

# WG1 hands-on workshop

## Learning materials through crowdsourcing: teachers, perspectives & scenarios

Gothenburg, 5-7 December 2018

# Crowdsourcing

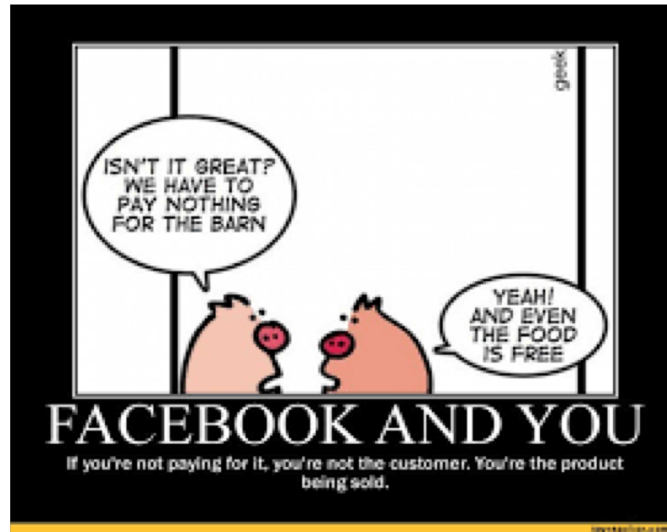
# The Rise of Crowdsourcing

Remember *outsourcing*? Sending jobs to India and China is so 2003. The new pool of *cheap labor*: everyday people using their spare cycles to create content, solve problems and even do corporate R & D.

*Jeff Howe, 2006, Wired magazine*

# About the (crowd)sourcing

Isn't it great?  
We have to  
pay nothing  
for the barn

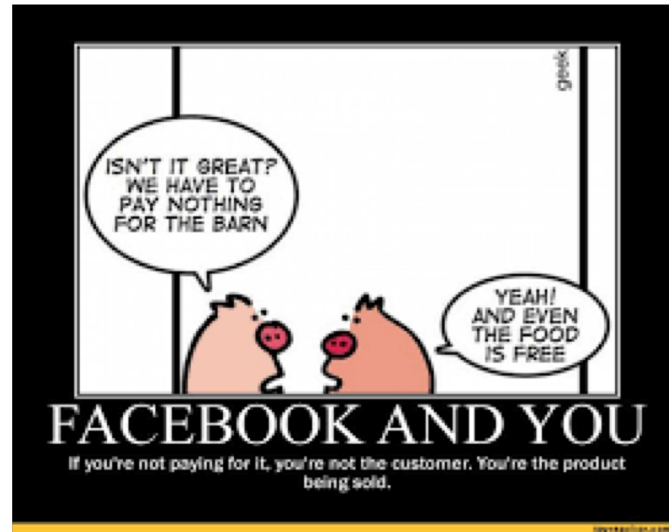
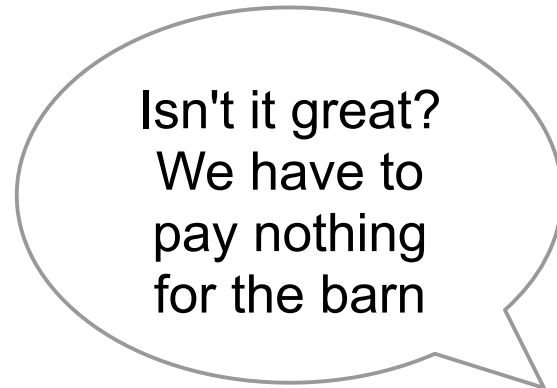


YEAH!  
And even  
the food  
is free

If you are not paying for it, you are not the customer.  
You are the product being sold.



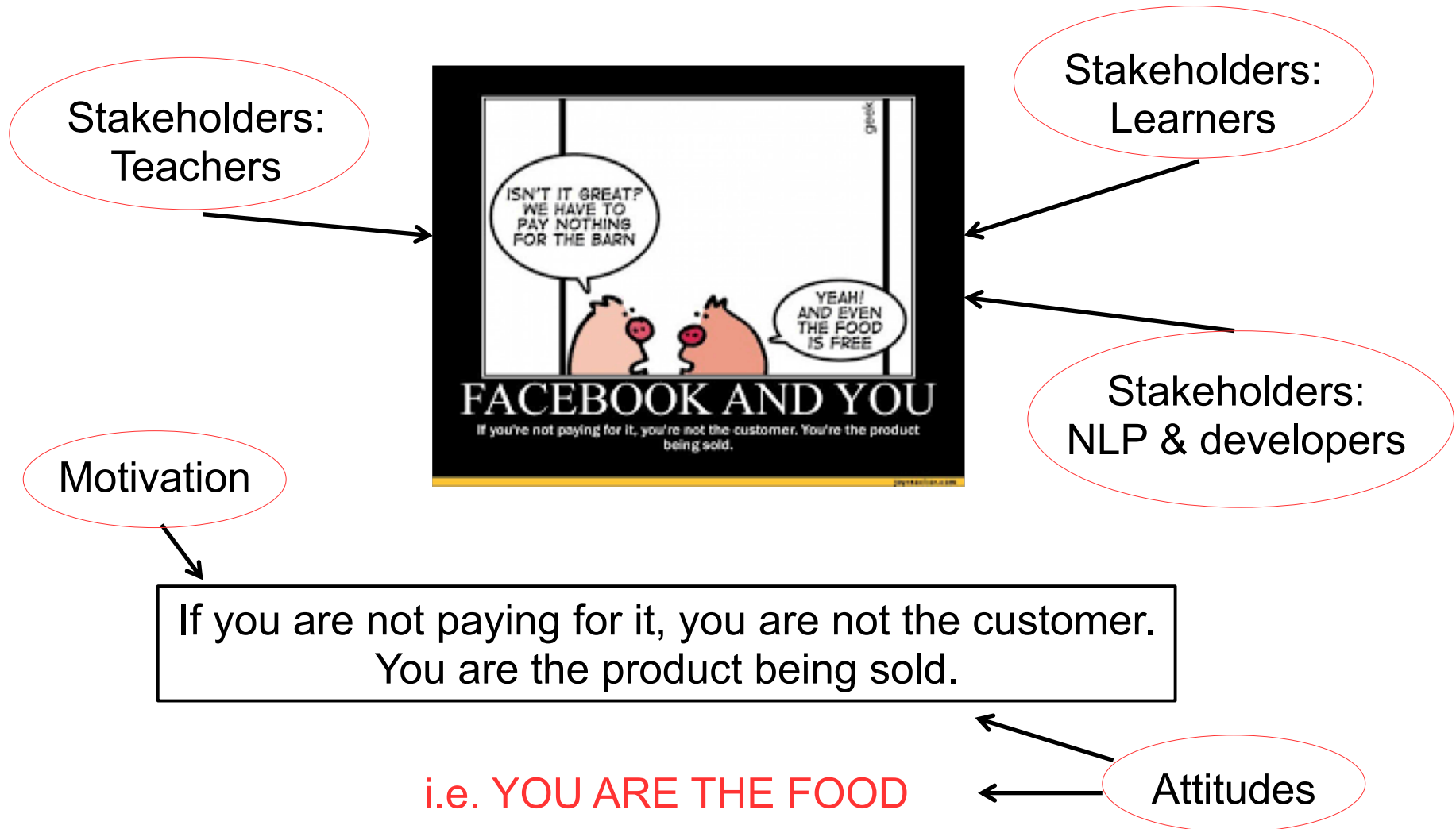
# About the (crowd)sourcing



If you are not paying for it, you are not the customer.  
You are the product being sold.

i.e. YOU ARE THE FOOD

# About the (crowd)sourcing



# About the crowdsourcing

*for language learning*

<https://www.youtube.com/watch?v=-Ht4qiDRZE8>



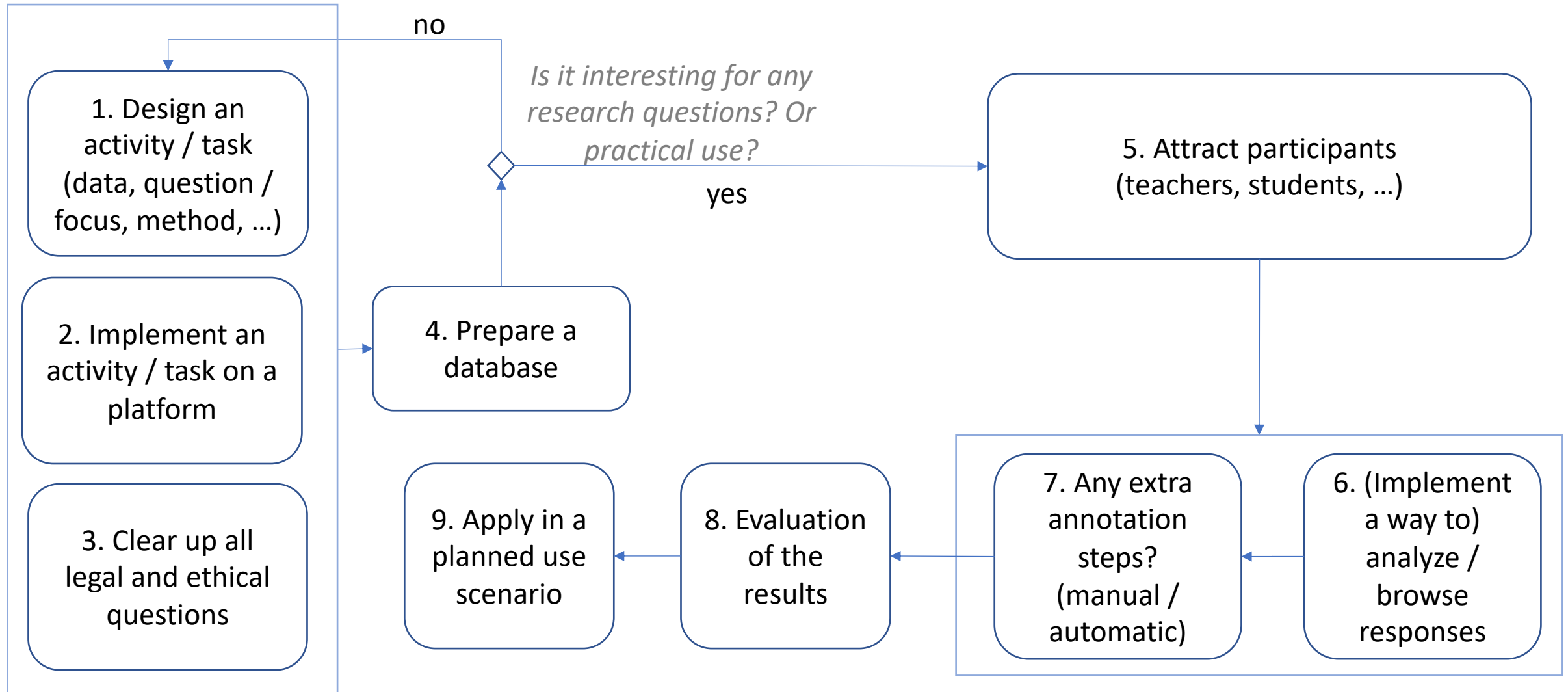
Luis von Ahn

2011

Massive-scale online collaboration  
DuoLingo

0:00-5:26, 8:30-13:00

# Crowdsourcing L2 resources/materials: steps



# Motivation for this hands-on meeting

- Crowdsourcing for language learning
  - Assessing the quality of text questions (*Chinkina & Meurers, 2017*)
  - Feedbook (*Ramon Ziaj, Bjoern Rudzewitz, Kordula De Kuthy, Florian Nuxoll & Detmar Meurers, 2018*)
  - DuoLingo (*Settles, Brust, Gustafson, Hagiwara & Madnani, 2018*)
  - ...
  - but --> not much to go, in fact, to validate crowdsourcing for LL
- We need to have a proof-of-concept that
  - crowdsourcing is valid for annotating/creating language learning data

# Experiment on Multi-Word Expressions

- Fall term 2018, enetCollect, Work Group 1
  - Elena Volodina, Iztok Kosem, Jaka Cibej, David Alfter
  - Linking Multi-Word Expressions (MWE) to L2 levels of proficiency



David  
Alfter



Jaka  
Čibej



Iztok  
Kosem



Elena  
Volodina

# MWE in L2 learning

- Granger: MWEs/formulaic language characterize learners of more advanced levels
- Define a scope for generation of exercises/tests and testing reading materials for appropriateness
- Where to get this data? And how?
  - Automatic annotation
  - Manual annotation
  - Crowdsourcing

# MWEs

- English Vocabulary Profiles
- Pybossa
- Best-worst scaling
- 5 votes for each task
- Clustering & ranking

British English

American English

Choose level:

- ☐ A1
- ☐ A1-A2
- ☐ A1-B1
- ☐ A1-B2
- ☐ A1-C1
- ☒ A1-C2
- ☐ A2 only
- ☐ B1 only
- ☐ B2 only
- ☐ C1 only
- ☐ C2 only

[Browse A-Z](#)

OR

Enter a word or phrase

ADVANCED SEARCH

Category

phrases

Part of speech

--Any--

Grammar

--Any--

Usage

--Any--

Topic

--Any--

Prefix

--Any--

Suffix

--Any--

Clear filters

☐ Hide culturally sensitive words

Search

Search results for A1-C2 (max. number of

1 2 3 ... 40 > Page 1 of 40

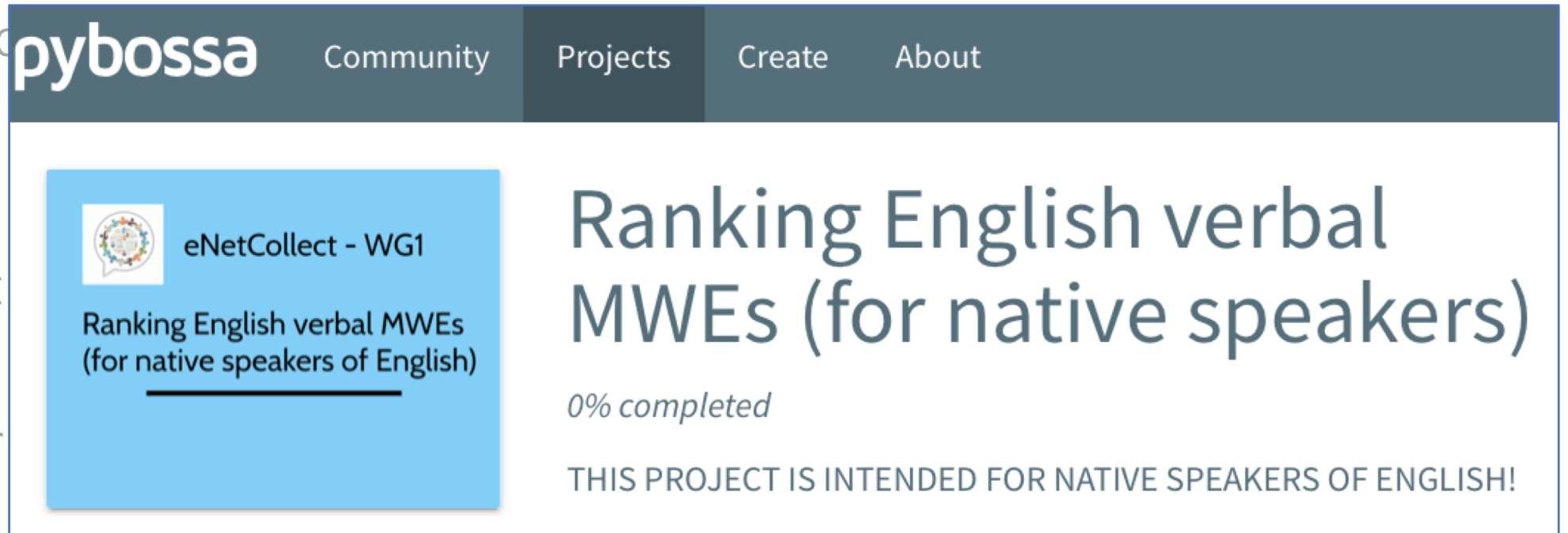
Core results:

- be able to do sth A2
- How/What about ...? A2
- How/What about ...? B1
- be (just) about to do sth B1
- above all B1
- Absolutely not. C2
- be absorbed in sth B2
- accept responsibility/blame B2
- by accident B1
- of your own accord C2
- in accordance with sth C1
- take account of sth B2
- take into account sth B2
- on account of sth B2
- by all accounts C1
- on no account; not on any account C2
- accustomed to sth/doing sth C1
- legal action C1
- out of action C1
- course of action C1
- in actual fact B2
- add insult to injury C2
- in addition (to) B1



# MWEs

- English Vocabulary
- Pybossa
- Best-worst
- 5 votes for
- Clustering & ranking



The screenshot shows the Pybossa website interface. The top navigation bar includes the 'pybossa' logo and links for 'Community', 'Projects', 'Create', and 'About'. The 'Projects' tab is selected. Below the navigation bar, a project card is displayed. The card has a light blue background and features a circular icon with a globe and the text 'eNetCollect - WG1'. The project title is 'Ranking English verbal MWEs (for native speakers of English)', which is underlined. Below the title, it says '0% completed'. At the bottom of the card, it states 'THIS PROJECT IS INTENDED FOR NATIVE SPEAKERS OF ENGLISH!'.

pybossa Community Projects Create About

eNetCollect - WG1

Ranking English verbal MWEs  
(for native speakers of English)

0% completed

THIS PROJECT IS INTENDED FOR NATIVE SPEAKERS OF ENGLISH!

# MWEs

- English Vocabulary I
- Pybossa
- Best-worst scaling
- 5 votes for each task
- Clustering & ranking

Easiest	Expression	Hardest
<input type="radio"/>	it goes without saying	<input type="radio"/>
<input type="radio"/>	kill time	<input type="radio"/>
<input type="radio"/>	lose your temper	<input type="radio"/>
<input type="radio"/>	beat about/around the bush	<input type="radio"/>

Save

Current task ID number: 688569 .

You have solved 0 task(s) out of a total of 326 . You are expected to solve 82 .

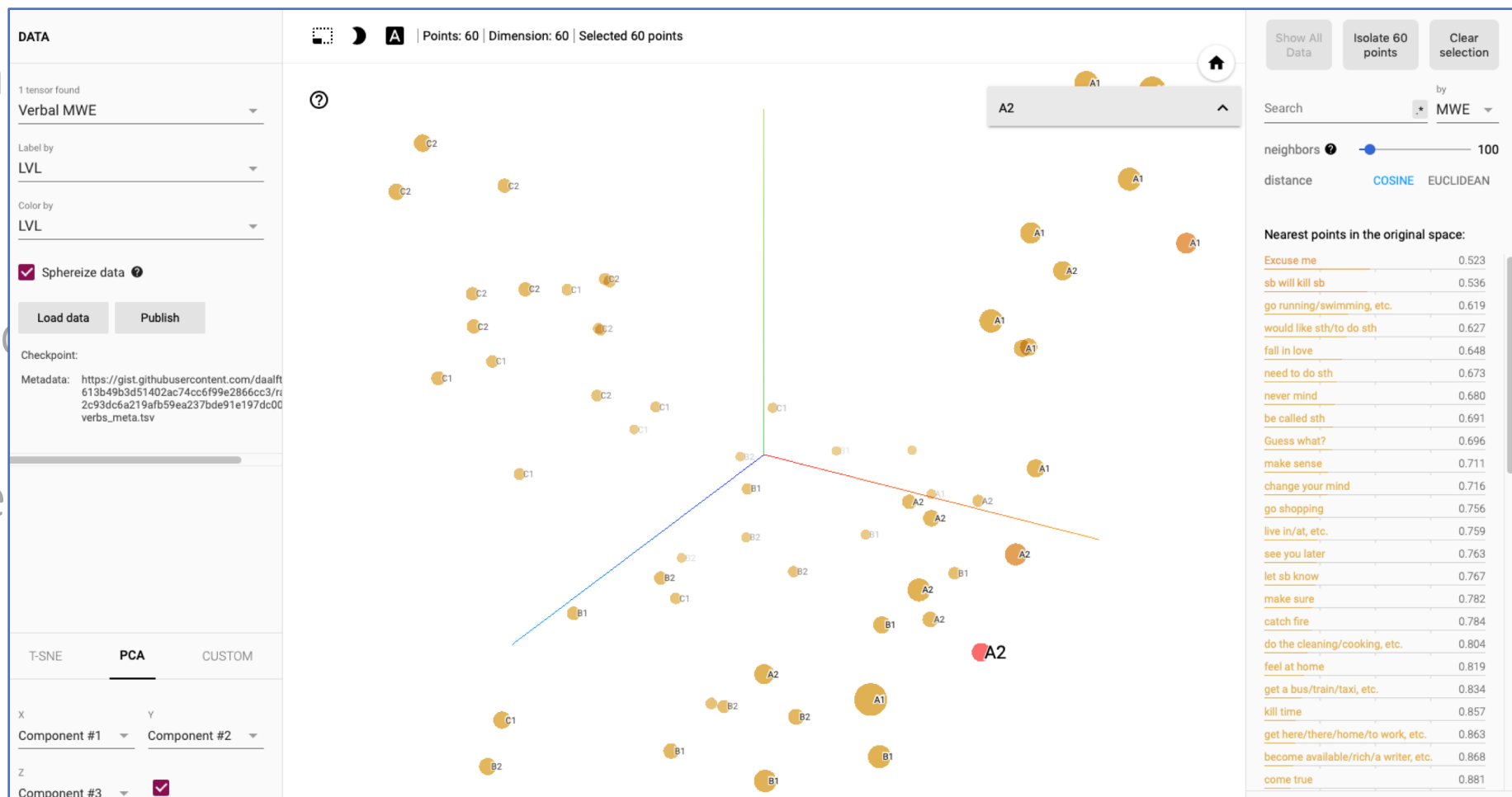
You can fill in [the feedback questionnaire](#) to describe how you made your decisions.

# MWEs

- English Vocabulary Profiles
- Pybossa
- Best-worst scaling
- 5 votes for each task. Voter profile: non-native English speaker
- Clustering  
& ranking

# MWEs

- English Vocabulary
- Pybossa
- Best-worst scaling
- 5 votes for each
- Clustering & ranking



# MWE clustering

MWE	Excuse me	Guess what?	be called sth	beat about/around the bush	become available/rich/a writer, etc.	break the ice	break the law	bring a lump
Excuse me	0.0	1.0	1.2857142857142858	1.7142857142857142	0.8571428571428571	1.0	1.0	1.4285714285714286
Guess what?	1.0	0.0	1.0	1.2857142857142858	0.5714285714285714	1.2857142857142858	0.8571428571428571	1.4285714285714286
be called sth	1.2857142857142858	1.0	0.0	1.1428571428571428	0.8571428571428571	0.5833333333333334	0.5714285714285714	1.1428571428571428
beat about/around the bush	1.7142857142857142	1.2857142857142858	1.1428571428571428	0.0	1.4285714285714286	1.3333333333333333	0.42857142857142855	1.1666666666666667
become available/rich/a writer, etc.	0.8571428571428571	0.5714285714285714	0.8571428571428571	1.4285714285714286	0.0	1.0	0.8571428571428571	1.0
break the ice	1.0	1.2857142857142858	0.5833333333333334	1.3333333333333333	1.0	0.0	0.8571428571428571	0.85
break the law	1.0	0.8571428571428571	0.5714285714285714	0.42857142857142855	0.8571428571428571	0.8571428571428571	0.0	1.0
bring a lump to your throat	1.4285714285714286	1.4285714285714286	1.1428571428571428	1.1666666666666667	1.0	0.85	1.0	0.0
burn the midnight oil	2.0	2.0	1.2	1.0	1.5	1.0526315789473684	1.4	0.7142857142857142
can't/couldn't help doing sth	1.7142857142857142	1.5714285714285714	1.0	1.0	1.1666666666666667	0.42857142857142855	0.5714285714285714	1.2857142857142858
catch fire	1.0	1.0	0.7142857142857143	1.5714285714285714	1.1428571428571428	0.7142857142857143	0.8	1.0
change your mind	1.1428571428571428	0.14285714285714285	1.0	1.0	0.6428571428571429	1.2857142857142858	0.4	1.1666666666666667
come true	1.0	1.0	1.0	1.4285714285714286	0.5	1.0714285714285714	0.7894736842105263	0.8333333333333333
crack a joke	1.8	1.1428571428571428	1.1428571428571428	0.5714285714285714	1.4285714285714286	1.0	0.42857142857142855	0.8571428571428571
cross your mind	1.0	1.0	1.5	1.0	0.7142857142857143	0.5714285714285714	0.7142857142857143	0.5714285714285714
do the cleaning/cooking, etc.	0.8571428571428571	1.0	1.0	0.5714285714285714	1.0	0.7142857142857143	1.3333333333333333	1.0
draw a conclusion	1.037037037037037	0.5714285714285714	1.0	0.7142857142857143	1.0	1.1428571428571428	0.42857142857142855	1.0
drive sb mad/crazy, etc.	1.0	0.5714285714285714	0.14285714285714285	0.7142857142857143	1.0	0.2857142857142857	1.0	1.0
face the music	1.8571428571428572	1.5714285714285714	1.0	0.6428571428571429	1.1428571428571428	1.0	0.7142857142857143	1.0
fall flat	1.2857142857142858	0.8571428571428571	0.42857142857142855	1.0	1.7142857142857142	0.7142857142857143	0.4	1.0
fall in love	0.8	1.0	0.8571428571428571	1.1428571428571428	0.8571428571428571	1.1428571428571428	1.0	1.4285714285714286
feel at home	0.2	1.1428571428571428	0.14285714285714285	0.7142857142857143	0.6666666666666666	0.8571428571428571	0.7142857142857143	1.3571428571428571
follow suit	1.6428571428571428	1.5714285714285714	1.5714285714285714	1.0	1.2857142857142858	0.5714285714285714	0.8571428571428571	1.0
get a bus/train/taxi, etc.	0.7142857142857143	0.5714285714285714	1.4285714285714286	0.5714285714285714	0.4	1.0	0.5714285714285714	2.0

# MWE ranking

MWE	CEFR	average_rank
burn the midnight oil	C2	2,777027027
grasp the nettle	C2	2,769736842
go against the grain	C2	2,75862069
throw in the towel	C2	2,68707483
beat about/around the bush	C1	2,647058824
follow suit	C2	2,644295302
keep sb on their toes	C2	2,620437956
nothing ventured, nothing gained	C2	2,6125
go from strength to strength	C1	2,584507042
bring a lump to your throat	C2	2,553333333
face the music	C1	2,549668874
fall flat	C1	2,503311258
get a grip (on yourself)	C1	2,493421053
hit the roof	C2	2,455172414
let off steam	C2	2,434482759
keep a low profile	C1	2,364238411
crack a joke	C1	2,353333333
get sth straight	C1	2,268115942
take it for granted	B2	2,253424658
lose your temper	B2	2,208333333
it goes without saying	B2	2,2
keep sb posted	C1	2,190789474
to cut a long story short	C1	2,15625
make up your mind or make your mind up	B1	2,151898734

# Data

- English Vocabulary Profile (EVP)
  - <http://vocabulary.englishprofile.org/staticfiles/about.html>
  - user: englishprofile, password: vocabulary

- MWEs are defined in terms of "phrases" in EVP

- Verbal MWEs: 10 per CEFR level = 60 items

- Adverbial MWES: 10 per CEFR level = 60 items

The screenshot displays the English Vocabulary Profile (EVP) search interface. On the left, there are tabs for 'British English' and 'American English'. Below these, a 'Choose level:' section lists CEFR levels from A1 to C2, with 'A1-C2' selected. There is also a 'Browse A-Z' link and an 'Enter a word or phrase' input field. An 'ADVANCED SEARCH' section contains dropdown menus for 'Category' (set to 'phrases'), 'Part of speech', 'Grammar', 'Usage', 'Topic', 'Prefix', and 'Suffix', all currently set to '--Any--'. A 'Clear filters' button is located below these menus. At the bottom left, there is a checkbox for 'Hide culturally sensitive words' and a red 'Search' button. On the right, the search results are displayed under the heading 'Search results for A1-C2 (max. number of)'. It shows a pagination bar with '1 2 3 ... 40' and 'Page 1 of 40'. Below this, a 'Core results:' section lists various phrases, each followed by a colored box indicating its CEFR level. Examples include 'be able to do sth' (A2), 'How/What about ...?' (A2), 'be (just) about to do sth' (B1), 'above all' (B1), 'Absolutely not.' (C2), 'be absorbed in sth' (B2), 'accept responsibility/blame' (B2), 'by accident' (B1), 'of your own accord' (C2), 'in accordance with sth' (C1), 'take account of sth' (B2), 'take into account sth' (B2), 'on account of sth' (B2), 'by all accounts' (C1), 'on no account; not on any account' (C2), 'accustomed to sth/doing sth' (C1), 'legal action' (C1), 'out of action' (C1), 'course of action' (C1), 'in actual fact' (B2), 'add insult to injury' (C2), and 'in addition (to)' (B1).

# CEFR – Common European Framework of Reference





# EVP data

- Labeled by lexicographers
- Based on Cambridge Learner Corpus (CLC) - a corpus of L2 English
- Using our system of ranking – how well can “a crowd” perform ?

The screenshot displays the Cambridge Learner Corpus (CLC) search interface. It features a search bar at the top with a dropdown menu for 'British English' and 'American English'. Below the search bar, there are options to 'Choose level' (A1, A1-A2, A1-B1, A1-B2, A1-C1, A1-C2) and 'Browse A-Z'. The search results are displayed on the right side, showing a list of phrases with their corresponding levels (A1, A2, B1, B2, C1, C2) indicated by colored boxes. The interface also includes an 'ADVANCED SEARCH' section with filters for Category, Part of speech, Grammar, Usage, Topic, Prefix, and Suffix. A 'Search' button is located at the bottom right.

British English American English

Choose level:

- ☐ A1
- ☐ A1-A2
- ☐ A1-B1
- ☐ A1-B2
- ☐ A1-C1
- ☒ A1-C2

Browse A-Z

OR

Enter a word or phrase

ADVANCED SEARCH

Category: phrases

Part of speech: --Any--

Grammar: --Any--

Usage: --Any--

Topic: --Any--

Prefix: --Any--

Suffix: --Any--

Clear filters

☐ Hide culturally sensitive words

Search

Search results for A1-C2 (max. number of)

1 2 3 ... 40 > Page 1 of 40

Core results:

- be able to do sth A2
- How/What about ...? A2
- How/What about ...? B1
- be (just) about to do sth B1
- above all B1
- Absolutely not. C2
- be absorbed in sth B2
- accept responsibility/blame B2
- by accident B1
- of your own accord C2
- in accordance with sth C1
- take account of sth B2
- take into account sth B2
- on account of sth B2
- by all accounts C1
- on no account; not on any account C2
- accustomed to sth/doing sth C1
- legal action C1
- out of action C1
- course of action C1
- in actual fact B2
- add insult to injury C2
- in addition (to) B1

# MWEs and crowdsourcing workshop

- <https://spraakbanken.gu.se/eng/wg1-dec18-gbg>
  - Analysis of the experiment results
  - Assessment – can crowdsourcing be used for language learning resources and materials?
  - Who should be the crowd?
  - How many are a crowd?
  - Is it possible to set up a multi-lingual experiment?
  - Etc-etc

Over - to set up and results

# Conclusions about MWE experiment

- Together during the WG1 workshop – analysis & conclusions
- Likely to see clashes between teachers, SLA and NLP researchers

# Reliability of tools & algorithms



# Evaluation needed

