EVALUATION OF RUSSIAN NOUN WORD EMBEDDINGS FOR CASE AND NUMBER

Anastasia Yablokova
anyab100@hhu.de
Computerlinguistik BA, HHU

Agenda

- Word Embedding
- Previous Research
- Project Idea
- Results

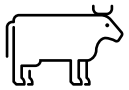
Word Embedding

- A Neural Probabilistic Language Model (Bengio et al., 2003)
- Word2vec (Mikolov et al., 2013)
- Glove (Pennington et al, 2014)
- FastText (Joulin et al, 2016)

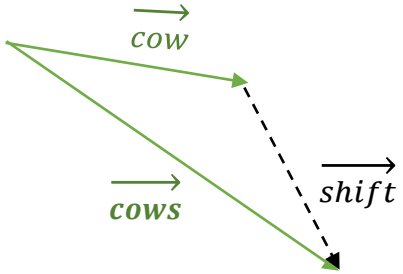
PREVIOUS RESEARCH

Semantic properties of English nominal pluralization: Insights from word embeddings

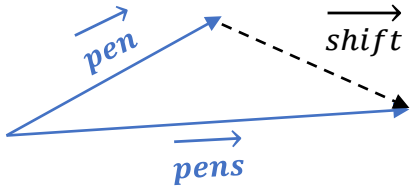
(Elnaz Shafaei-Bajestan, Masoumeh Moradipour-Tari, Peter Uhrig and R. Harald Baayen, 2022)



- cow = $[\![cow]\!]$
- cows = $[\![cow]\!] + [\![plural]\!]$



- pen = $[\![pen]\!]$
- pens = $[\![pen]\!] + [\![plural]\!]$

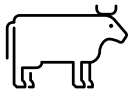


average shift vector: the difference vector between the average of plural vectors and the average of singular vector

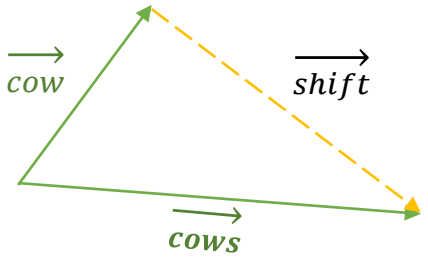
$$[\![plural]\!]_{\text{cows-cow}} \approx [\![plural]\!]_{\text{pens-pen}}$$

Semantic properties of English nominal pluralization: Insights from word embeddings

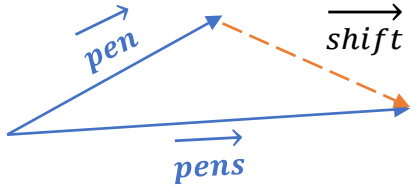
(Elnaz Shafaei-Bajestan, Masoumeh Moradipour-Tari, Peter Uhrig and R. Harald Baayen, 2022)



- cow = $[\![cow]\!]$
- cows = $[\![cow]\!] + [\![plural_{animal}]\!]$



- pen = $[\![pen]\!]$
- pens = $[\![pen]\!] + [\![plural_{item}]\!]$



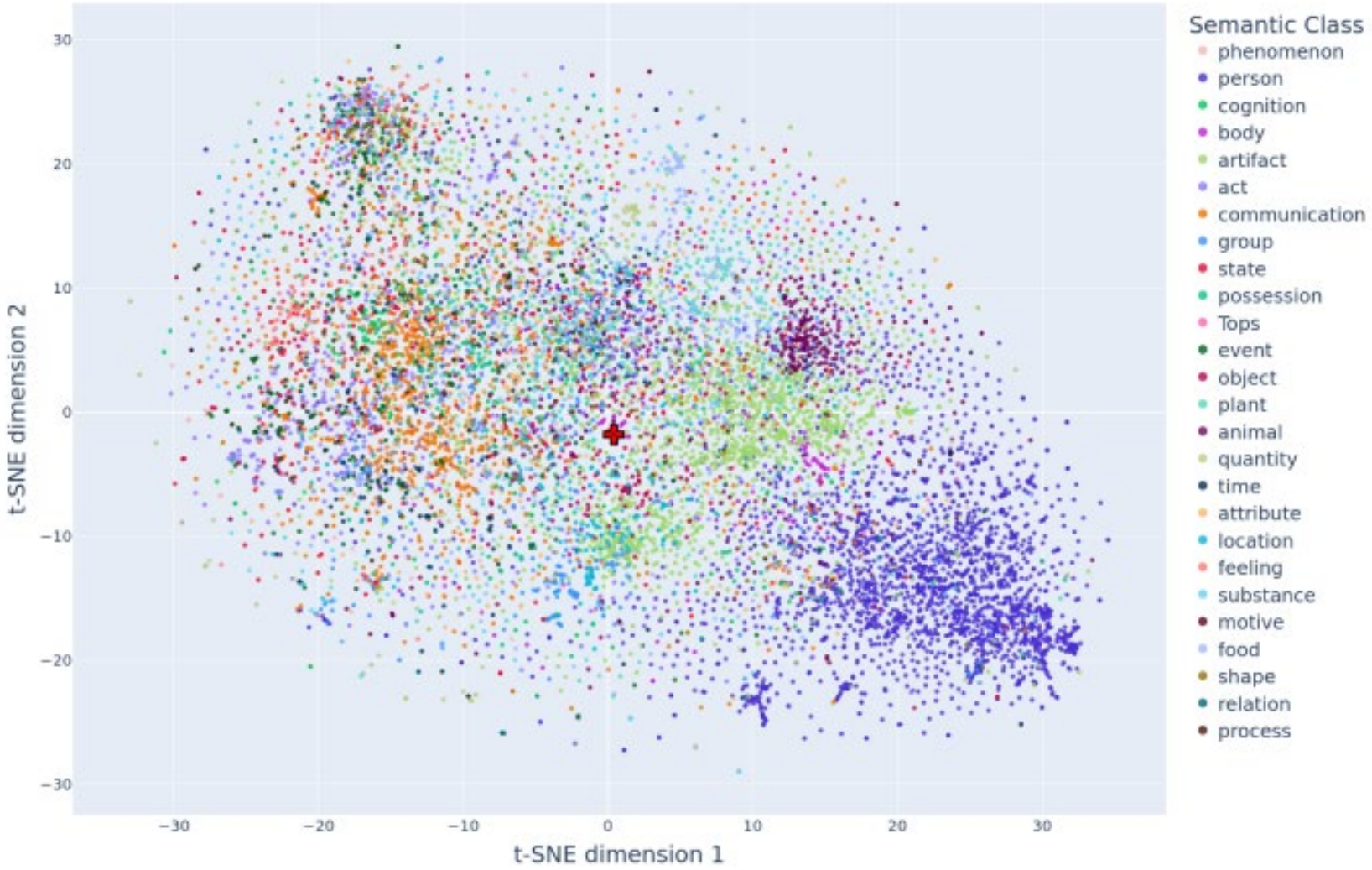
average shift vector: the difference vector between the average of plural vectors and the average of singular vector

$$[\![plural_{animal}]\!]_{\text{cows-cow}} \neq [\![plural_{item}]\!]_{\text{pens-pen}}$$

Semantic properties of English nominal pluralization: Insights from word embeddings

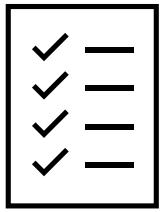
(Elnaz Shafaei-Bajestan, Masoumeh Moradipour-Tari, Peter Uhrig and R. Harald Baayen, 2022)

$$\overrightarrow{bananas} = \overrightarrow{banana} + \overrightarrow{AVG{-}SHIFT\ fruit}$$



PROJECT IDEA

Morphologically rich language



at least six cases

singular/plural

at least 12 noun forms

Gender: masculine, feminine, neuter

Examples of noun declension

case	singular	plural
nom 	jablok-o	jablok-i
acc	jablok-o	jablok-i
gen	jablok-a	jablok -ø
dat	jablok-u	jablok-am
instr	jablok-om	jablok-ami
prep	jablok-e	jablok-ach

ENG: apple, neuter

case	singular	plural
nom 	sol’	sol-i
acc	sol’	sol-i
gen	sol-i	sol-ei
dat	sol-i	sol-jam
instr	sol-’ju	sol-jami
prep	sol-i	sol-jach

ENG: salt, feminine

Examples of noun declension

case	singular	plural
nom 	kukuz-a (corn)	kukuruz-y
acc	kukuruz-u	kukuruz-y
gen	kukuruz-y	kukuruz-ø
dat	kukuruz-e	kukuruz-am
instr	kukuruz-oi	kukuruz-ami
prep	kukuruz-e	kukuruz-ach

ENG: corn, feminine

case	singular	plural
nom	mango	mango
acc	mango	mango
gen	mango	mango
dat	mango	mango
instr	mango	mango
prep	mango	mango

ENG: mango, neuter

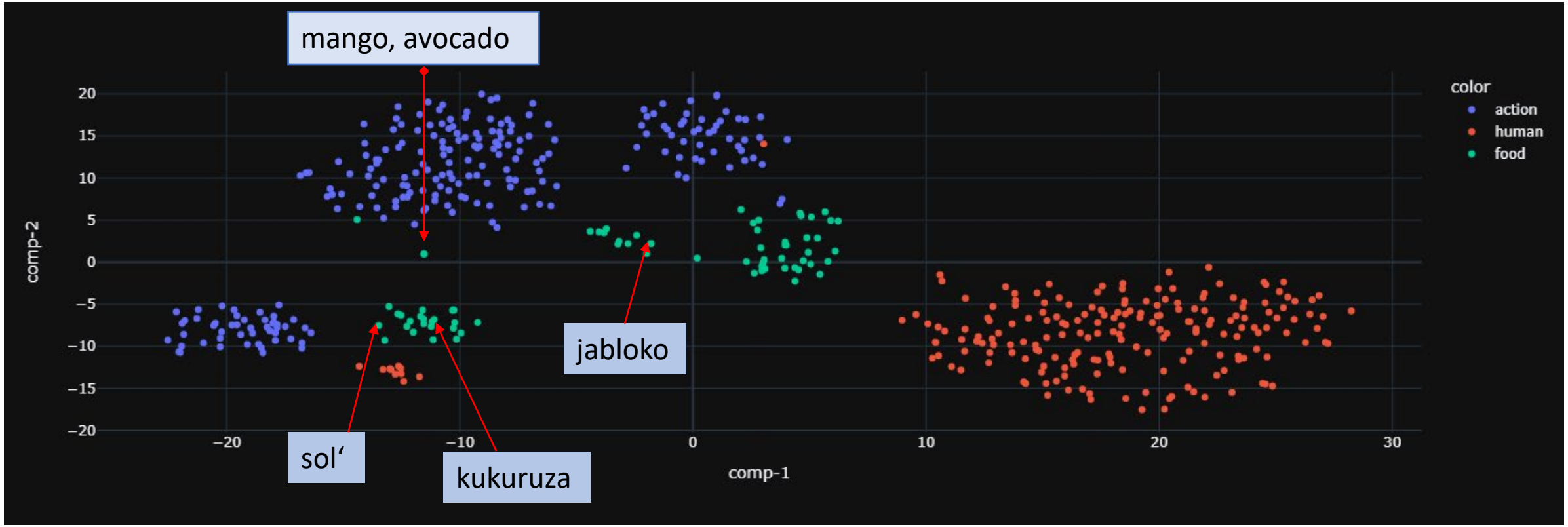
$$\overrightarrow{banana}_{gen\ sg} - \overrightarrow{banan}_{nom\ sg} = \overrightarrow{AVR\ SHIFT general}$$

$$\overrightarrow{banana}_{gen\ sg} - \overrightarrow{banan}_{nom\ sg} = \overrightarrow{AVR\ SHIFT food}$$

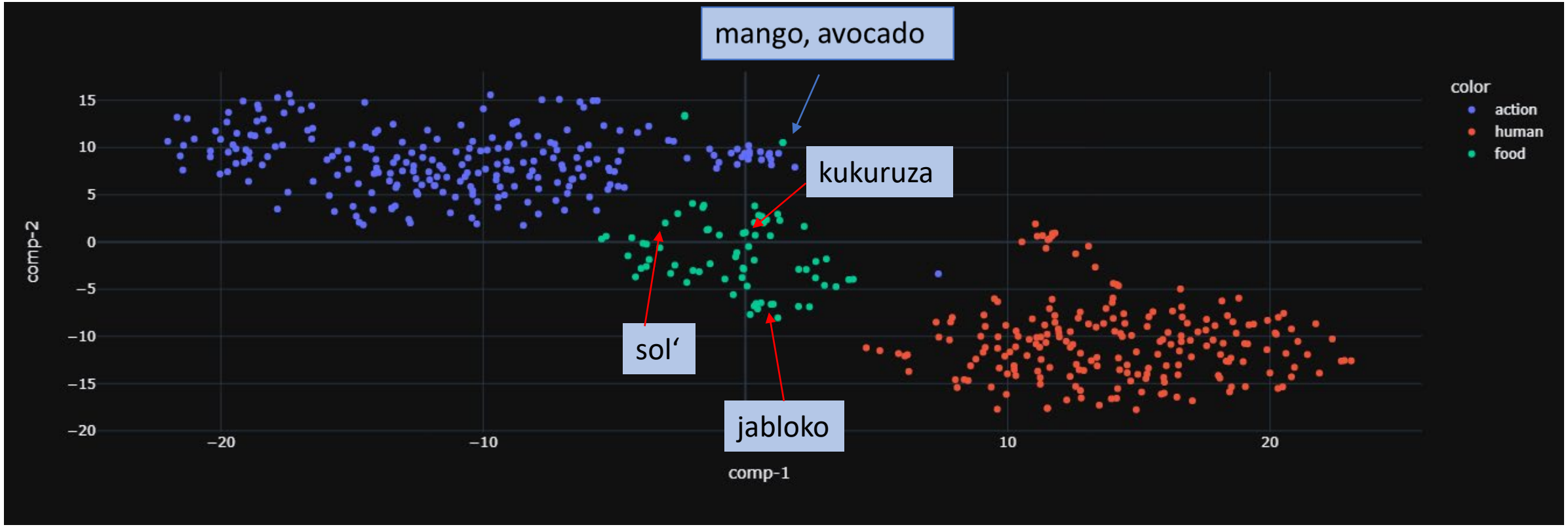
$$\overrightarrow{shift} = \overrightarrow{case} - \overrightarrow{NOM}$$

RESULTS

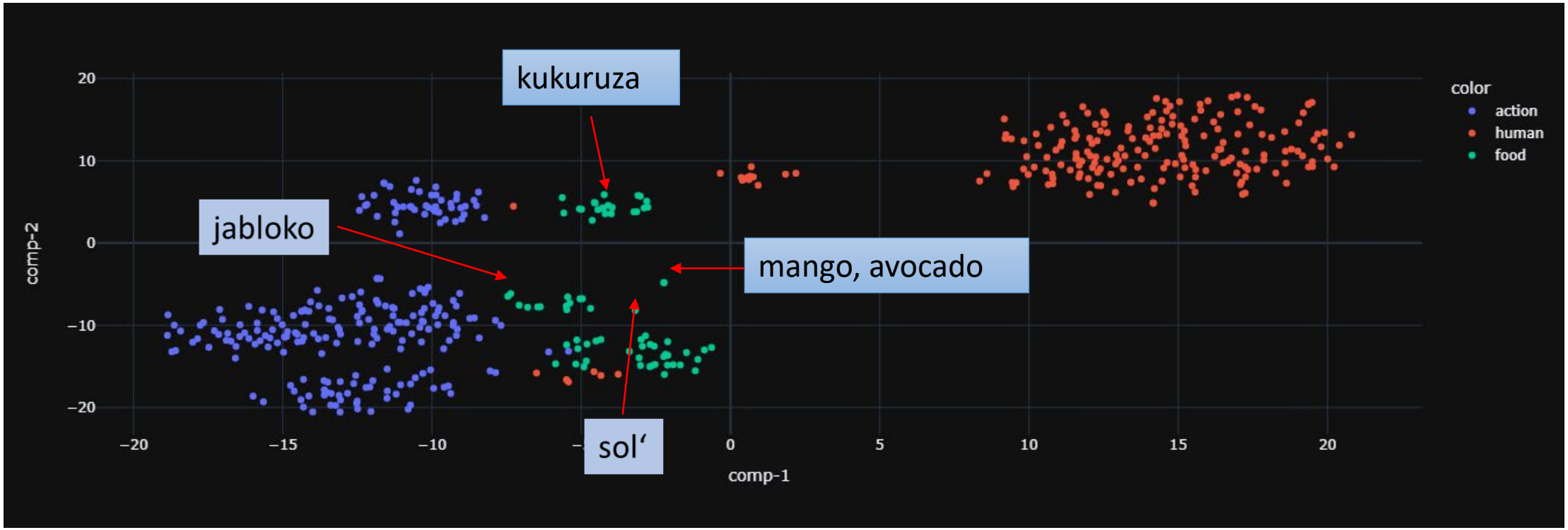
Shift Vector: gen_sg – nom_sg



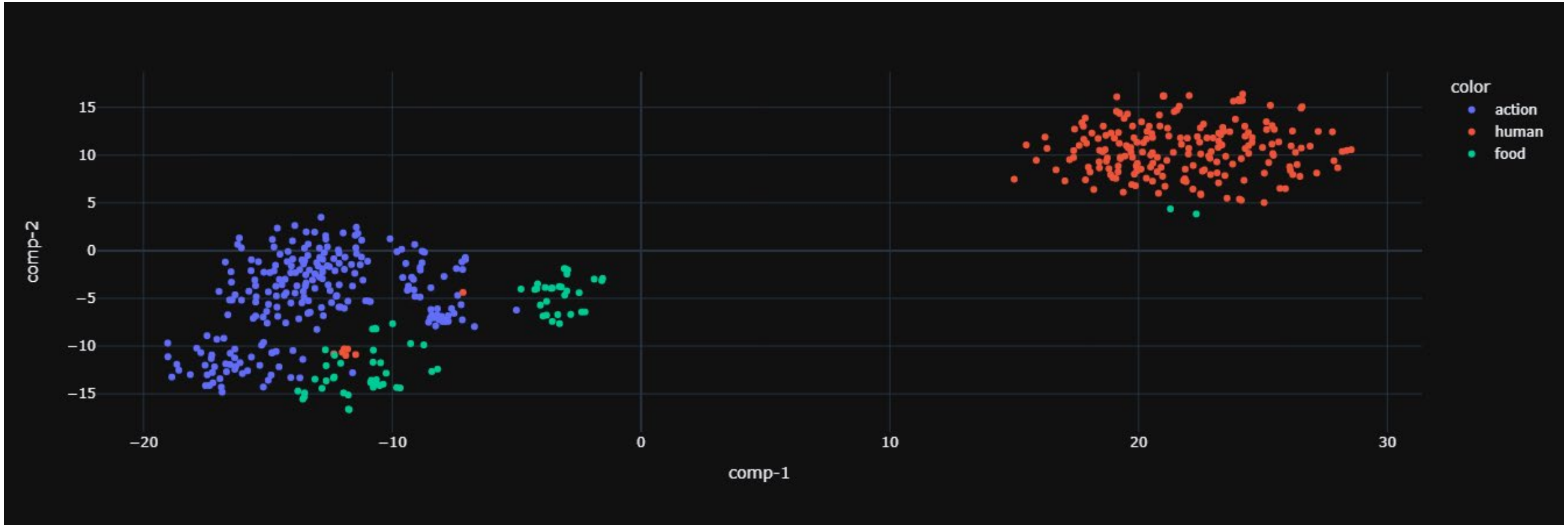
Shift Vector, Genitive Singular – Base



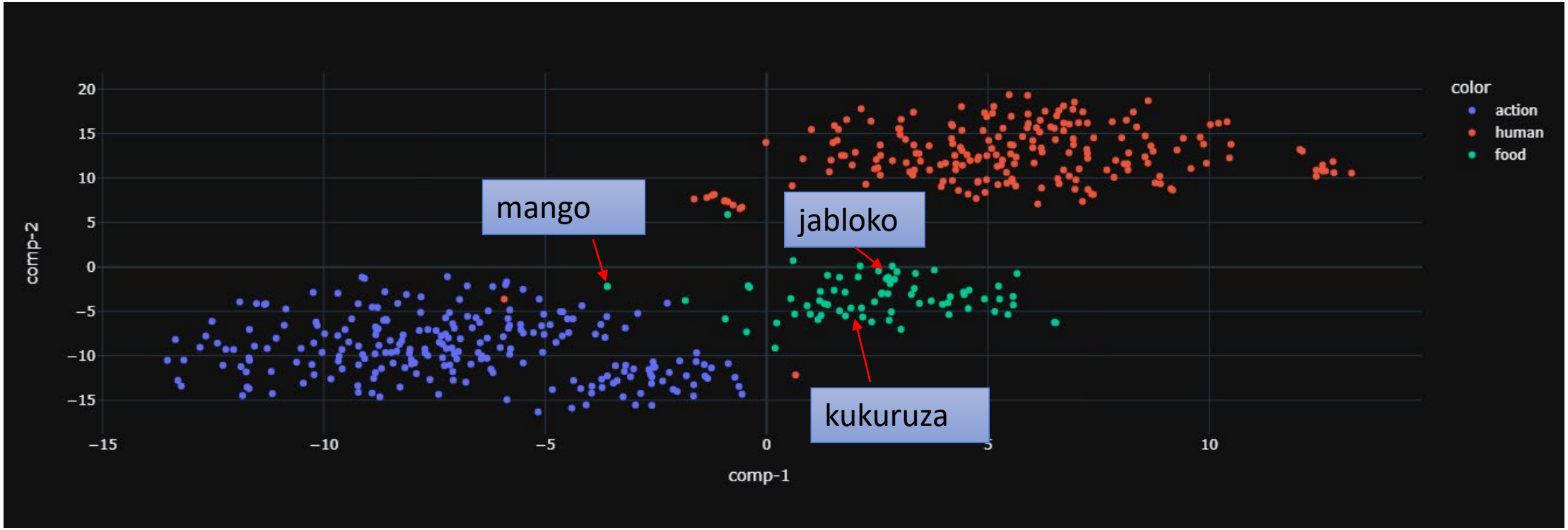
Shift Vector, Accusative Singular



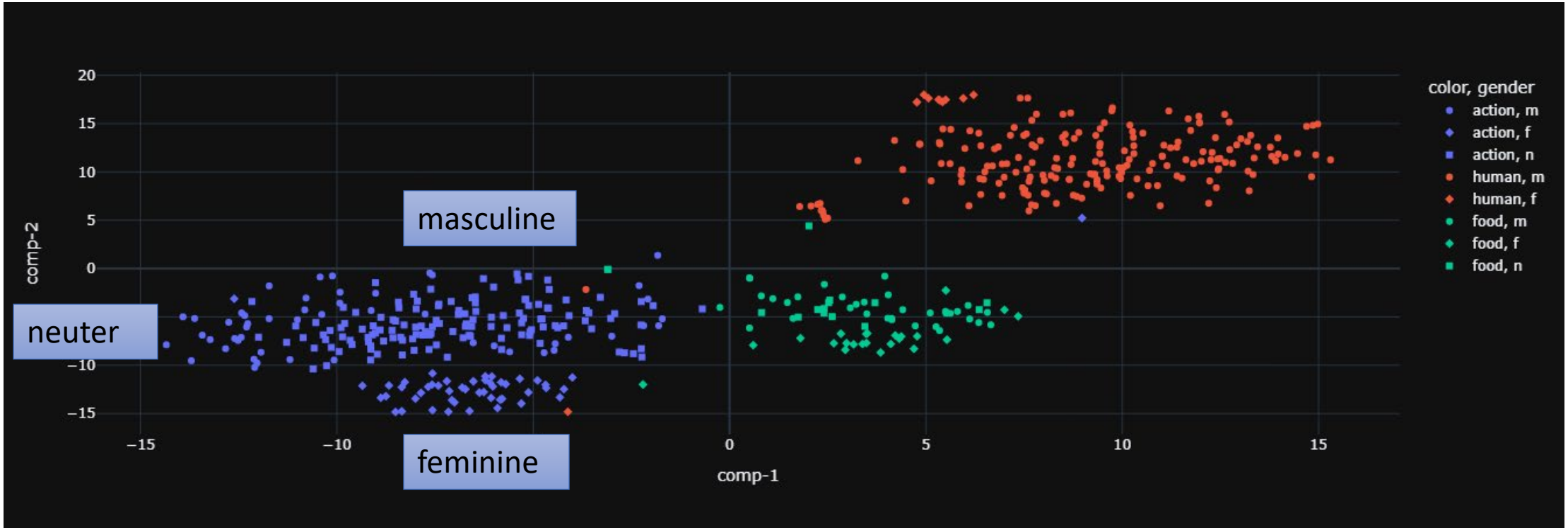
Shift Vector, Accusative Plural



Shift Vector, Instrumental Singular



Shift Vector, Instrumental Singular, Gender



Summary

Base vector = average(word paradigm)

Shift vectors can be divided into clusters according to the semantic groups

Shift vectors have more features inside (gender, declension)

THANK YOU!

Any Questions?

Anastasia Yablokova, anyab100@hhu.de