Procedure for identifying metaphorical scenes (PIMS): the case of spatial and abstract relations. - Study 1: Reliability testing for corpus linguistic data for the preposition into

Table of contents

1. Description of methods used for collection/generation of data:

We applied PIMS to the identification of metaphorical constructions that include prepositions in two studies. The experimental designs of the two studies are described in:

Johansson Falck, M. and L. Okonski (2022). Procedure for Identifying Metaphorical Scenes (PIMS): The Case of Spatial and Abstract Relations." *Metaphor and Symbol, 38*(1), 1-22. https://doi.org/10.1080/10926488.2022.2062243

This post is related to the data discussed in Study 1: Reliability testing for corpus linguistic data for the preposition *into*

Study 1: Reliability testing for corpus linguistic data for the preposition into

In a first study, we used PIMS to identify metaphorical 'into relations' that are evoked by sentences that include the preposition *into*.

2. Methods for processing the data:

The data were excepted from (Davies, 2008). First, we searched for the 100 most frequent nouns that collocate with the preposition *into* in a one-word window to the right of the preposition). Second, we retrieved 100 instances of each of the 100 *into* + noun collocations from the corpus platform[1]. Third, we identified metaphorical *into* relations by means of applications of PIMS individually.

Datafile, Study 1:

- Study_1_application_PIMS_into_coll.xlsx
  The file "Study_1_application_PIMS_into_coll_2022_10_05.xlsx" includes two versions of our codings of the instances of the *into* collocations. In version 1 of our analysis (sheet into coll heaven metaphor), the concept of heaven was understood as an abstract concept. In the other version (sheet into coll heaven place) 'heaven' was understood as place that people or their souls can literally travel to after death.

3. Instrument- or software-specific information needed to interpret the data:

---

[1] Searches for 100 instances of collocations in COCA occasionally generate fewer than 100 instances of each collocation. Moreover, when the data was excerpted (201220), there were fewer than 100 instances of the 56th-100th most frequent *into* + noun collocations in COCA.

Fourth, we compared our coding of metaphorical vs. non-metaphorical *into* relations by means of Fleiss kappa using the irr-package (Gamer, Lemon, & Singh, 2012) and Gwet's Kappas using the irrCAC package (Gwet, 2019).

We tested our interrater reliability to provide information about the reliability of the method based on the raw data in the "Study_1_applications_PIMS_into_coll_2022_10_05.xlsx".

The Fleiss kappa measure was calculated using the template for the R code available from the Open Science Framework of the volume Metaphor Identification in Multiple Languages: MIPVU Around the World (Nacey, Dorst, Krennmayr, & Reijnierse, 2019). The code uses the 'boot' function, a bootstrap for the 'irr' package in R and is available at https://osf.io/vw46k/files/osfstorage.

Gwet's Kappa was calculated using the agreement coefficient from raw ratings available from Kilem L. Gwet's home page for Calculating Chance-corrected Agreement Coefficients (CAC) https://cran.r-project.org/web/packages/irrCAC/vignettes/overview.html . The code uses the irrCAC package in R. Version 1.0.

4. People involved with sample collection, processing, analysis and/or submission:

The data were excerpted from Davies, M. COCA. Corpus of Contemporary American English. on May 8 2019 by Marlene Johansson Falck. The data were analyzed by Marlene Johansson Falck and Lacey Okonski.

5. DATA-SPECIFIC INFORMATION FOR:

[Study_1_application_PIMS_into_coll_2022_10_05.xlsx]

1. Number of variables: 3 variables were coded by two researchers (M and L)

2. Number of cases/rows: 7540

3. Variable List:

METM = metaphorical into relation, coded by Researcher M
NONMM = nonmetaphorical into relation, coded by Researcher M
AMBM = ambiguous into relation, coded by Researcher M

METL = metaphorical into relation, coded by Researcher L
NotML= nonmetaphorical into relation, coded by Researcher L
AMBL = ambiguous into relation, coded by Researcher L

4. Specialized formats or other abbreviations used:

ID = ID of collocation
COLL = collocation (i.e., noun collocating with *into*)
into coll heaven metaphor = analysis version 1 (in which *heaven* was interpreted as a metaphorical concept)
into coll heaven place = analysis version 2 (in which *heaven* was interpreted as a place)

References:

Davies, M. (2008). COCA. Corpus of Contemporary American English. In.

Gamer, M., Lemon, J., & Singh, I. F. P. (2012). irr: Various Coefficients of Inter-rater Reliability and Agreement, R package version 0.84. http://CRAN.R-project.org/package=irr <p class="EndNoteBibliography" style="margin-left:35.45pt;text-align:justify;

text-justify:inter-ideograph;text-indent:-35.45pt;line-height:150%">. In.

Gwet, K. L. (2019). irrCAC: Calculating Chance-corrected Agreement Coefficients (CAC). https://cran.r-project.org/web/packages/irrCAC/vignettes/overview.html.

Nacey, S., Dorst, A. G., Krennmayr, T., & Reijnierse, W. G. (2019). *Metaphor Identification in Multiple Languages : MIPVU Around the World*. Amsterdam/Philadelphia: John Benjamins Publishing Company.