfilers > English

Web VP v 2.5 Classic (300% speed-up on Jan 6, 2006)

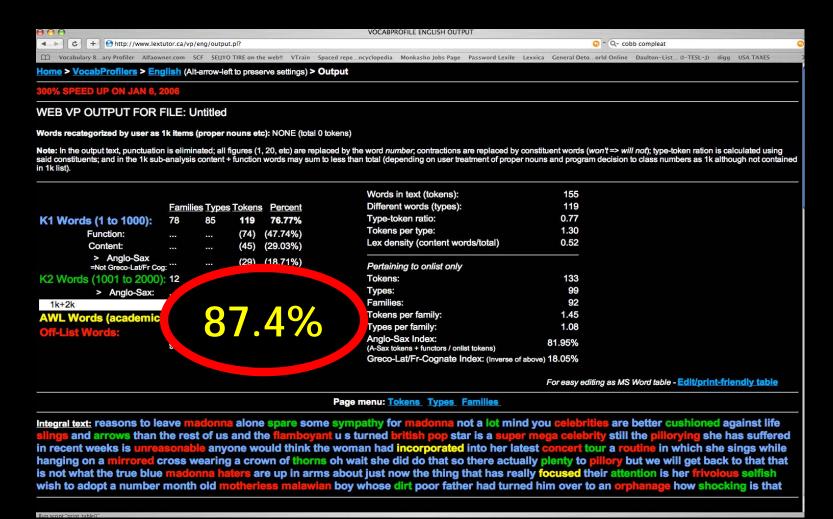
Input method A Type or paste smaller text (<2000 words) below and click Submit_window to see its Frequency Profile.



Original VP by A Heatley & P. Nation, VUW New Zealand & B. Laufer, U Haifa, Israel; AWL by A. Coxhead VUW; WebVP adapted by T. Cobb, UQAM Canada. How to reference

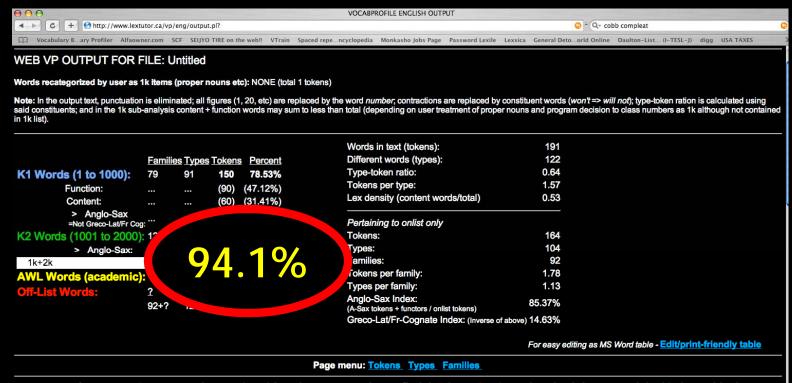
Typical English Newspaper in Japan

85% expected for 2000 words



Harry Potter Chapter 2

85% expected for 2000 words



Integral text: chapter two harry went down to breakfast the next morning to find the three dursleys already sitting around the kitchen table they evision a welcome home for the summer present for dudley who had been complaining loudly about the long walk were watching a brand new te on in the living room <mark>dudley</mark> had spent most of the summer in the kitchen his p little eves fixed on the between the fr sat down between dudley and uncle vernon a large be as he ate continually ha fv man with very little neck screen and his five chins and a lot of mustache far from wishing harry a happy birthday none of the dursleys made any sign that they had noticed harry enter the room but harry was far too used to this to care he helped himself to a piece of toast and then looked up at the reporter on the television who was halfway through a report on an escaped com rict the public is warned that black is armed and extremely dangerous a special ho ine has been set up and ing of black should be reported immediately

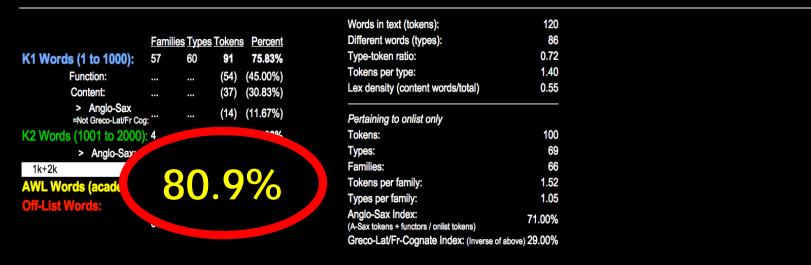
Typical Time Magazine Article

85% expected for 2000 words

WEB VP OUTPUT FOR FILE: Untitled

Words recategorized by user as 1k items (proper nouns etc): NONE (total 0 tokens)

Note: In the output text, punctuation is eliminated; all figures (1, 20, etc) are replaced by the word number, contractions are replaced by constituent words (won't => will not); type-token ration is calculated using said constituents; and in the 1k sub-analysis content + function words may sum to less than total (depending on user treatment of proper nouns and program decision to class numbers as 1k although not contained in 1k list).



For easy editing as MS Word table - Edit/print-friendly table

Page menu: Tokens Types Families

Integral text: the postponement of the first session of a high stakes summit between president bush and iraq prime minister nuri al maliki wednesday night in jordan may indeed be as officials on both sides took pains to stress simply a matter of logistics but reports on the outlook that each man was going to bring to the table suggests that the two may be on a collision course the meeting was touted as a crisis summit designed to set a new course for tackling iraq mounting violence civil war or whatever one chooses to call it the salient point is that iraq has spun so dangerously out of control that existing policies appear to offer no way out of the mayhem

Japanese High School Textbook (Spectrum Unit 16)

85% expected for 2000 words

WEB VP OUTPUT FOR FILE: Untitled

Words recategorized by user as 1k items (proper nouns etc): NONE (total 0 tokens)

Note: In the output text, punctuation is eliminated; all figures (1, 20, etc) are replaced by the word number, contractions are replaced by constituent words (won't => will not); type-token ration is calculated using said constituents; and in the 1k sub-analysis content + function words may sum to less than total (depending on user treatment of proper nouns and program decision to class numbers as 1k although not contained in 1k list).

					Words in text (tokens):	131
	Families	s Types	Tokens	<u>Percent</u>	Different words (types):	81
K1 Words (1 to 1000):	47	55	100	76.34%	Type-token ratio:	0.62
Function:			(71)	(54.20%)	Tokens per type:	1.62
Content:					Lex density (content words/total)	0.46
> Anglo-Sax =Not Greco-Lat/Fr Cog	g: ···		(15)	(11.45%)	Pertaining to onlist only	
K2 Words (1001 to 2000);	2			**	Tokens:	116
> Anglo-Sa					Types:	68
1k+2k				01	Families:	60
AWL Words (acad	/	6	X	%	Tokens per family:	1.93
Off-List Words:		.		/0	Types per family:	1.13
OII-List Words.					Anglo-Sax Index: (A-Sax tokens + functors / onlist tokens)	75.86%
					Greco-Lat/Fr-Cognate Index: (Inverse	of above) 24.14%

For easy editing as MS Word table - Edit/print-friendly table

Page menu: Tokens Types Families

Integral text: spectrum unit number to put it concisely walking is an inherent biological function of the human species not so language it is true that in a certain sense the individual is predestined to talk but that is due entirely to the circumstance he is born in and the lap of a societal situations that are certain to lead him to its traditions eliminate society and it is justified to reason that he will learn to walk if he survives at all or again remove the newborn individual from the societal environment into which he has come and transplant him to an utterly alien one he could possibly develop the skill of walking in this novel environment but his speech will be utterly at variance with the speech of his native environment

Summary: Difficulty of Typical Types of Text

Type of Text	% Coverage from knowing 2000 High Frequency Words
English Daily Newspaper	87%
Harry Potter	94%
Time Magazine	81%
Unsimplified Native Text	85%

Difficulty of Japanese Textbooks

(Browne, 1996, 1998, and in press)

Text	Coverage from 2000 High Frequency Words
Spectrum	71%
Milestone	78%
Unicorn	79%
Unsimplified Native Texts	85%

Difficulty of Japanese Entrance Exams

(Kikuchi, 2006, Browne & Kikuchi, 2008)

Text of Entrance Examinations for:	% Coverage from 2000 High Frequency Words
Keio Univ.	69%
Sophia Univ.	72%
Waseda Univ.	72%
Kyoto Univ.	77%
Nagoya Univ.	68%
Tokyo Univ.	80%

Examples of low frequency words found in high school textbooks:

- anemone
- hearth
- syncopate
- cockcrow
- neigh
- insularism
- cataclysm

- seedle
- prodigious
- lappilus
- whin
- pitchblende
- mazurka
- polonaise

Solution Number One:

Find out which high frequency words learners don't know and teach them

V-Check online vocabulary test

www.lexxica.com



- Identifies both <u>size</u> of vocabulary and <u>specific</u> words known
- Uses IRT to identify each word item's statistical difficulty
- A patented Computer Aided Test using Lexical Decision Tasks, IRT, and Signal Detection Theory

Lexical Decision Task

(Meara, 1992)

Do you know this word?



Yes

No

Lexical Decision Task

(Meara, 1992)

Do you know this word?



Yes

No

V-Check answers the questions:

What words do you already know? Which high frequency words are you missing?

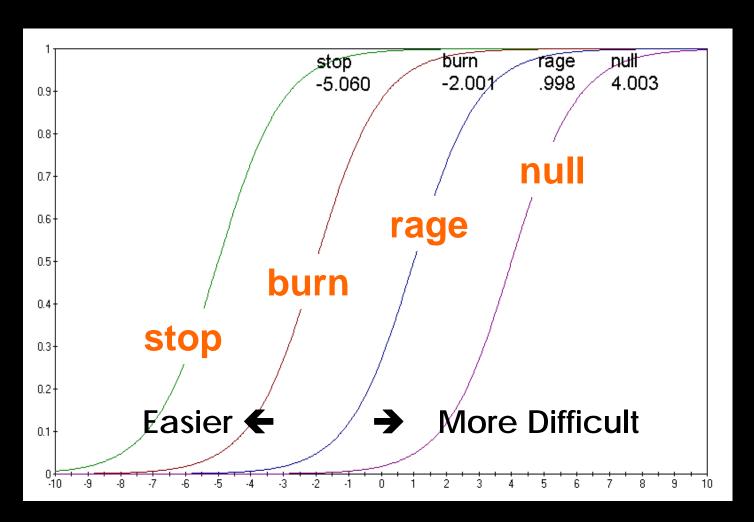




Methods for Estimating Lexical Size

Test	Gathers data on:	To find probability of knowing (<i>N</i>):	Formula for proportion of known words	Formula for Estimating Lexical Size
Paper-based Levels Test (NATION)	18 items per 1,000 words	1,000 words	$p = \frac{X_j}{N}$	pN
Paper-based Yes\No (MEARA)	40 items per 1,000 words	1,000 words	$p = \frac{P(H) - P(FA)}{1 - P(FA)}$	pN
CAT-based Yes\No (CULLIGAN/ BROWNE)	1,000 items per 1,000 words (IRT-based)	1,000 words	$P = \frac{e^{(\theta - b)}}{1 + e^{(\theta - b)}}$	$\sum_{i=1}^{N} P_i$

IRT generates item difficulty for each word as to each specific population





Word Frequency vs. Difficulty

injured hurt

Frequency

25x

55x

Average occurrences per million words

Word Frequency vs. Difficulty

	injured	hurt
Frequency	25x	55x
Difficulty	1.33	2.34
	People with a 1600 word vocabulary will know injured	People with a 2500 word vocabulary will know hurt



Which is more difficult?

tulip lily

snake crow

peanut <u>onion</u>

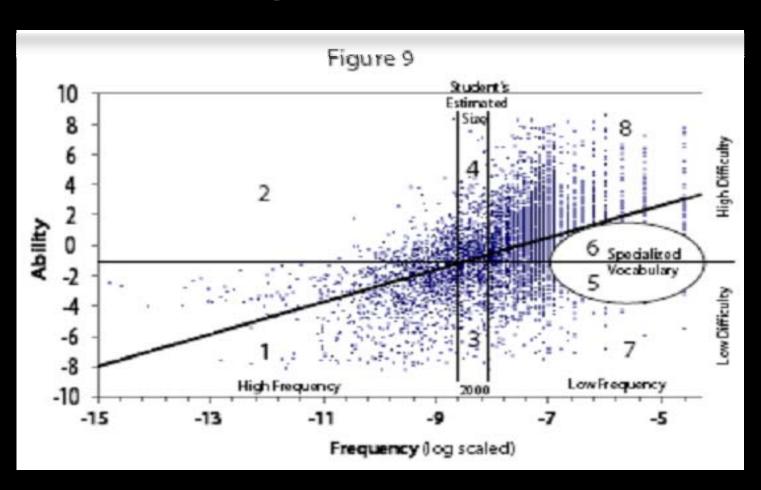
More difficult

tulip lily

snake crow

peanut onion

Word difficulty regression for Japanese learners



www.lexxica.com

Word Engin€ My personal coach



Increase your

Word Engine offers special purpose programs for subjects like TOEIC and TOEFL.

Learn with any of these popular rograms - rn. FEI



The 2874 high-frequency words that provide 95% coverage of all English in all situations!



The ultimate list of 7931 high-frequency words. Covers 99% of all English in all situations!



The 3 high-frequency words that provide 95. Horage of all Francisco TOEIC exams



The 2426 high-frequency words that provide 99% coverage of all Interchange textbooks.

is the comprehension and performance!

high-frequency words are those that of ur most often in each subject. K bwing them is essential to your secess.

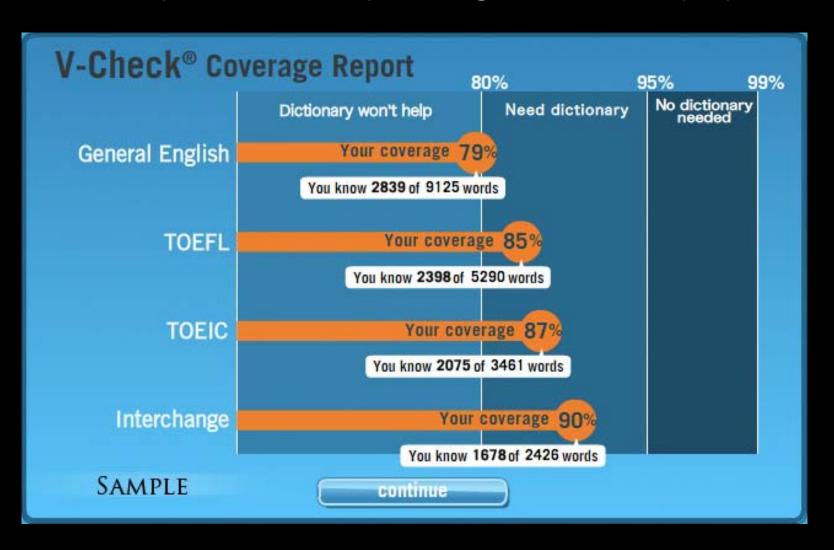
The patented V-Check system identifies all the specific high-frequency words you are missing and automatically prepares a personalized list for you to learn!

High speed learning tools help you develop instant recognition ability, and the built-in spaced repetition system insures long-term memory retention.

Finally, there's a fast and easy way to acquire high-frequency vocabulary!

Sample V-Check Coverage Report

V-Check reports vocabulary coverage for different purposes

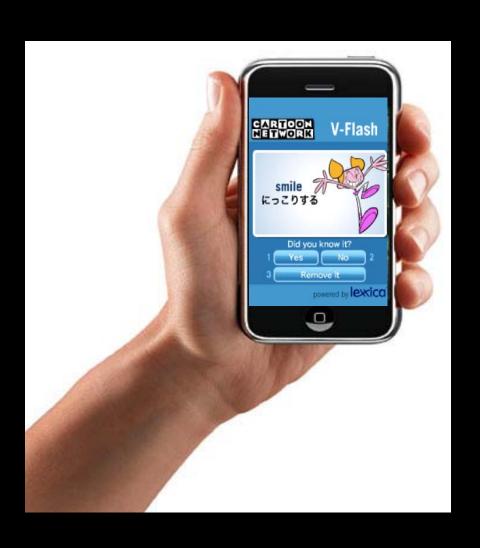


Solution Number Two:

High-speed vocabulary learning system

Personal target words based on each learner's actual needs

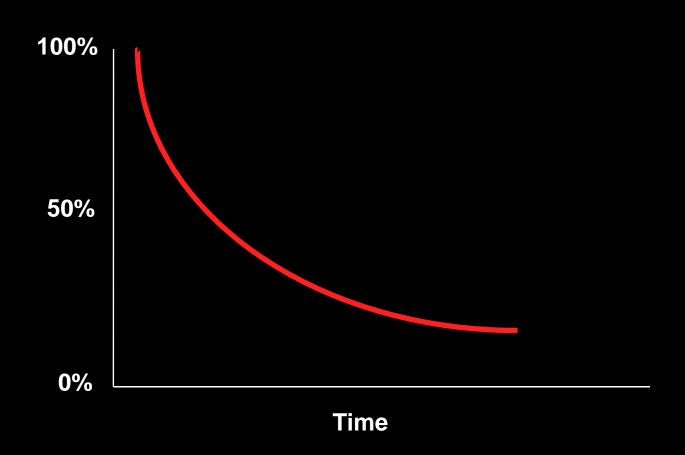
PC and Mobile Learning Games



Cell phones enjoy 100% penetration among Japanese university students versus just 54% for PCs

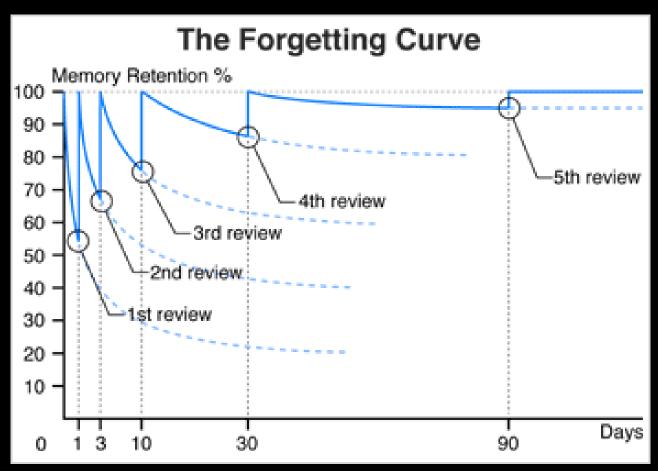
(Browne, 2007)

How do we overcome short term memory loss?



Learning via spaced repetition

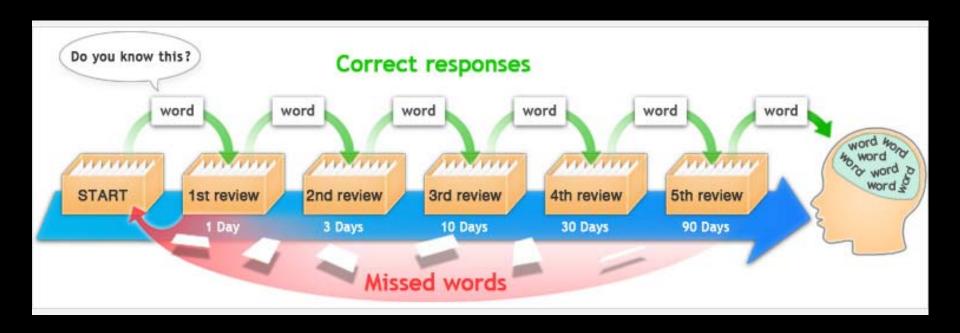
Ebbinhaus (1885), Leitner (1972), Pimsleur (1967), Mondria, (1994)



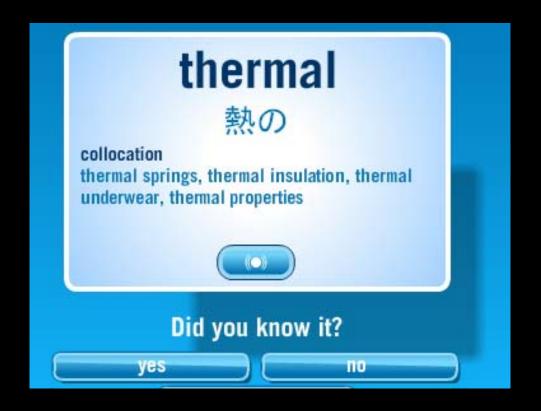
Repeated viewings foster long-term retention

A personal spaced repetition database

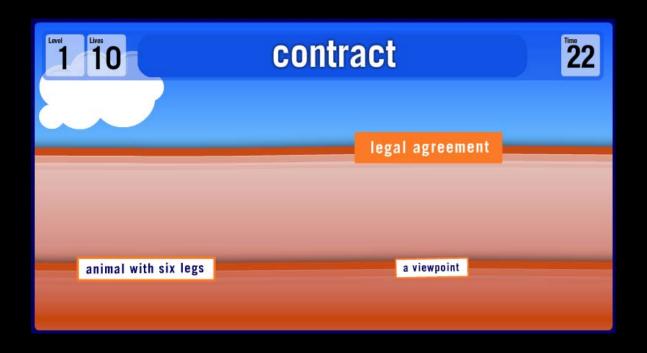
Words are automatically selected and repeated at increasing time intervals until the student acquires long-term retention



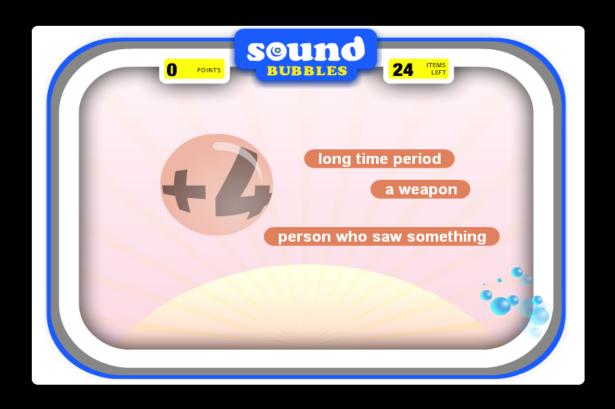
Flashcards for comprehension



SightWords for visual automaticity



SoundBubbles for aural automaticity



Mobile games for visual automaticity



Solution Number Three:

Graded reading online

Reading materials at student's 95% comprehension level

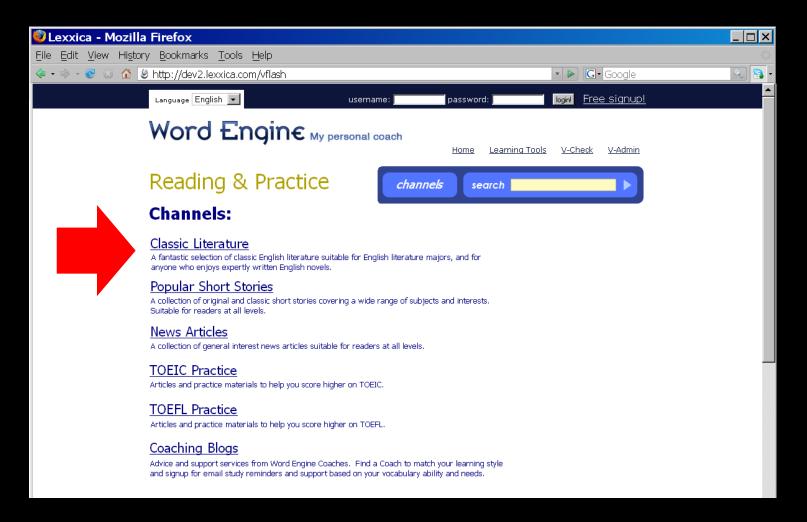
Graded Reading

Lexxica . com Available Spring 2009



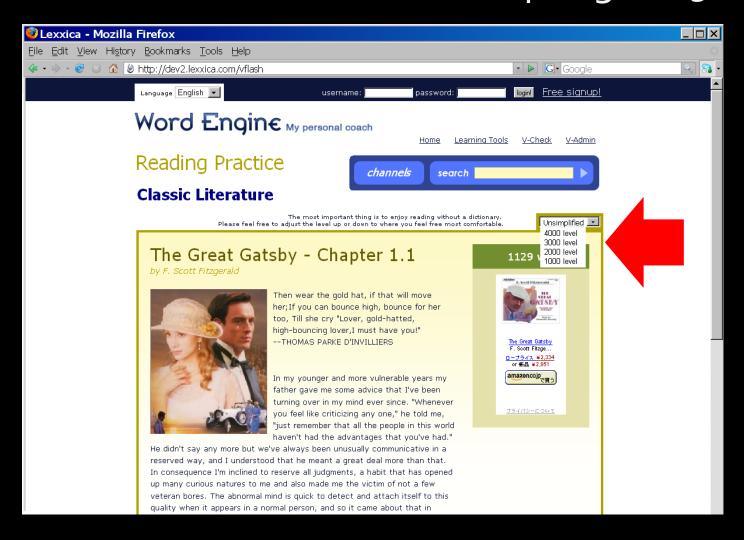
Graded Reading

Lexxica.com Available Spring 2009



Extensive Reading

Lexxica . com Available Spring 2009



Thank you!

Dr. Charles Browne, Professor of Linguistics Meiji Gakuin University, Dept. of English