MWE for Essay Scoring English as a Foreign Language

Rodrigo Wilkens*, Daiane Seibert†, Xiaoou Wang*, Thomas François*
*CENTAL, IL&C, UCLouvain, †KU Leuven

Introduction

- Multiword expressions (MWE) are associated with proficiency.
- MWEs are usually neglected in works concerning automated scoring of language learners.

Objectives:

1. Verify MWE impact on learners' proficiency.
2. Compare MWE-based features with classic linguistic ones.

MWE features

MWE$_{cnt}$ Occurrence of MWE, based on a list of more than 62 thousand MWEs (Muraki et al., 2022).

MWE$_{conc}$ The concreteness ratings of the MWE perceived by native speakers.
Snow shower v. *Skeleton in the closet*

Corpus

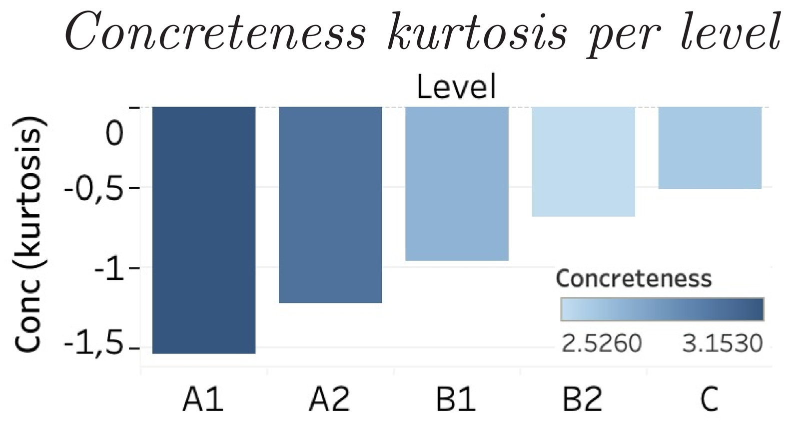
- EFCAMDAT, 10 most represented nationalities.
- Unification of C1 and C2 levels into C level since the quantity of texts decreases considerably in higher levels.
- Truncation of the larger nationalities to reduce bias.

Nationality	Texts	%Texts
Brazil	2469	22.99
Germany	2469	22.99
Italy	1238	11.53
Russia	1195	11.13
France	818	7.62
Mexico	762	7.09
China	555	5.17
Saudi Arabia	468	4.36
Japan	420	3.91
Taiwan	347	3.23

Profiling MWE usage

- Beginners produce more concrete MWEs, more abstract ones are found in higher levels.
- Kurtosis is highly relevant to MWE's concreteness.

Concreteness kurtosis per level



Relation between MWE and level

- The concreteness shows better correlation than the occurrence.
- Higher correlated feature does not imply that the corresponding feature family is highly correlated.

Correlation of different features and families of features

Family	best score	correlation	
		best	family
DVR	STTR (all surface tks)	0.81	0.42 (0.25)
DEV	depth	0.70	0.48 (0.17)
DEP	mark	0.62	0.29 (0.20)
POS	punct	0.59	0.27 (0.14)
LEN	word per sent.	0.58	0.50 (0.07)
NRM	AOA	0.58	0.43 (0.13)
FRQ	content words subtlex	0.57	0.28 (0.18)
PRH	SBAR	0.54	0.20 (0.16)
TNS	use past	0.51	0.18 (0.12)
MOR	finite verb	0.47	0.26 (0.14)
NGH	phonologic dist	0.47	0.20 (0.17)
SOP	verbs	0.46	0.32 (0.14)
MWE	MWE$_{conc}$	0.36	-
COH	PPMI (lemma)	0.29	0.14 (0.09)
GRD	C1	0.24	0.21 (0.04)
MWE	MWE$_{cnt}$	0.21	-

Other features

LEN Length-based (4)
GRD Graded resource (6)
FRQ Frequency features (18)
NGH Orthographic neighbor (8)
NRM Lexical Norms (8)
SOP Lexical sophistication (6)
POS Part-of-speech tags (17)
MOR Morphology features (56)
DEP Dependency relations (37)
TNS Verb tense (19)
PRH Phrase usage (34)
DEV Language development (5)
DVR Lexical diversity (112)
COH Coherence features (15)

MWE usage statistics

Correlation of MWE features aggregators

MWE	Kurt	Avg	Q3	Median	Q1	Min
CONC	0.40	0.36	-0.29	-0.35	-0.37	-0.50
CNT	-	0.21	-0.02	-	-	-

Level Classification

- Parser (MOR, POS, DEP, PRH and TNS)
- NRM$_{all}$, is the lexical norms-based features (NRM and MWE$_{conc}$)
- All-MWE, all features excluding MWE ones

Machine learning models using different features

Feature set	RandForest		SLogistic	
	ACC	F1	MAE	RMSE
LEN	0.553	0.553	0.897	1.364
FRQ	0.682	0.682	0.739	1.200
GRD	0.490	0.490	1.014	1.487
NGH	0.561	0.560	1.053	1.520
NRM	0.624	0.624	0.744	1.158
SOP	0.498	0.498	0.869	1.294
DVR	0.745	0.745	0.410	**0.789**
DEP	0.736	0.736	0.630	1.065
PRH	0.645	0.645	0.941	1.406
DEV	0.726	0.726	0.694	1.075
POS	0.745	0.744	0.772	1.235
MOR	**0.775**	**0.775**	**0.682**	1.126
TNS	0.565	0.559	0.731	1.161
COH	0.519	0.519	1.170	1.628
MWE	0.428	0.425	1.455	1.916
MWE$_{cnt}$	0.454	0.447	1.660	2.121
MWE$_{conc}$	0.418	0.413	1.499	1.946
Parser	0.835	0.835	0.425	0.857
NRM$_{all}$	0.640	0.640	0.734	1.153
All	0.843	0.843	**0.535**	0.697
All-MWE	**0.844**	**0.844**	0.534	**0.699**

Conclusions

1. MWE statistical measures show promising results.
2. We profiled MWE concreteness usage across CEFR's levels.
3. We compared MWE and classic scores.

UCLouvain

CENTAL centre de traitement automatique du langage