

# MWEs and Crowdsourcing Outline and Results

---

Jaka Čibej, David Alfter, Iztok Kosem, Elena Volodina  
eNetCollect Annual Meeting  
14 March 2019, Lisbon



**enetCollect**  
COST Action CA16105  
<http://enetcollect.eurac.edu/>  
[enetcollect@gmail.si](mailto:enetcollect@gmail.si)



COST is supported by the EU  
Framework Programme  
Horizon 2020



**Språk-**  
**BANKEN**

# What and why?

- A **crowdsourcing experiment** to rank English multi-word expressions (MWE) according to their difficulty (L2 levels of proficiency)
  - *to burn the midnight oil*
  - *to be absorbed in something*
  - *to add insult to injury*
  - *to be able to do something*
- Manual annotation with CEFR-levels (A1 ... C2) is time-consuming and difficult
- Is there a simpler way to generate a ranked list?
  - Through crowdsourcing, perhaps?
  - Let's try and see.



# Who and when?

- STSM (Elena Volodina, Ljubljana, June 2018)
  - planning the experiment
- STSM (Jaka Čibej, Gothenburg, September 2018)
  - setting up the experiment
- Preparations and WG1 Workshop (Gothenburg, October–December 2018)
  - conducting the experiment and presenting the results



David Alfter



Jaka Čibej



Iztok Kosem



Elena Volodina



# English Vocabulary Profile

- 10 expressions per CEFR level
- 60 verbal MWEs
  - *to burn the midnight oil*
  - *it goes without saying*
- 60 adverbial MWEs
  - *Happy New Year!*
  - *by all accounts*



**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

**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

**ADVANCED SEARCH** ^

Category	phrases	<input type="button" value="v"/>
Part of speech	--Any--	<input type="button" value="v"/>
Grammar	--Any--	<input type="button" value="v"/>
Usage	--Any--	<input type="button" value="v"/>
Topic	--Any--	<input type="button" value="v"/>
Prefix	--Any--	<input type="button" value="v"/>
Suffix	--Any--	<input type="button" value="v"/>

Hide culturally sensitive words

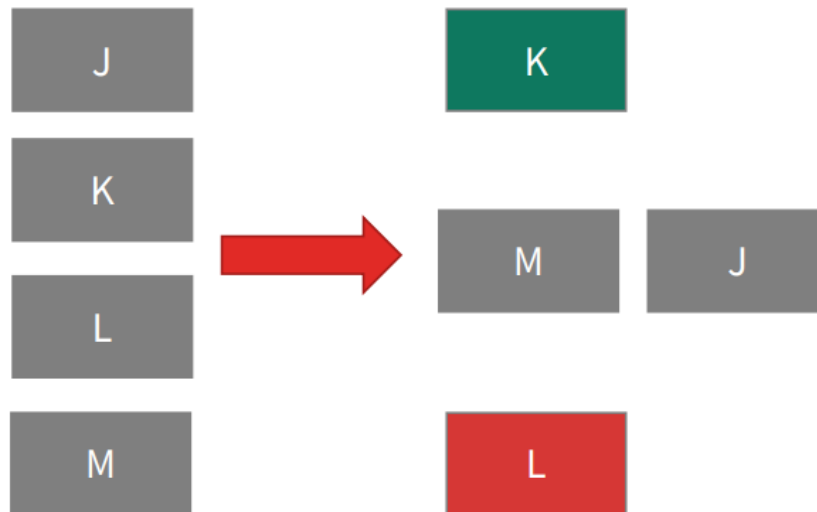
# How to rank?

- Ranking the entire list?
  - Task can't be divided between multiple participants.
- Ranking a subset of tasks?
  - Combinations might affect results.
  - Still not very user-friendly.
  - Difficult to merge?
  - Which combinations?



# Best-Worst Scaling

- Ranking method
- Choosing the best and worst unit in a combination of (ideally) 3–4 candidates
- Example:



- 6 possible binary relations between the 4 elements
  - $J \sim K, J \sim L, J \sim M, K \sim L, K \sim M, L \sim M$
- **BWS with 4 elements**
  - $K = 3, M = 2, J = 2, L = 1$
  - $J < K, J > L, J \sim M, K > L, K > M, L < M$
  - 5 out of 6 relations (83 %)
  - (at least) 2 clicks
- **Ranking all 4 elements:**
  - 6 out of 6 relations (100 %)
  - (at least) 4 clicks
  - twice the workload!



# Tasks

- 60 expressions per project
  - 487,635 combinations (for combinations of 4 units)
  - 1,770 binary relations
- 326 tasks per project (to include all binary relations between the expressions)
- 77% are non-repetitive.
- 23% are partially repetitive (as little as
- possible).



# Predictions

- **IF:**
  - Number of crowdsourcers: 20
  - Average response time: 30 seconds
  - Responses per task: 5
- **THEN:**
  - Tasks per crowdsourcer: 82 (per project)
  - Time per crowdsourcer: 0.68 hours, which equals 40.75 minutes





# PyBossa Interface

Easiest	Expression	Hardest
<input type="radio"/>	a lot	<input type="radio"/>
<input type="radio"/>	once upon a time	<input type="radio"/>
<input type="radio"/>	as it happens	<input type="radio"/>
<input type="radio"/>	deadly dull/serious, etc.	<input type="radio"/>

Save

*as it happens*

**Meaning:** something that you say in order to introduce a surprising fact

**Example:** As it happens, her birthday is the day after mine.

---

Current task ID number: **689222** .

You have solved **1** 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.



# Guidelines

- Decide which expression is the most difficult/easiest for a language learner **to produce**.
- In case of a tie, choose one.
- Do not overthink the decision.
- Try not to spend more than 30 seconds per task.
- **(No mention of the English Vocabulary Profile OR CEFR-levels!)**
  - crowdsourcers only relied on their intuition
- 26 participants, mostly linguists and NLP experts
  - 24 non-native speakers of English, 2 native speakers



# Merging the Results

- **Method 1: Linear scale using average ranks**
  - a more brute-force approach
  - take all annotations for a specific expression (regardless of the expressions it appears with)
  - average the sum to get the expression's average rank
  - the premise: harder/easier expressions should more frequently be annotated as more difficult (rank 3) or easier (rank 1)



# Linear Scale

MWE	CEFR	average_rank
burn the midnight oil	C2	2,781818182
go against the grain	C2	2,771428571
grasp the nettle	C2	2,745454545
follow suit	C2	2,681818182
throw in the towel	C2	2,657142857
beat about/around the bush	C1	2,636363636
keep sb on their toes	C2	2,63
nothing ventured, nothing gained	C2	2,608695652
go from strength to strength	C1	2,59047619
face the music	C1	2,545454545
bring a lump to your throat	C2	2,536363636

have a rest/shower/walk, etc.	A2	1,438095238
get here/there/home/to work, etc.	A1	1,409090909
do the cleaning/cooking, etc.	A1	1,4
get a bus/train/taxi, etc.	A1	1,381818182
go running/swimming, etc.	A2	1,32
see you later	A1	1,260869565
live in/at, etc.	A1	1,217391304
Excuse me	A1	1,209090909
go shopping	A1	1,18

MWE	CEFR	average_rank
burn the midnight oil	C2	2,781818182
go against the grain	C2	2,771428571
grasp the nettle	C2	2,745454545
follow suit	C2	2,681818182
throw in the towel	C2	2,657142857
beat about/around the bush	C1	2,636363636
keep sb on their toes	C2	2,63
nothing ventured, nothing gained	C2	2,608695652
go from strength to strength	C1	2,59047619
face the music	C1	2,545454545
bring a lump to your throat	C2	2,536363636
fall flat	C1	2,527272727
get a grip (on yourself)	C1	2,481818182
let off steam	C2	2,466666667
hit the roof	C2	2,428571429
crack a joke	C1	2,381818182
keep a low profile	C1	2,372727273
take it for granted	B2	2,247619048
lose your temper	B2	2,241666667
get sth straight	C1	2,228571429
it goes without saying	B2	2,19047619
keep sb posted	C1	2,172727273
to cut a long story short	C1	2,139130435
make up your mind	B1	2,130434783
cross your mind	B2	2,127272727
draw a conclusion	B2	2,09552174
break the ice	B2	2,091666667
can't/couldn't help do sth	B1	2,07826087
get rid of sth	B1	2,057142857
break the law	B2	2,019047619
catch fire	B1	2
drive sb mad/crazy	B2	1,980952381
kill time	B2	1,963636364
keep in touch	B1	1,904761905
help yourself (to sth)	B1	1,904761905
fall in love	B1	1,754545455
change your mind	B1	1,752380952
be called sth	A1	1,75
make sense	B2	1,747826087
let sb know	B2	1,747826087
come true	B1	1,745454545
give sb a calling	B2	1,73
never mind	B2	1,7
feel at home	B1	1,660809525
make sure	B1	1,652173913
become available/accessible	B2	1,647619048
give a party	B2	1,6
would like sth/to do sth	A1	1,6
sb will kill sb	B2	1,580952381
need to do sth	A1	1,539130435
Guess what?	B2	1,514285714
have a rest/shower/walk	B2	1,438095238
get here/there/home	A1	1,409090909
do the cleaning/cooking	A1	1,4
get a bus/train/taxi	A1	1,381818182
go running/swimming	B2	1,32
see you later	A1	1,260809525
live in/at, etc.	A1	1,217391304
Excuse me	A1	1,209090909
go shopping	A1	1,18



# Merging the Results

- **Method 2:** Clustering and multi-dimensional visualization using vector embeddings
  - 60x60 matrix of average distances between expressions

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 to your th
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	0.7142857142857143
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	0.8333333333333334
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	1.3571428571428572
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



# Clustering

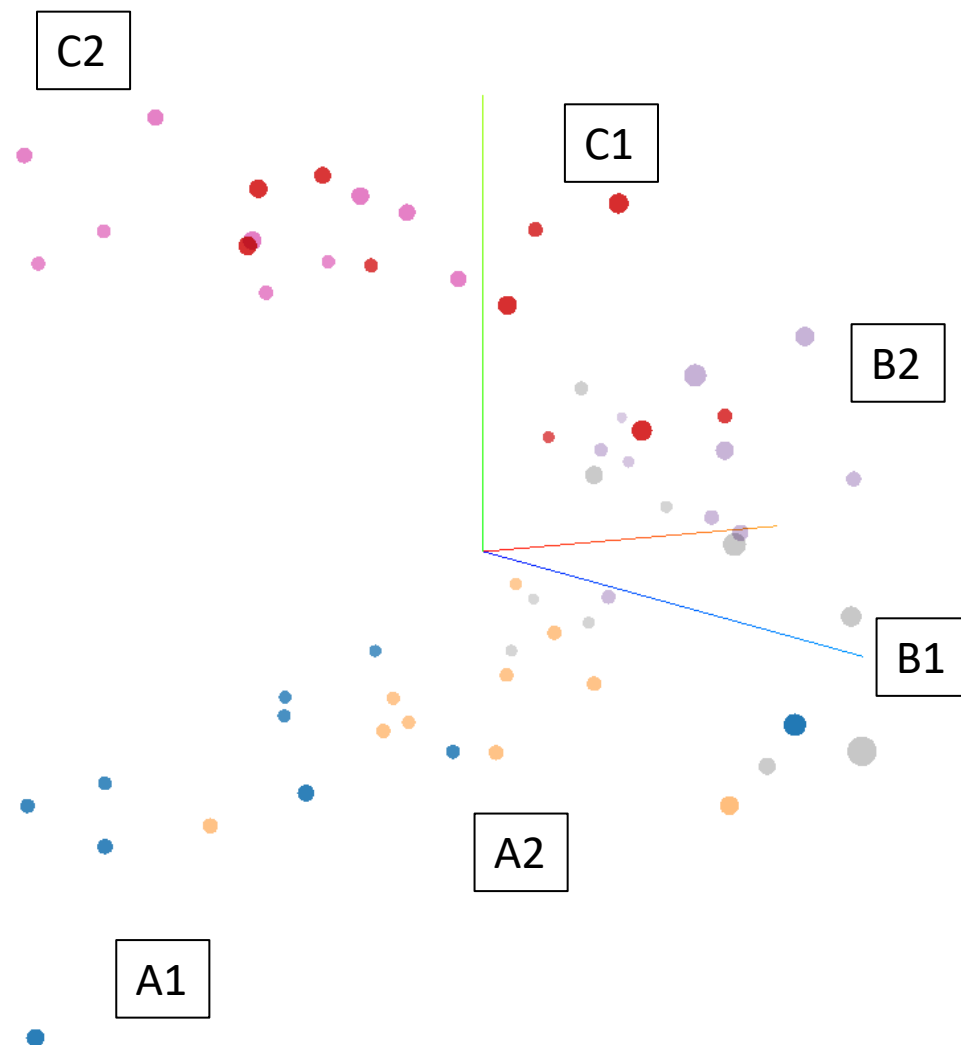
- **Method 1: Linear scale using average ranks**
  - k-means with  $n = 6$
  - Adverbial MWEs: 41.7% accuracy
  - Verbal MWEs: 50.0% accuracy
  - **most misclassifications between neighboring levels!**
    - e.g. A1 ~ A2, C1 ~ C2, but no A1 ~ C2

Assigned ↓ True ->	A1	A2	B1	B2	C1	C2
A1	6	1	2	1	0	0
A2	2	5	3	0	0	0
B1	0	1	3	5	1	0
B2	0	1	1	5	3	0
C1	0	0	0	3	2	5
C2	0	0	0	1	0	9

**Verbal  
MWEs**

# Visualization

- **Tensorflow embedding projector**
- <https://tinyurl.com/enetCollectVerbalMWE>
- <https://tinyurl.com/enetCollectAdverbialMWE>



# Conclusion

- **Crowdsourcing for generating language learning resources?**
  - possible
  - minimal crowdsourcer training
  - results comparable to expert annotations
- **Future work**
  - difference between native and non-native crowdsourcers?
  - what about language learners?
  - similar experiments for other languages





# Thank you.

---

Jaka Čibej  
jaka.cibej@cjvt.si