

*SemEval 2022 Task 2:
Multilingual Idiomaticity Detection and
Sentence Embedding*

Dr Harish Tayyar Madabushi



The
University
Of
Sheffield.

Motivation for a Shared Task on Idioms

State of the art NLP systems have achieved extraordinary success across several tasks such as sentences classification (Zhang et al., 2019) and reading comprehension (Raffel et al., 2019)

However, they cannot effectively represent idiomatic multiword expressions (MWEs) (Yu and Ettinger, 2020; Garcia et al., 2021).

Addressing this is non-trivial due to their heavy reliance on compositionality at both the word and sub-word levels.

Proposed Tasks

Subtask A: Identification of Idiomaticity

Given a sentence with a potentially idiomatic MWE, classify the use of the MWE in the sentence as either: a) Idiomatic or b) Non-idiomatic

Subtask B: Idiom Representation

The effective representation of potentially idiomatic expressions so they can be used in downstream tasks

Subtask A: Data and Task

MWE	Target Sentence	Previous Sentence	Next Sentence	Label	Idiomatic?	Paraphrase
gold mine	This means that search data is a <i>gold mine</i> for marketing strategy. (marketingweek.com)	The data that those searches generate builds...	It reveals which types of product...	Idiomatic 1	Yes	source of fortune
gold mine	The hashtag “Qixia <i>gold mine</i> incident” has been viewed many million of times on the social media site Weibo. (wsws.org)	The rescue operation took place...	A week after the explosion...	Literal	No	mine
gold mine	The <i>Gold Mine</i> ’s plain frontage & sparse, white-walled dining room suggest that it’s a quick-fix refuelling stop rather than a place to linger. (squaremeal.co.uk)	SquareMeal Review of Gold Mine	The menu touts a bewildering array of dishes...	Proper Noun	No	Proper Noun

Subtask B: Data and Task

Sentence (E)	Correct Replacement ($E_{MWE \rightarrow c}$)	Wrong Replacement ($E_{MWE \rightarrow i}$)	Expected
<p>When removing a <i>big fish</i> from a net, it should be held in a manner that supports the girth. (newsdakota.com)</p> <p>To pay attention only to new housing and houses I think skews the <i>big picture</i>. (streets.mn)</p>	<p>When removing a <i>fish</i> from a net, it should be held in a manner that supports the girth.</p> <p>To pay attention only to new housing and houses I think skews the <i>whole situation</i>.</p>	<p>When removing a <i>important person</i> from a net, it should be held in a manner that supports the girth.</p> <p>To pay attention only to new housing and houses I think skews the <i>large image</i>.</p>	<p>$sim(E, E_{\rightarrow c}) = 1$ $sim(E, E_{\rightarrow i}) = sim(E_{\rightarrow c}, E_{\rightarrow i})$</p> <p>$sim(E, E_{\rightarrow c}) = 1$ $sim(E, E_{\rightarrow i}) = sim(E_{\rightarrow c}, E_{\rightarrow i})$</p>

Baseline, Languages and Setup

Models struggle in the zero shot setting, which is the most interesting.

Representations of idiomatic expression without annotated data are poor

To this end we propose:

English

Zero shot, One shot

Portugese

Zero shot, One shot

Galician

Zero shot, One shot (Transfer learning only)

Task Organisers

Aline Villavicencio

University of Sheffield, UK

Marco Idiart

Federal University of Rio Grande do Sul, Brazil

Carolina Scarton

University of Sheffield, UK

Marcos Garcia

Universidade de Santiago de Compostela, Spain

Harish Tayyar Madabushi

University of Sheffield, UK

Visit the Task Website (& join the Google Group)

<https://sites.google.com/view/semEval2022task2-idiomaticity>

