# Technology and Language Learning

## Online Tools for Reading Skills and Vocabulary Development

Dr. Charles Browne, Professor of Linguistics Meiji Gakuin University, Dept. of English Director, Lexxica Corp. browne@ltr.meijigakuin.ac.jp

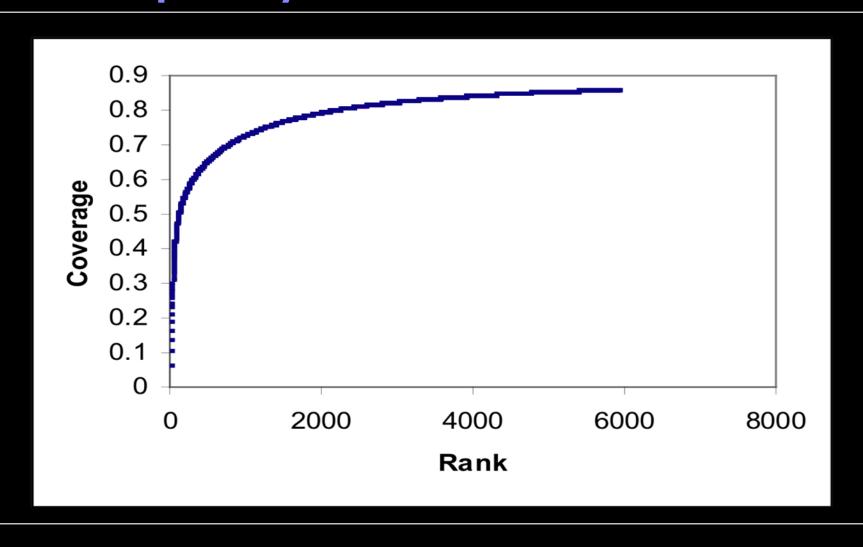
### **Outline of Presentation**

- 1. Basic "facts" regarding vocabulary research
- 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

Provided by 1000 high frequency words

16 missing words

### **Forestry**

<b>If</b> _	*	_ planti	ng rates	are _	*	_ with p	lanting	*	
	*	in eac	h <u>*</u>	; 	*	and the	forests	*	at
the	e ear	liest opp	portunity	, the _	*	_ wood	supplie	s coulc	
fur	ther	increas	e to abo	out 36	milli	on*	mete	ers <u>*</u>	
in i	the _	* 20	01-201	5. Th	e ad	ditional	* V	vood	
SU	pply	should	greatly <sub>-</sub>	*		*	<u>*</u>	even if	
mι	ıch i	s used f	or <u>*</u>	_ prod	ducti	ion.			

## 85% Coverage

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

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?

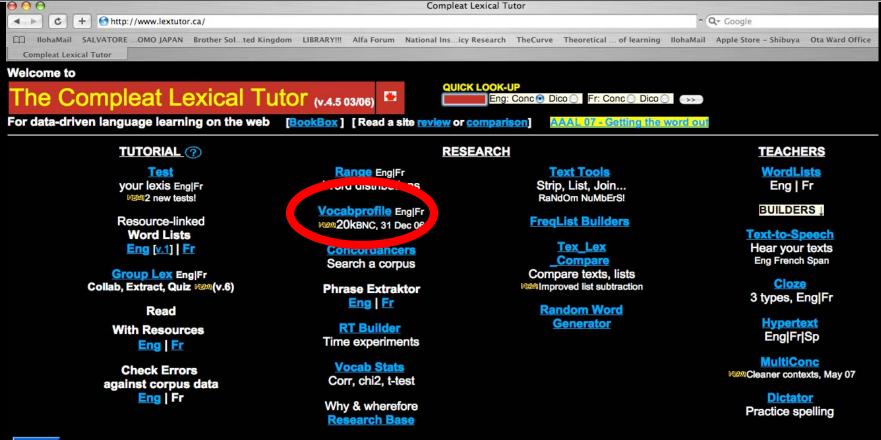
#### **Junior High:**

- 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



UQÀM

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 Graded Reader Results (1200 word level) 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).



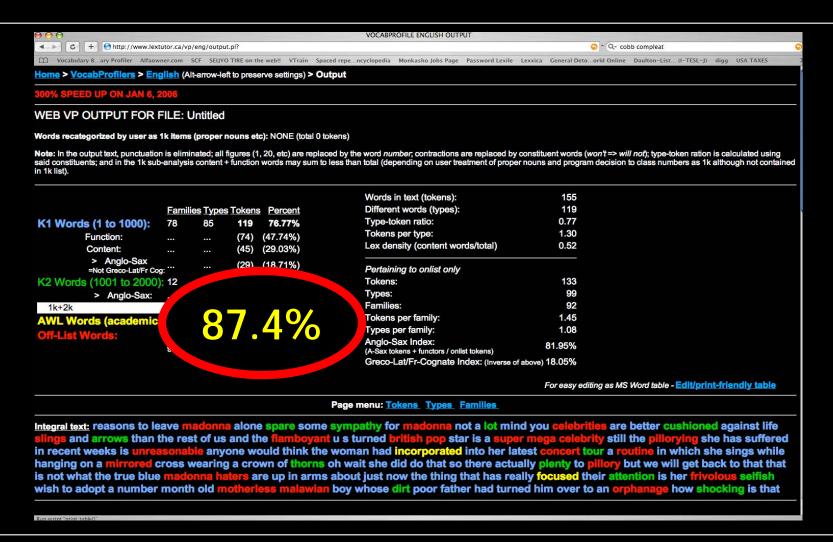
For easy editing as MS Word table - Edit/print-friendly table

Page menu: Tokens Types Families

Integral text: no thank you i do not want any ice just water please thank you he was your friend k said yes a very good friend too the colonel said putting his head between his hands dead michael gray dead i can not believe it it is sad k said but it is true tell me about him we were in the army together that was twenty five years ago then michael left the army married elizabeth and went into cavell and company that is lady elizabeth's family business she was a cavell i left the army five years ago i am not in the army dead michael gray dead i can not believe it

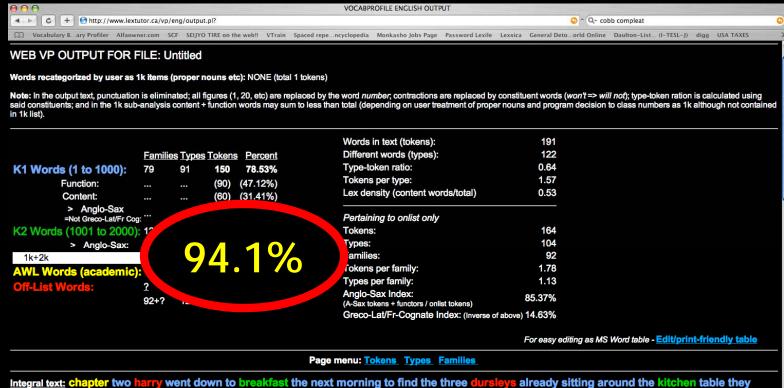
### Typical Yomiuri Newspaper Article

### 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 were watching a brand new television a welcome home for the summer present for dudley who had been complaining loudly about the long walk between the fridge and the television in the living room dudley had spent most of the summer in the kitchen his piggy little eyes fixed on the screen and his five chins wobbling as he ate continually harry sat down between dudley and uncle vernon a large beefy man with very little neck 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 convict the public is warned that black is armed and extremely dangerous a special hotline has been set up and any sighting of black should be reported immediately

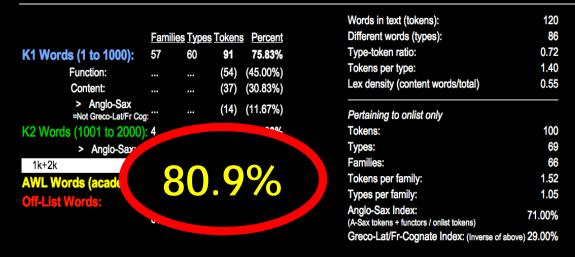
### **Typical Time Magazine Article**

### 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).

	Families	Types T	okens	Percent	Words in Different
K1 Words (1 to 1000): Function: Content:		55 	100 (71) (29)	<b>76.34%</b> (54.20%) (22.14%)	Type-tol Tokens Lex den
> Anglo-Sax =Not Greco-Lat/Fr Cog: K2 Words (1001 to 2000): > Anglo-Say			(15)	(11.45%)	Pertainii Tokens: Types:
1k+2k AWL Words (acade Off-List Words:	7	6.	8	%	Families Tokens Types p Anglo-S (A-Sax tok Greco-L

Words in text (tokens):	131
Different words (types):	81
Type-token ratio:	0.62
Tokens per type:	1.62
Lex density (content words/total)	0.46
Pertaining to onlist only	
Tokens:	116
Types:	68
Families:	60
Tokens per family:	1.93
Types per family:	1.13
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 of Vocab-Profile Results for Various Texts

Type of Text	% Coverage for 2000 High Frequency Words
Native Speaker Texts	85%
Graded Reader (1200 words)	99%
Yomiuri Newspaper	87%
Harry Potter	94%
Time Magazine	81%
Japanese HS textbook	77%

# 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	85%

### Are universities testing the right vocabulary?

(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

### Nation's Vocabulary Levels Test

(Nation, 1990)

- Estimates vocabulary size through "random" sampling of 36 words per word band (1000 ~5000 words)
- Doesn't make predictions about specific words known...
- Online version can be found at:

www.lextutor.ca

Vocabulary Levels Test (Recognition version)

#### 2000 level

matepond

<ol> <li>original</li> </ol>	
2. private	complete
3. royal	first
4. slow	not public
5. sorry	
6. total	
1. blame	
2. hide	keep away from sight
3. hit	have a bad effect on something
4. invite	ask
5. pour	
6. spoil	
1. basket	
2. crop	money paid regularly for doing a job
3. flesh	heat
4. salary	meat
5. temperature	
6. thread	
3000 Level	
1. administration	
2. angel	managing business and affairs
3. front	spirit who serves God
4. herd	group of animals
	g. cap or arminato

### **Lexical Decision Task**

(Meara, 1992)

Do you know this word?



### Lexical Decision Task

(Meara, 1992)

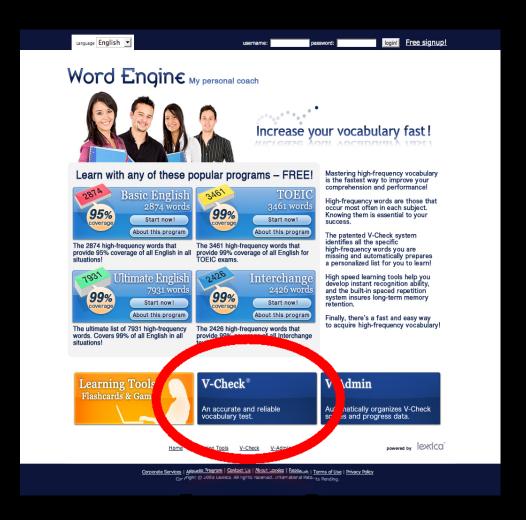
Do you know this word?



### V-Check online vocabulary test

#### www.lexxica.com

#### **Available June 2008**



- 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

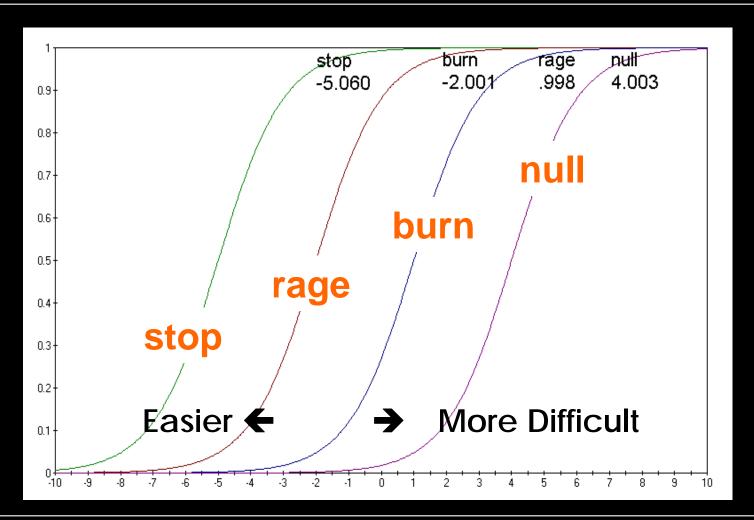
### V-Check answers the questions:

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





# IRT generates item difficulty for 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 <b>1600</b> word vocabulary will know injured	People with a <b>2500</b> word vocabulary will know hurt

### Which is more difficult?

tulip lily

snake crow

peanut <u>onion</u>

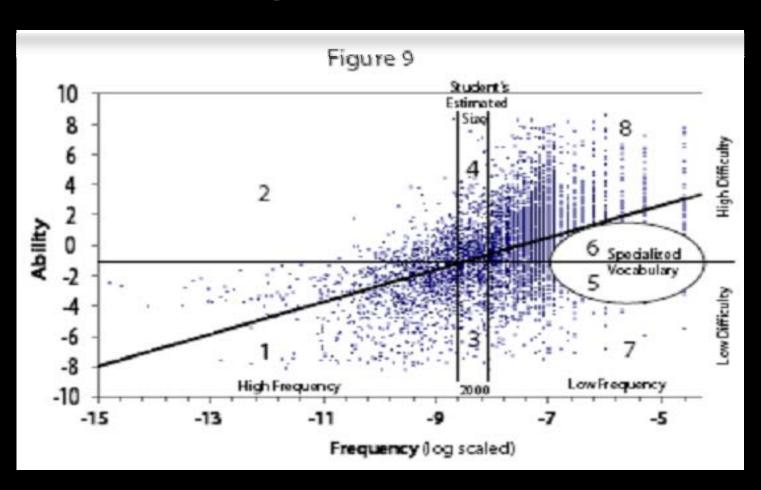
### More difficult

tulip lily

snake crow

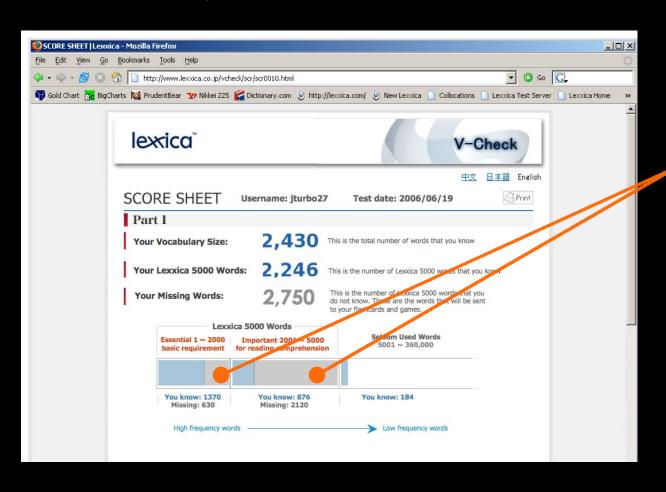
peanut onion

## Word difficulty regression for Japanese learners



### Sample V-Check Results Sheet

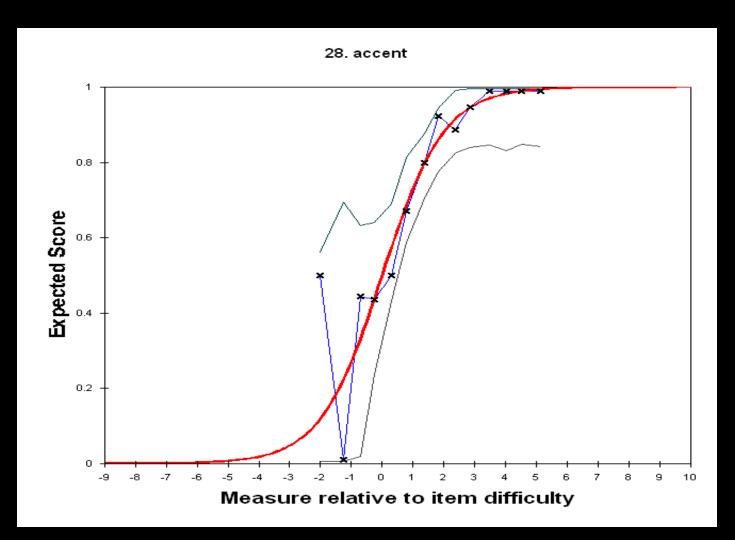
- 1) Estimates vocabulary size
- 2) Identifies specific known/unknown words



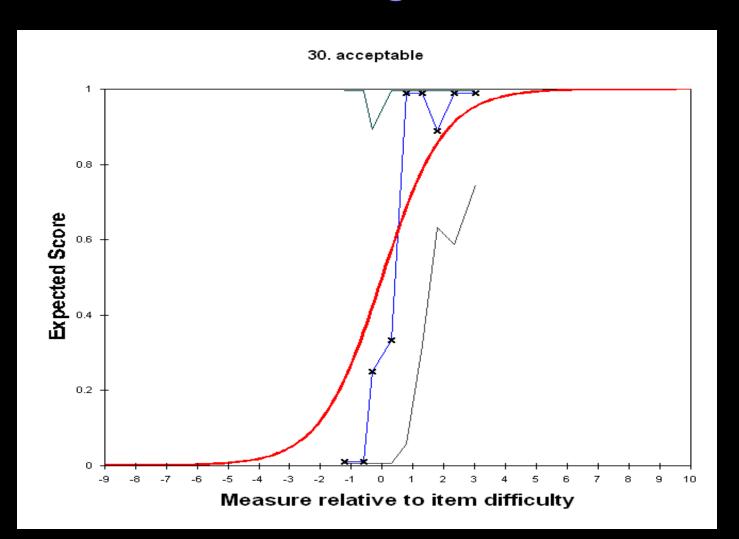
Note the gaps in knowledge of high frequency words



# Limitations to a Paper-based Test N = 385



# Limitations to a Paper-based Test N = 30



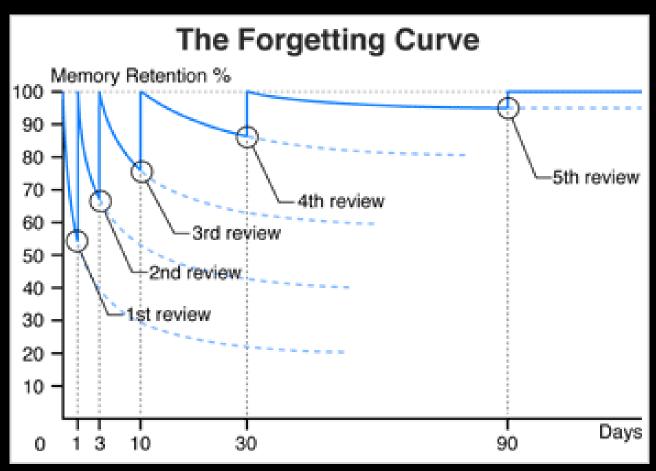
#### **Solution Number Two:**

Spaced repetition vocabulary learning system

Personal target word lists based on actual need

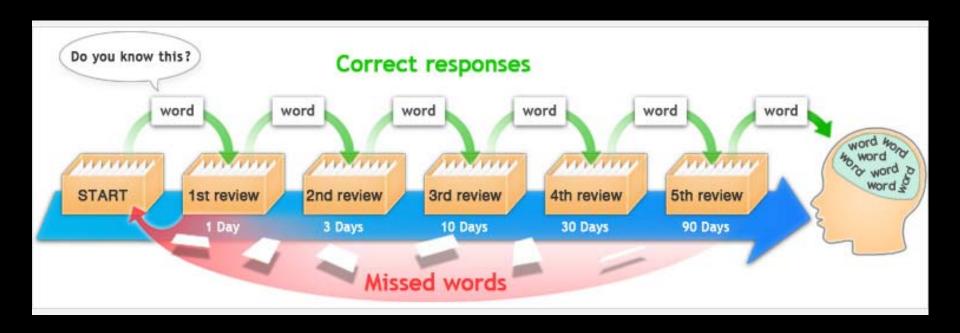
#### The Forgetting Curve

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

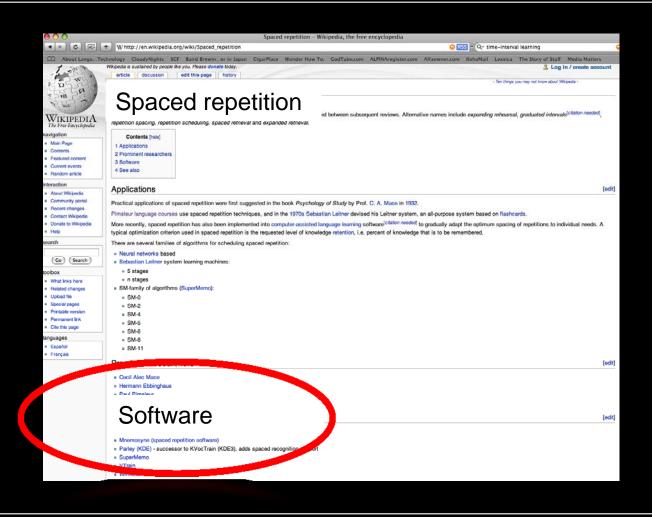


Repeated viewings foster long-term retention

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



# Programs using spaced-repetition mentioned in Wikipedia





Intro - Why Anki? - Screenshots - FAQ - Wiki - Forum Download - Study online

#### **Download Anki**



Latest version: 0.9.5.7

- Windows
- Mac OSX
- Debian
- Source (Linux/BSD)

Debian packages are in 'unstable' and 'testing' distributions. The testing packages lag behind by at least 10 days.

For more info and previous releases, see the sourceforge page.

#### Development

A development release is available, too. Install Mercurial and then grab the following two projects and put them in the same directory:

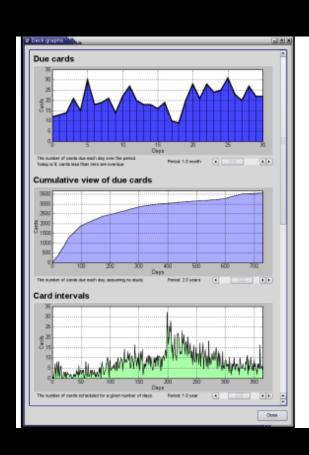
- \$ hg clone http://ichi2.net/hg/libanki
- \$ hg clone http://ichi2.net/hg/ankigt





#### Multimedia Support

Sound and graphics are also supported.



#### Statistics and Graphs

Anki provides a number of graphs related to your deck: cards due by date in a per-day and cumulative graph, interval distribution of cards, cards added a day, cards answered a day, and ease distribution over new, young and mature cards.



#### SuperMemo Screenshots

Here you can view 26 sample screenshots.

Point your mouse to a picture to see its description.

You can find more than 150 screenshots in the SuperMemo manual.

The manual can be read online or downloaded to your computer for offline reading.

















#### Question phase

With each card in a drill, you will be shown a question first. When you read it, you should try to recall the answer.

When you think you know the answer or when you cannot recall any answer at all, it is time to tap the *Show Answer* button. This will reveal the correct answer and you can proceed to the <u>Answer phase</u>.



#### Answer phase

When you see the correct answer, compare it to the answer you prepared in your mind before and grade its quality on the Right/Wrong scale. The evaluation is much easier than in the tests:

#### Right (perfect answer)

The card will be removed from the drill

#### Wrong (inaccurate or no answer)

The card will remain in the drill and will reappear later



#### Reverse drills

With SuperMemo you can also reverse your cards in drills. This is a very useful feature and it allows you to learn your cards in both directions. In reversed mode, SuperMemo will ask you showing the answer side of the card, and you have to reply recalling the question side.

Of course such mode is not suitable for all kinds of data. It is good for learning pairs such as Country to Capital (Capital to Country in reversed mode) or word pairs such as English to German (German to English in reversed mode).

To turn the reversed drills on or off for a particular database, tap the bottom left silksreen menu button and select Card / Reverse. You will see a dialog offering you Normal or Reversed drills.





#### lexica

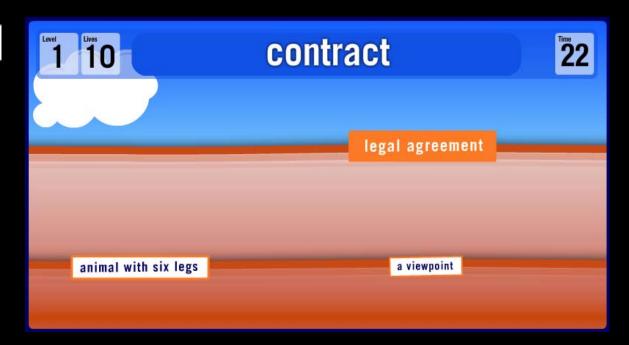
Flashcards for comprehension



Available June 2008

### lexica

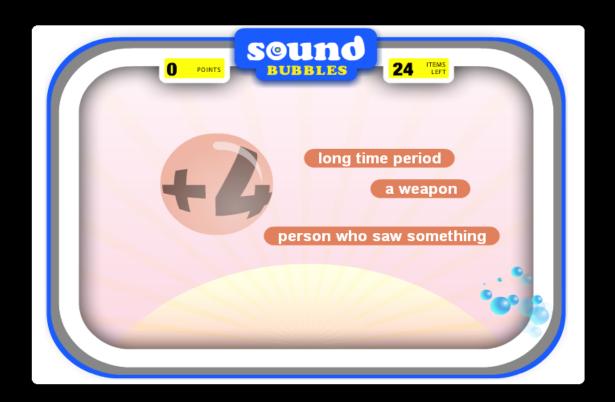
Visual automaticity games



Available June 2008

#### lexica

Aural automaticity games



Available June 2008

#### **Solution Number Three:**

Graded reading online

Reading materials at student's 95% comprehension level

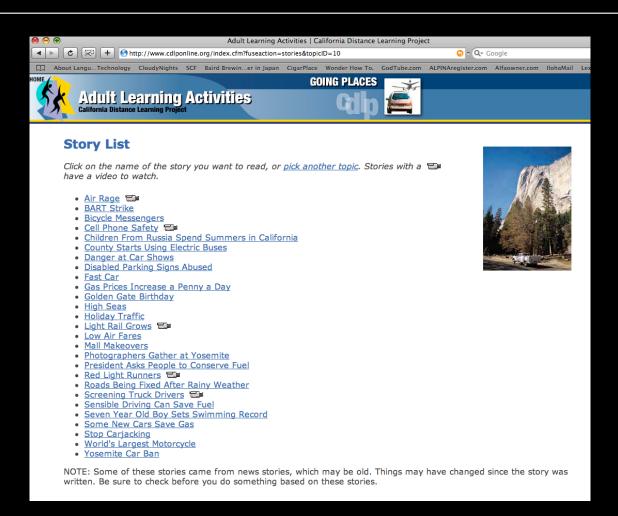
Graded Readers (books)

- Cambridge
- Penguin
- Oxford





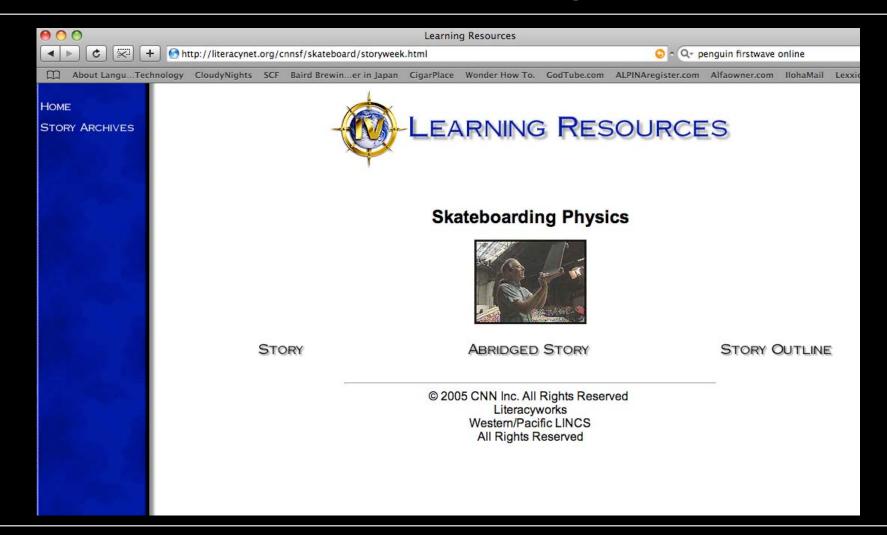
#### California Distance Learning Project



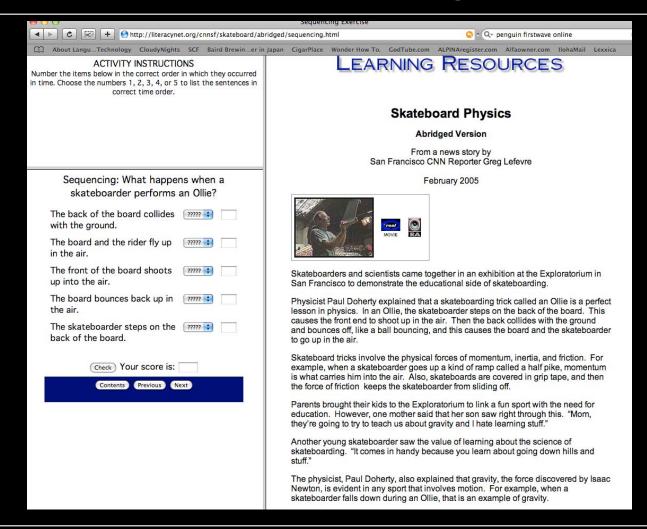
#### California Distance Learning Project



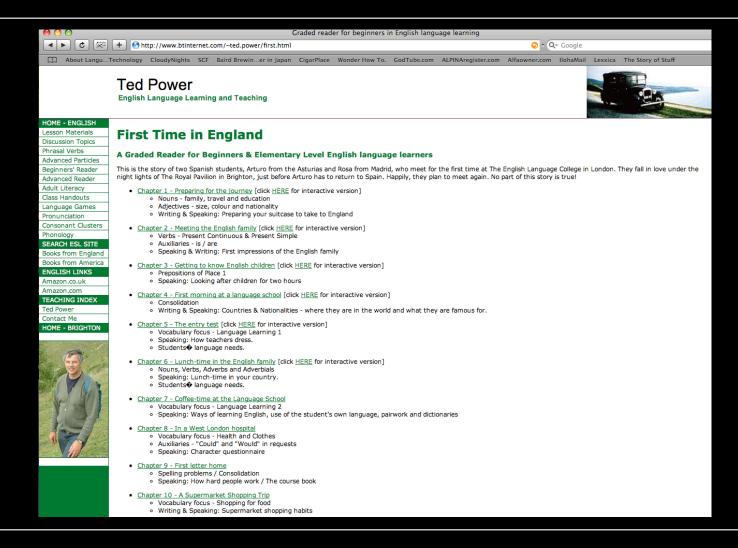
Literacy Net . Org



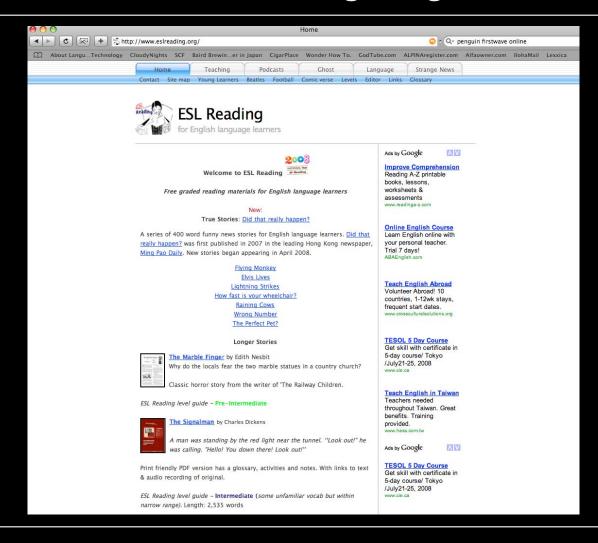
#### Literacy Net . Org



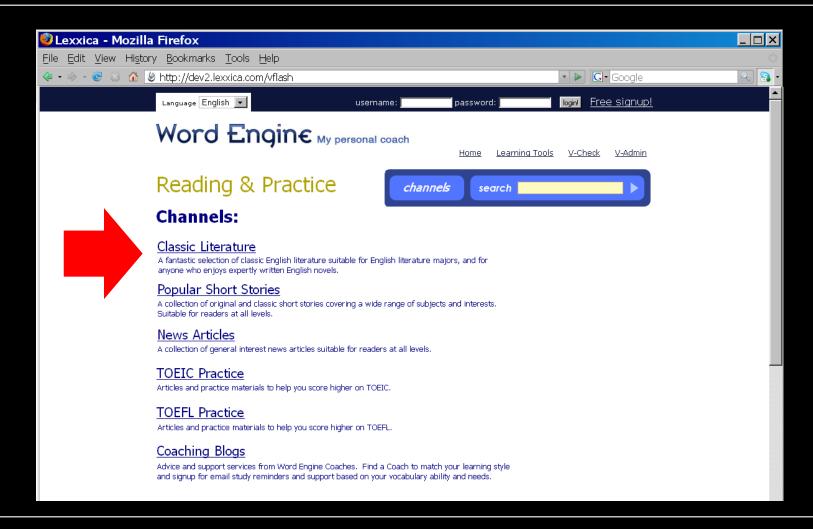
#### Ted Power's Website

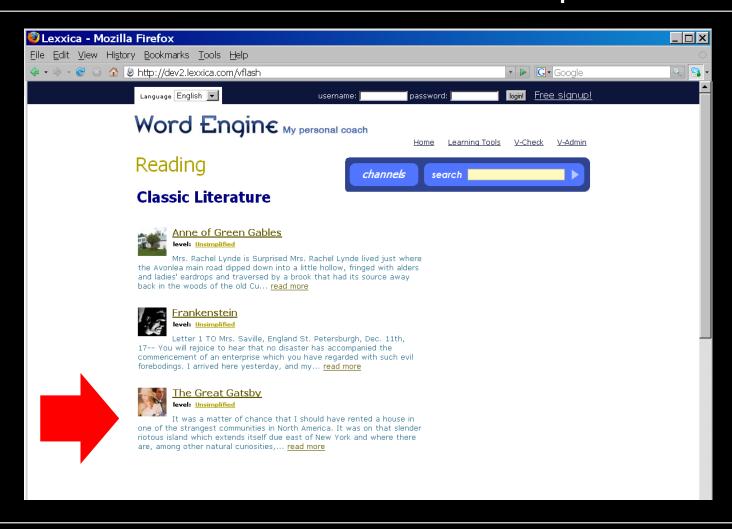


#### ESL Reading . Org



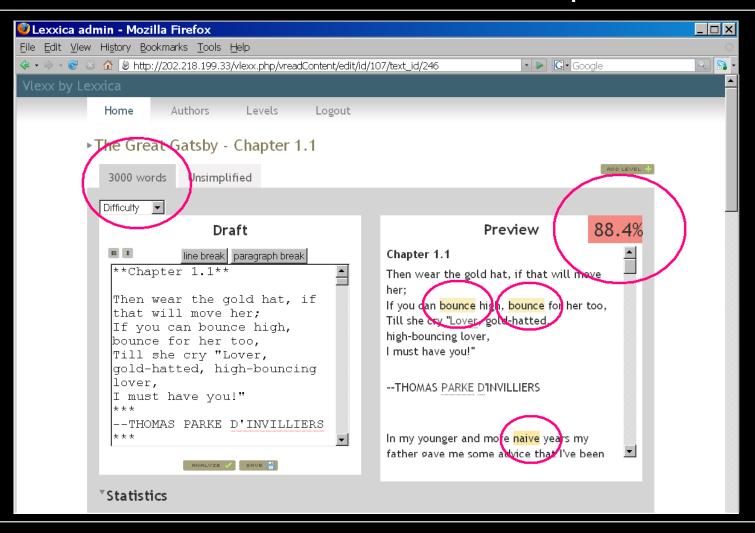




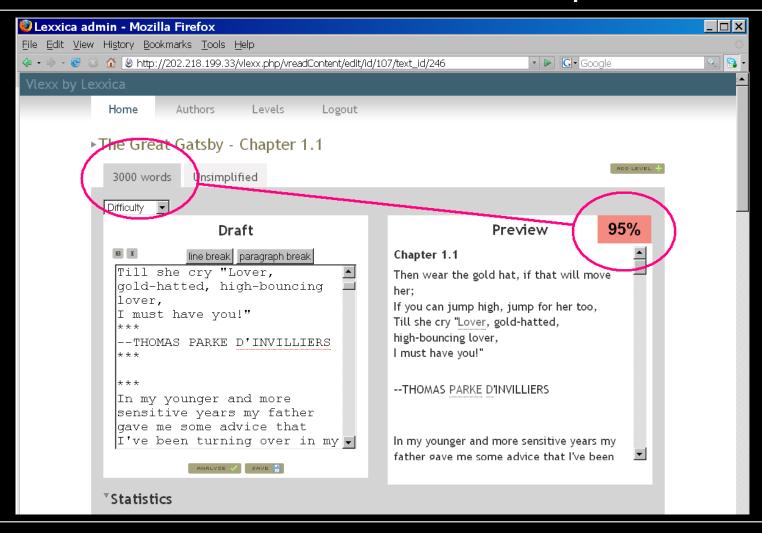




### V-Lexx Editing Application



# V-Lexx Editing Application



### www.lexxica.com

#### Word Engin€ My personal coach



Increase your vocabulary fast!

Learn with any of these popular programs – FREE!



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



pre 3461 pre 39% coverage of all English-EIC exams.



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



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

Example of a special purpose V-Check test and teaching lexicon developed specifically to assist learners using the Interchange textbook.

high-fr mis automatically prepares sonalized 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 retent on.

Final there's a fast and easy way to a quire high-frequency vocabulary!

# Thank you!

Dr. Charles Browne, Professor of Linguistics Meiji Gakuin University, Dept. of English Director, Lexxica Corp. browne@ltr.meijigakuin.ac.jp