

# Not the Word I Wanted? How Online English Learners' Dictionaries Deal with Misspelled Words

Robert Lew & Roger Mitton  
Bled, eLEX2011

# Spelling and dictionary use

- even the best content is useless if ... the user can't find it
- what a user needs to access a word:
  - familiarity with access alphabet
  - spelling representation
- in e-dictionaries, alphabetic sequencing is not so crucial, but spelling representation is still quite important
- but knowing the spelling can be a problem

# When is spelling a problem?

- word mistyped (e.g. *spellign* → performance error)
- word misspelled (e.g. *speling* → competence error, erroneous representation)
- word heard only, spelling guessed (e.g. *clime* for climb → may be no lexical representation, sound-mediated)
- special needs of non-native speakers (e.g. learners): they are often *poor spellers*, though not necessarily in ways typical of native speakers

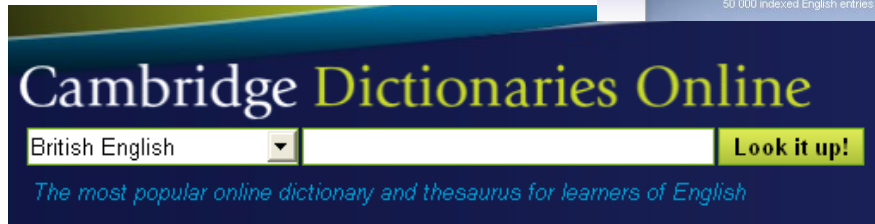
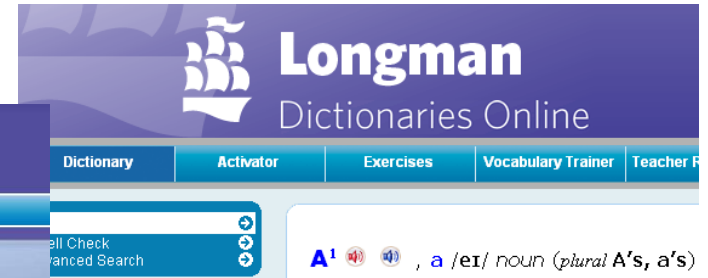
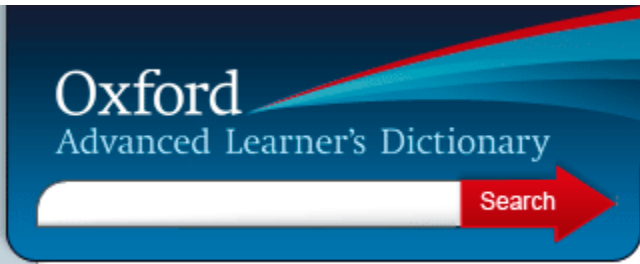
# Dictionaries and misspelled words

- a good dictionary interface should be able to guess the users' intention even if they misrepresent the orthographic form of the word
- but, in an analysis of three online German dictionaries (Bank 2010), only one dictionary was found to be at all 'rechtschreibtolerant'
- a dictionary should:
  - make reasonable guesses at the intended word
  - order the guesses with most likely ones at the top
- while no system can guess all misspellings correctly due to their unpredictable nature, it should make an honest attempt to maximize the success

# Research questions

- how good are online English monolingual learners' dictionaries at guessing headwords misspelled by learners of English?
- do the dictionaries perform similarly?
- can the performance be improved?

# Dictionaries tested



# Misspellings

- our corpus of misspellings consisted of three sets:
  1. **Polish:** two first-year Polish university students wrote down items from Nordquist's list of 200 common misspellings, presented in audio (100 items)
  2. **Japanese:** taken from the SAMANTHA Error Corpus created by Takeshi Okada at Tohoku University (50 items)
  3. **Finnish:** obtained from the Birkbeck spelling error corpus (Mitton 1985) originally collected by Suomi (1984) (50 items)

# Procedure

- misspellings of all 200 target words were looked up manually in the seven dictionaries
- position of the target word on the suggestions list was noted, relative to other (irrelevant) hits





## Results for tempori

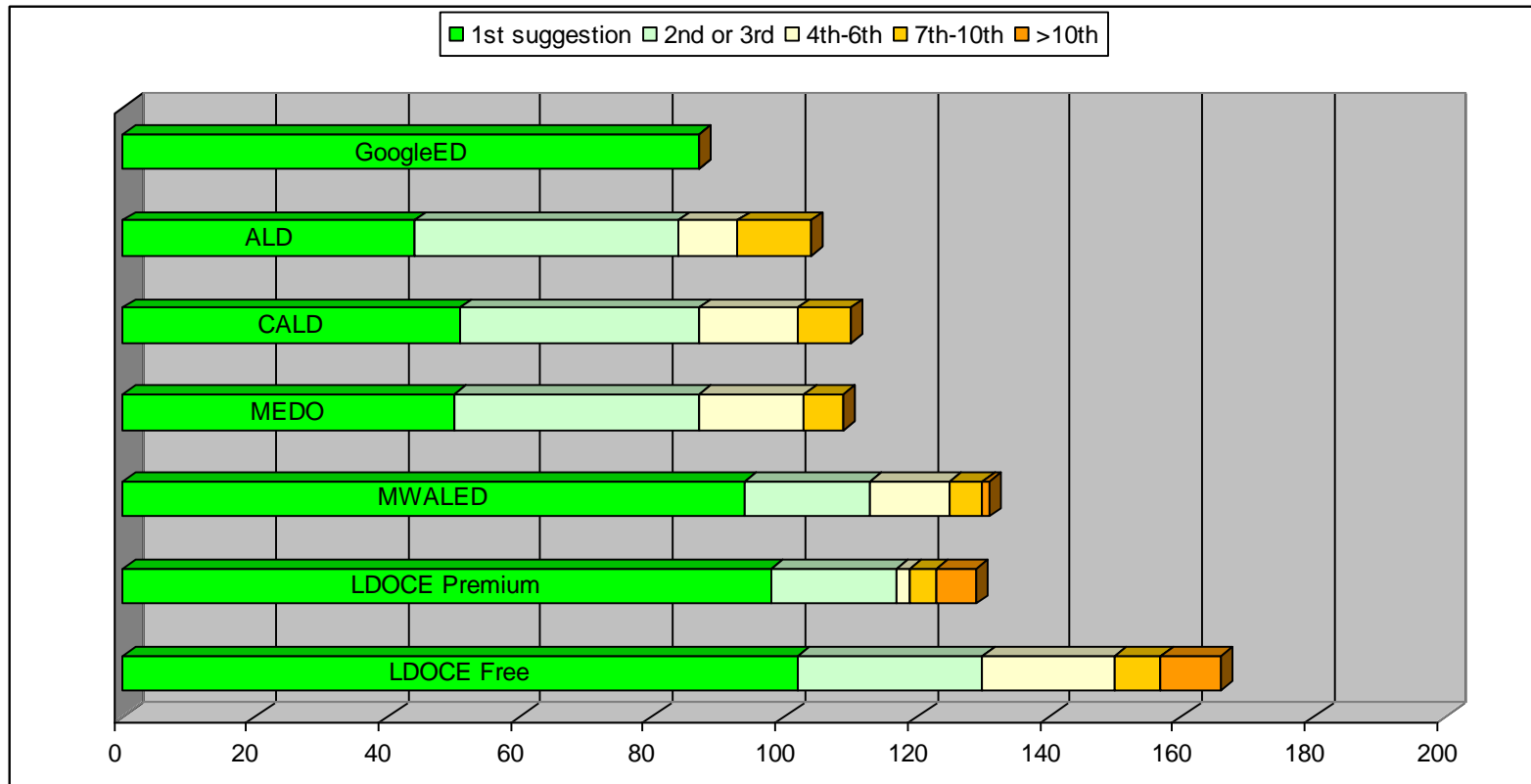
**tempori** was not found

Did you spell it correctly? Here are some alternatives:

- [temporise](#)
- [temporize](#)
- [emporia](#)
- [temporal](#)
- [temporizes](#)
- [temporized](#)
- [temporised](#)
- [temporises](#)
- [temporary](#)
- [tempering](#)

target word at  
position 9

# How the dictionaries performed



# Where the dictionaries failed

- ALD: too much weight attached to substring matching

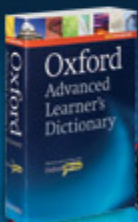
Learn on the move  
with the OALD App



Learn how to plan, write and  
check your written work

Oxford **iWriter**

1 million copies sold




Find out why people  
choose it, use it,  
love it

Put the OALD search box  
on your own website.

## Sorry, no search result for sakology.

Did you mean:

- apology
- sexology
- sinology
- ecology
- zoology
- ufology
- urology
- geology
- cytology
- tautology



ALD does not give much regard to the first letter (which is known to be correct most of the time)

[HOME](#)[WORDLISTS](#)[USING OALD](#)

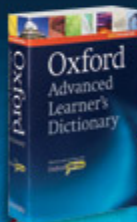
Learn on the move  
with the OALD App



Learn how to plan, write and  
check your written work

Oxford **iWriter**

1 million copies sold




Find out why people  
choose it, use it,  
love it

Put the OALD search box  
on your own website.

## Sorry, no search result for lajbery.

Did you mean:

- robbery
- rubbery
- lumberyard
- bribery



ALD, CALD, and MEDO tend to give preference to plural suggestions for no particular reason

---

Sorry, no search result for **sitizen**.

---

*Did you mean:*

- **citizens**
- **citizen**
- **citizenry**
- **siting**
- **netizens**
- **sitins**
- **sitin**
- **netizen**
- **sanitizing**
- **sanitizes**



Search

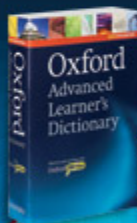
Learn on the move  
with the OALD App



Learn how to plan, write and  
check your written work

Oxford **iWriter**

1 million copies sold



Find out why people  
choose it, use it,  
love it

staycation

Search

Put the OALD search box  
on your own website.

## Sorry, no search result for spaghetti.

Did you mean:

- spaghetti
- spaghetti
- spacesuit
- spaceship
- spatted
- spating
- spatter
- spastic
- spaghetti western
- suffragettes



Why would all of ALD, CALD, and MEDO be affected by this overeagerness to tag on –s?

- all three use the same software for dictionary compilation and publication: the DPS Writing System from IDM
- but LDOCE also uses the same system and yet it does not exhibit the –s problem



At times, suggestions are bewildering

Search: 

Learn



Practice



Blog



Shop



Help

Learner's

My Dictionary

Search History

3,000 Words

Manage All

The word you have entered is not in the dictionary. Click on a spelling suggestion below or try your search again.

[archdiocese](#)

## Flash Cards — Drag and Drop

Manage

Practice

Title:



### Instructions

- 1.) Register and sign in to [LearnersDictionary.com](#).
- 2.) Give your new set of Flash Cards a title or select a set you have already created. Click the arrow button to save the title of a new set of Flash Cards.
- 3.) Drag words from your MyDictionary list and drop them here to add them to your Flash Cards.
- 4.) Click "Practice" to begin using your Flash Cards.

### [Learn about the book...](#)

#### **Merriam-Webster's Advanced Learner's Dictionary**

It's here! The first learner's dictionary from America's language experts is the perfect resource for teachers and students of English as a second language, with over 100,000 words and phrases, loads of drawings and illustrations, and the most example sentences of any learner's dictionary. It also includes a free eBook download. [Learn more](#)



Search:

- [Learner's](#)
- [My Dictionary](#)
- [Search History](#)
- [3,000 Words](#)
- [Manage All](#)

3 entries found for **das**.

- [cream soda \(noun\)](#)
- [giant panda \(noun\)](#)
- [piña colada \(noun\)](#)

## cream soda noun

*plural* ~ -das

[*count, noncount*] : a type of sweet, bubbly soft drink that is flavored with vanilla

Share this entry:

Link to this page:



### Flash Cards — Drag and Drop

Manage
Practice

Title:

- Instructions**
- 1.) Register and sign in to LearnersDictionary.com.
  - 2.) Give your new set of Flash Cards a title or select a set you have already created. Click the arrow button to save the title of a new set of Flash Cards.
  - 3.) Drag words from your MyDictionary list and drop them here to add them to your Flash Cards.
  - 4.) Click "Practice" to begin using your Flash Cards.

**Browse the Learner's Dictionary**  
**A B C D E F G H I J K L M N O P Q R S T U V W X Y Z**



Search:



Learn



Practice



Blog



Shop



Help

- Learner's
- My Dictionary
- Search History
- 3,000 Words
- Manage All

The word you have entered is not in the dictionary. Click on a spelling suggestion below or try your search again.

- [spigot](#)
- [spectate](#)
- [spotted](#)
- [spotlight](#)
- [speculate](#)
- [spectacle](#)
- [septet](#)
- [aseptic](#)
- [sabotage](#)
- [septic](#)
- [sceptical](#)
- [sceptic](#)
- [seepage](#)
- [sceptically](#)
- [slippage](#)
- [spatula](#)

## Flash Cards — Drag and Drop

Manage
Practice

Title:

➔ ✕


### Instructions

- 1.) Register and sign in to LearnersDictionary.com.
- 2.) Give your new set of Flash Cards a title or select a set you have already created. Click the arrow button to save the title of a new set of Flash Cards.
- 3.) Drag words from your MyDictionary list and drop them here to add them to your Flash Cards.
- 4.) Click "Practice" to begin using your Flash Cards.

### Learn about the book...

## Merriam-Webster's Advanced Learner's Dictionary

It's here! The first learner's dictionary from America's language experts is the perfect resource for teachers and students of English as a second language, with over 100,000 words and phrases, loads of drawings and illustrations, and the most example sentences of any learner's dictionary. It also includes a free eBook download. [Learn more](#)



MWALED's algorithm seems to focus excessively on transpositions or even scrambling

Search: 

Learn



Practice



Blog



Shop



Help

[Learner's](#)
[My Dictionary](#)
[Search History](#)
[3,000 Words](#)
[Manage All](#)

The word you have entered is not in the dictionary. Click on a spelling suggestion below or try your search again.

[crane](#)[cream](#)[crone](#)[croon](#)[crown](#)[crayon](#)[careen](#)[creme](#)[cram](#)[cornea](#)[cringe](#)[cairn](#)[crime](#)[crying](#)[krone](#)[crony](#)[cranny](#)[corn](#)[creamy](#)[Korean](#)

## Flash Cards — Drag and Drop

Manage

Practice

Title: 

## Instructions

- 1.) Register and sign in to LearnersDictionary.com.
- 2.) Give your new set of Flash Cards a title or select a set you have already created. Click the arrow button to save the title of a new set of Flash Cards.
- 3.) Drag words from your MyDictionary list and drop them here to add them to your Flash Cards.
- 4.) Click "Practice" to begin using your Flash Cards.

*Learn about the book...*
**Merriam-Webster's  
Advanced Learner's Dictionary**

It's here! The first learner's dictionary from America's language experts is the perfect resource for teachers and students of English as a second language, with over 100,000 words and phrases, loads of drawings and illustrations, and the most example sentences of any learner's dictionary. It



# Rare and marginal words

- a largish wordlist (of correct spellings) can affect negatively the performance of a spellchecker
- a particular problem for GoogleED, which gives the following unhelpful suggestions among the 'best' guesses: *deferens* (probably from *vas deferens*), *etyma*, *xylem*, *inf*, *umbrae*, *commis*, as well as proper names like *Du Bois*, *Tok Pisin* and *Wat Tyler*

# Oddball suggestions

- GoogleED sometimes provided suggestions which are clearly *not* genuine words



sinsirli



Advanced search

Everything

Images

Videos

News

Shopping

More

**Any time**

Past hour

Past 24 hours

Past week

Past month

Past year

Custom range...

All results

Sites with images

Related searches

Timeline

**Dictionary**

Reading level

Social

Nearby

Custom Location

Translated foreign pages

Reset tools

**Dictionary**

Did you mean: [sinirli](#)

No standard web pages containing all your search terms were found.

Your search - **sinsirli** - did not match any documents. [Reset search tools](#)

Suggestions:

- Make sure all words are spelled correctly.
- Try different keywords.
- Try more general keywords.

[Google Home](#)

[Advertising Programs](#)

[Business Solutions](#)

[Privacy](#)

[About Google](#)



tempreczer



Advanced search

Everything

Images

Videos

News

Shopping

More

**Any time**

Past hour

Past 24 hours

Past week

Past month

Past year

Custom range...

All results

Sites with images

Related searches

Timeline

**Dictionary**

Reading level

Social

Nearby

Custom Location

Translated foreign pages

Reset tools

**Dictionary**

▶ Did you mean: [tempricher](#)

No standard web pages containing all your search terms were found.

Your search - **tempreczer** - did not match any documents. [Reset search tools](#)

Suggestions:

- Make sure all words are spelled correctly.
- Try different keywords.
- Try more general keywords.

[Search Help](#)

[Give us feedback](#)

[Google Home](#)

[Advertising Programs](#)

[Business Solutions](#)

[Privacy](#)

[About Google](#)



tempreczer



[Advanced search](#)

Everything

Images

Videos

News

Shopping

More

### Dictionary

Your search - **tempricher** - did not match any documents. [Reset search tools](#)

Suggestions:

- Make sure all words are spelled correctly.
- Try different keywords.
- Try more general keywords.

### Any time

Past hour

Past 24 hours

Past week

Past month

Past year

Custom range...

All results

Sites with images

Related searches

Timeline

### Dictionary

Reading level

Social

Nearby

Custom Location

Translated foreign pages

Reset tools

[Search Help](#)

[Give us feedback](#)

[Google Home](#)

[Advertising Programs](#)

[Business Solutions](#)

[Privacy](#)

[About Google](#)

# The *-ing* ending (example: *\*useing*)

- another cause of difficulty for these dictionaries
- only GoogleED was able to correct *\*useing* to the intended *using*
- LDOCE Free and MEDO offered *unseeing*

[Search](#)[Widgets](#)[How to use](#)[About LDOCE](#)

## Do you mean:

unseeing  
Seeing  
seeing  
icing  
sing  
assign  
busing  
asking  
ageing  
eyeing  
sign  
design  
nursing  
resign  
upswing  
singe  
eying  
sewing  
swing  
sin  
use

# The *-ing* ending (example: *\*useing*)

- another cause of difficulty for these dictionaries
- only GoogleED was able to correct *\*useing* to the intended *using*
- LDOCE Free and MEDO offered *unseeing*
- LDOCE Premium suggested *suing*
- MWALED *seeing*
- CALD is strangest of all





## Results for using

**using** was not found

Did you spell it correctly? Here are some alternatives:

- [using](#)
- [hoeing](#)
- [aweing](#)
- [beings](#)
- [usedn't](#)
- [dyeing](#)
- [cueing](#)
- [ageing](#)
- [eyeing](#)
- [eyeings](#)

**Do you mean:**

dining  
diving  
diggings  
doing  
ducking  
edging  
dialing  
doping  
during  
rigging  
decking  
ticking  
aging  
digit  
dignify  
dying  
djinn  
going  
Viking  
daring  
dating  
divine  
doggone  
gaging  
hiking  
liking  
raging  
tagging  
tiling  
timing  
tining  
tiring  
lodging  
skiing  
dung  
deign



Search: diging



Learn



Practice



Blog



Shop



Help

[Learner's](#) [My Dictionary](#) [Search History](#) [3,000 Words](#) [Manage All](#)

The word you have entered is not in the dictionary. Click on a spelling suggestion below or try your search again.

- [Diegan](#)
- [dating](#)
- [doting](#)
- [gliding](#)
- [editing](#)
- [jogging](#)
- [gelding](#)
- [doggone](#)
- [decking](#)
- [ducting](#)
- [Dijon](#)
- [dudgeon](#)
- [detain](#)
- [joking](#)
- [ticking](#)
- [diagonal](#)
- [dotingly](#)
- [guideline](#)
- [dignity](#)



## Flash Cards — Drag

Manage

Practice

Title:



### Instructions

- 1.) Register and sign in to LearnersDict
- 2.) Give your new set of Flash Cards a title. If you have already created a set, click the arrow to edit the title of a new set of Flash Cards.
- 3.) Drag words from your MyDictionary to the box here to add them to your Flash Cards.
- 4.) Click "Practice" to begin using your Flash Cards.

[Learn about the book...](#)

Merriam-Webster's

# One word or two words?

- LDOCE (especially LDOCE Free) likes to make two correct words by sticking a space in the middle of the misspelling
- a potentially useful strategy for spellchecking running text, but not so much for dictionary queries
- especially if the spellchecker does not care whether the resulting pair is a likely combination



# Longman

Dictionary of Contemporary English

[Search](#)[Widgets](#)[How to use](#)[About LDOCE](#)

## Do you mean:

of fen  
of-fen  
off en  
off-en  
offend  
offing  
oven  
offer  
often  
fen  
off  
offence  
offense  
effing  
omen  
open



# Longman

Dictionary of Contemporary English

Search

Widgets

How to use

About LDOCE

## Do you mean:

inter fir  
inter-fir  
interview  
interior  
interfere  
interfaith  
interwar  
interface  
inter  
anterior  
integer

[Pearson Longman](#)

[Longman Dictionaries](#)

[Copyright and legal](#)

[Privacy Policy](#)



# Longman

Dictionary of Contemporary English

Search

[Widgets](#)

[How to use](#)

[About LDOCE](#)

## Do you mean:

- good bay
- good-bay
- goodbye
- goody
- goodly
- goods
- Good
- good
- godly
- goodish

[Pearson Longman](#)

[Longman Dictionaries](#)

[Copyright and legal](#)

[Privacy Policy](#)

forward



Spell Check



Advanced Search



A-Z

CULTURE

FO

foal1 *noun*foal2 *verb*foam1 *noun*foam2 *verb*

foamy

fob1 *verb*fob2 *noun*

FOB

FOB airport

fob watch

focaccia

focal

focal length

focal point

fo'c'sle

focus1 *verb*focus2 *noun*

focus group

focus list

focussed

fodder

foe

FoE

foetal

Sorry, your search has returned no result. Try to change the criteria of your query, or try an alternative spelling:

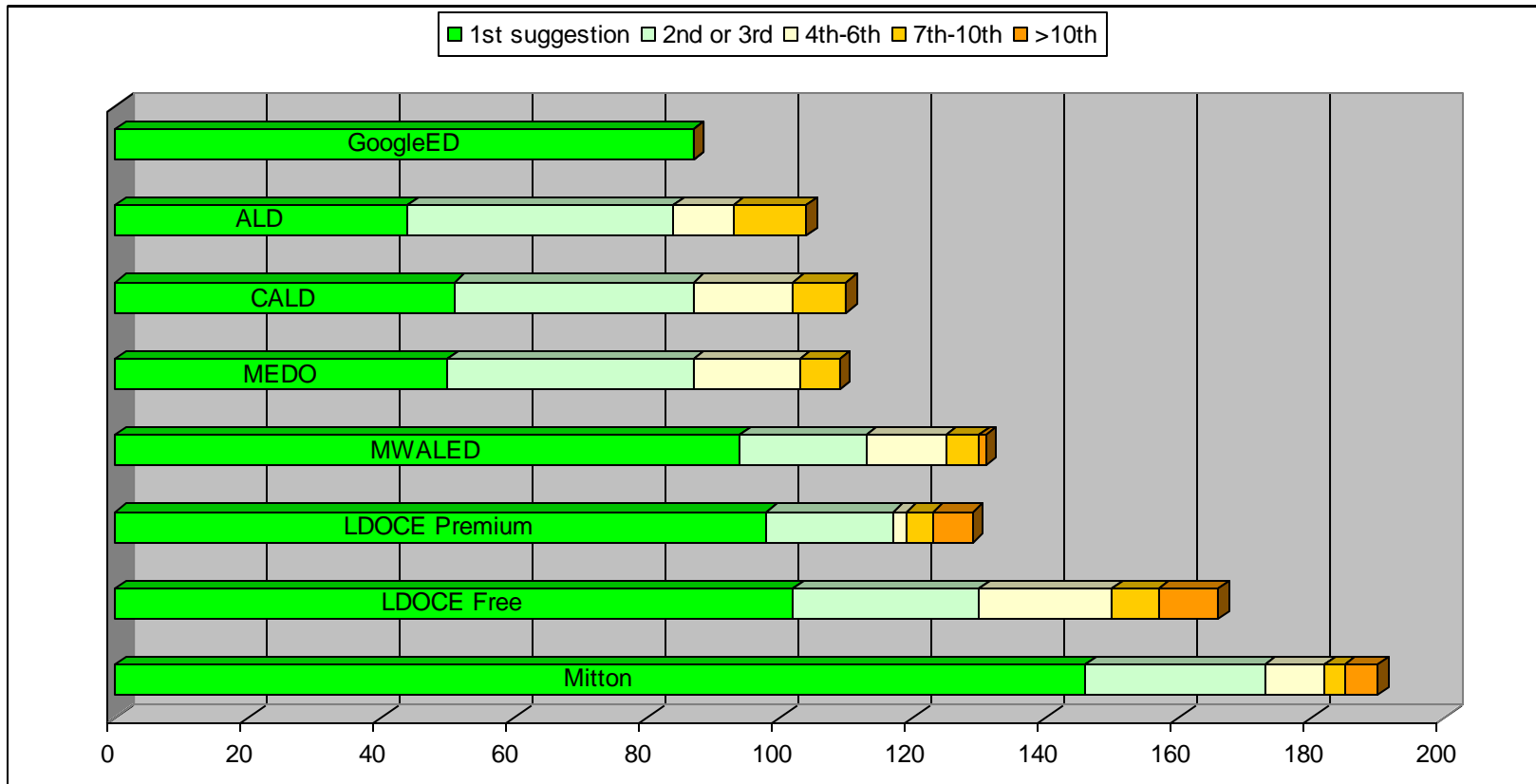
- fword
- ford
- Ford
- forward
- toward
- Coward
- coward
- Howard
- Forward
- ward
- Award
- award
- sward
- Board
- board
- hoard
- Ward
- Seward
- reward
- Howerd



# Can one do better than this?

- the dictionaries examined performed below expectation
- we decided to compare them with a context-free experimental spelling correction system (Mitton, 1996, 2009)
- at its heart is a dictionary primed with information about the quirks of English spelling
- it also makes use of word frequency in ordering its list of suggestions

# The dictionaries versus Mitton's spellchecker



# Ways to improve spelling correction in e-dictionaries

- misspelling patterns of non-native speakers differ (in quality and quantity) from those of native users of a language
  - customization by L1 (see Mitton & Okada 2007)
- however, we would not expect the influence of L1 to remain uniform across a wide range of L2 proficiency levels
  - individual (personalization)
- greater phonological awareness
  - many items where dictionaries had problems were phonologically motivated
  - as evidenced by L1-specific letter combinations (<sz>, <aj>, or <ej> for PL) to approximate English pronunciation
  - well handled by Mitton's spellchecker, though unaware of Polish-specific letter-to-sound correspondences
  - explicit inclusion of the most common such correspondences could improve performance significantly

# Dealing with real-word errors

- a proportion of misspellings turn a target word into another real word
- these are hard to catch except through contextual clues, but very rare words are potential misspellings
- in our corpus: \**wold* for *world*
- even though *wold* is a (very rare!) English word and most occurrences in texts *will* be misspellings of *world*, *would*, or *wild*
- still, risky to assume that *wold* as entered in the search box of a dictionary is a misspelling:
  - De Schryver et al. (2006) find that though the core vocab are most often looked up, rare words also get a fair share of lookups
- how should a dictionary respond?
- depends on the type of dictionary: OED online should not hesitate to serve up the entry for *wold*
- a learner's dictionary might not have it at all (no problem then), but if it did, a happy compromise might be to display the entry, but with a warning:  
'Did you perhaps mean \_\_\_\_\_?'

# First things first

- we have suggested at least three ways to improve success in correcting misspelled dictionary search terms
- but the more basic problems need to be addressed first
- such fundamental problems are numerous and grave, they affect the most authoritative of English monolingual learners' dictionaries
- we should be able to do much better than that

Thank you for listening!



UNIwersytet  
IM. ADAMA MICKIEWICZA  
W POZNANIU

Robert Lew  
Adam Mickiewicz University  
[rlew@amu.edu.pl](mailto:rlew@amu.edu.pl)

Roger Mitton  
Birkbeck College, University of London  
[roger@dcs.bbk.ac.uk](mailto:roger@dcs.bbk.ac.uk)



**Birkbeck**  
UNIVERSITY OF LONDON