Shigeki Yoshida/Mai Hayashi/Sakura Ishikawa/Yuko Morokuma/Yuta Sakon/Yui Suzuki/Mizuki Tanigawa/Naonori Nagaya

# Uniplex/multiplex pairs and frequency asymmetries in general number languages

**Keywords:** number; frequency; plurative; singulative; general number language

## Background

This paper examines uniplex/multiplex pairs of nouns across languages. The terms *uniplex* and *multiplex* are used here as notional concepts (Talmy 1988). Uniplex nominals denote a single entity, while multiplex nominals denote a set of multiple entities. In number-marking languages, uniplex and multiplex nominals receive different marking. In English, for example, multiplex nominals are expressed by overt plurative forms, as in (1). In Welsh, by contrast, uniplex nominals are overtly marked by singulative forms, as in (2).

(1)   English

   a.   *dog-ø*          (uniplex meaning, basic form)
   b.   *dog-s*          (multiplex meaning, plurative form)

(2)   Welsh (Haspelmath/Karjus 2017: 1214)

   a.   *pys-en* 'pea'      (uniplex meaning, singulative form)
   b.   *pys-ø* 'peas'      (multiplex meaning, basic form)

Haspelmath/Karjus (2017) propose that these coding asymmetries can be explained with reference to usage frequency (see also Zipf 1935; Fenk-Oczlon 1991; Hawkins 2004; Haspelmath 2008). Namely, more coding is used for less frequent meanings: across languages, plurative-prominent meanings (i.e., noun meanings that are frequently expressed by plurative lexemes) tend to occur frequently in uniplex use, while singulative-prominent meanings (noun meanings that are frequently expressed by singulative lexemes) tend to occur frequently in multiplex use (Haspelmath/Karjus 2017: 1219). Empirically, Haspelmath/Karjus (2017) provide corpus evidence from five number-marking languages (English, Estonian, Latvian, Norwegian, and Russian).

## Research questions

This paper is a replication and extension study of Haspelmath/Karjus (2017) from a contrastive linguistic perspective. Crucially, we test their hypothesis not only against languages with obligatory number marking but also against general number languages, "in which the meaning of the noun can be expressed without reference to number" (Corbett 2000: 10).

The present paper seeks to address two research questions.

[A] Is Haspelmath/Karjus' (2017) hypothesis replicated in other languages with obligatory number marking, including singulative languages such as Sinhala?

[B] Are singulative-prominent lexemes more frequently used with a multiplex meaning than with a uniplex meaning even in general number languages?

## Methods

To answer these two questions, we examined large corpora from four number-marking languages (Hindi, Sinhala, Spanish, and Swedish) and seven general number languages (Bengali, Indonesian, Japanese, Korean, Tagalog, Turkish, and Quechua) (Table 1). Following Haspelmath/Karjus (2017), we analysed the frequencies of 18 lexemes in each language: EAR, LEG, LUNG, GLOVE, SHOE, SKI, APPLE, POTATO, STRAWBERRY, BEE, PIGEON, SHEEP, CHILD, BOY, GIRL, EUROPEAN, AMERICAN, SPEAKER OF (THE RESPECTIVE LANGUAGE). These lexemes are singulative-prominent in the sense that they crosslinguistically tend to receive singulative marking. In addition, we looked at 18 random lexemes in each language, with the hypothesis that random lexemes would not exhibit the same usage patterns as the 18 singulative-prominent lexemes.

| Language | Language family | Language type | Corpus |
|---|---|---|---|
| Hindi | Indo-European | plurative | hiTenTen17 |
| Spanish | Indo-European | plurative | esTenTen18 |
| Swedish | Indo-European | plurative | svTenTen14 |
| Sinhala | Indo-European | singulative | OpenSubtitles 2018 |
| Bengali | Indo-European | general number | bnWaC |
| Indonesian | Austronesian | general number | tufs_web_2012 |
| Japanese | Japonic | general number | BCCWJ |
| Korean | Koreanic | general number | koTenTen18 |
| Tagalog | Austronesian | general number | tlTenTen19 |
| Turkish | Turkic | general number | trTenTen12 |
| Quechua | Quechuan | general number | Wikipedia |

Table 1: Samples of languages analysed in this study

## Coding and annotation

For number-marking languages, we counted the number of basic and derived (i.e., plurative or singulative) forms of nouns. For general number languages, we took 40 random samples of each noun from the corpus and manually annotated the counts of semantically uniplex and multiplex nouns.

To capture the difference between the counts, an "asymmetry index" with a range of −1…1 was used, where negative values indicate dominant uniplex usage, and positive values dominant multiplex usage (Haspelmath/Karjus 2017: 1225).

## Results

The preliminary findings of our study are presented in Figures 1 and 2, where 'R' marks the random lexemes. Note that Figures 1 and 2 are tentative and subject to further analysis and validation.
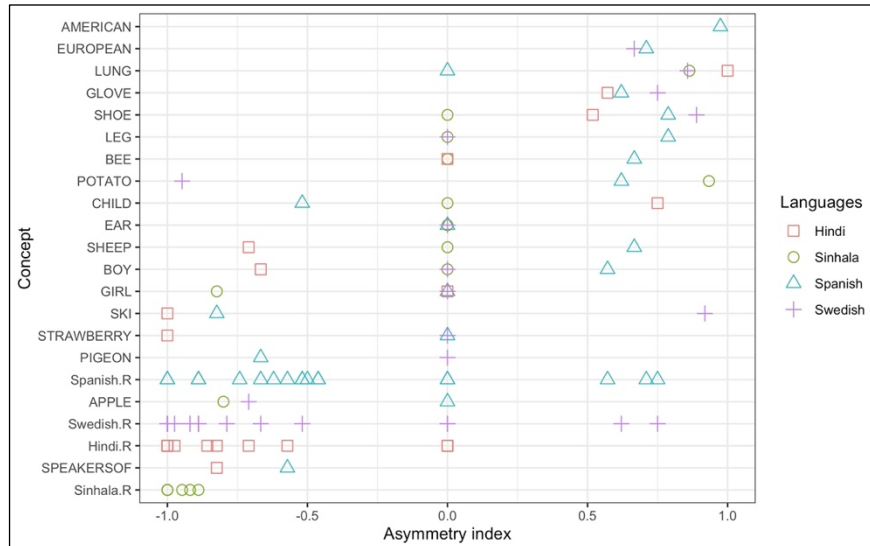


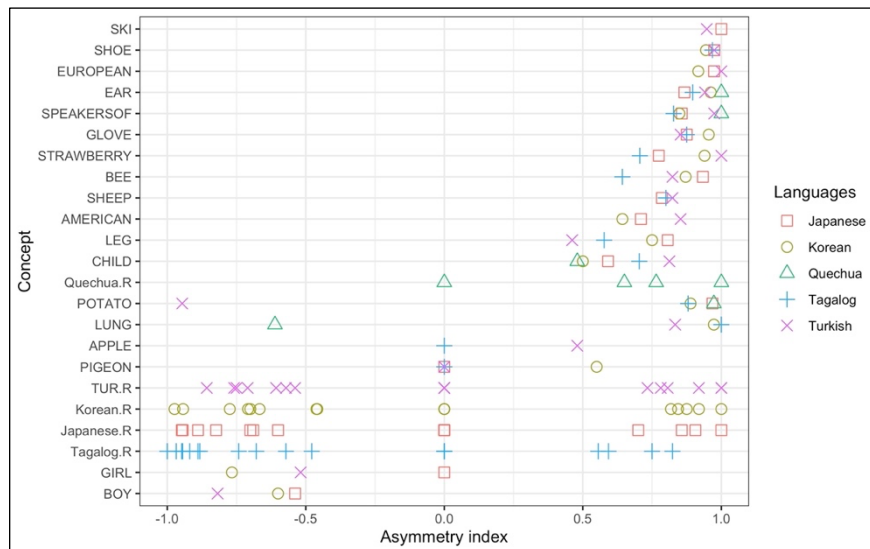Fig. 1: The asymmetry index in number-marking languages



Fig. 2: The asymmetry index in general number languages

Figure 1 shows that, in the number-marking languages we examined (Hindi, Spanish, and Swedish), most singulative-prominent lexemes tend to be more frequent in plurative forms than in basic forms, compared to randomly sampled nouns. It also indicates that, in Sinhala, the singulative-prominent lexemes tend to appear more frequently in basic forms than in singulative forms, compared to randomly sampled nouns.

Figure 2 summarises the results of the general number languages we examined (Japanese, Korean, Tagalog, Turkish, and Quechua). It shows that the singulative-prominent lexemes strongly tend to be more frequently employed in a multiplex sense than in a uniplex sense, compared to randomly sampled nouns.

## Discussion

The above results show that the answer to both research questions [A] and [B] is yes. First, Haspelmath/Karjus' (2017) hypothesis is replicated in Hindi, Sinhala, Spanish, and Swedish and proves to be a robust hypothesis. Importantly, this is the first study to demonstrate that singulative-prominent lexemes are more frequently expressed by basic forms than by singulative (derived) forms in a singulative language like Sinhala. These findings are consistent with the predictions made by Haspelmath/Karjus (2017), providing further support for the validity and generalisability of their findings.

Second, this study also shows that general number languages exhibit the same kind of frequency asymmetries as number-marking languages, even though they lack form distinctions between uniplex and multiplex forms. Thus, by contrasting two different types of languages, this study suggests that frequency asymmetries between uniplex and multiplex nouns universally exist, although they do not always result in coding asymmetries.

## References

Corbett, Greville G. (2000): Number. Cambridge: Cambridge University Press.

Fenk-Oczlon, Gertraud (1991): Frequenz und Kognition–Frequenz und Markiertheit. In: Folia Linguistica, 25(3–4), pp. 361–394. https://doi.org/10.1515/flin.1991.25.3-4.361.

Haspelmath, Martin (2008): Frequency vs. iconicity in explaining grammatical asymmetries. In: Cognitive Linguistics, 19(1), pp. 1–33. https://doi.org/10.1515/COG.2008.001.

Haspelmath, Martin/Karjus, Andres (2017): Explaining asymmetries in number marking: Singulatives, pluratives, and usage frequency. In: Linguistics, 55(6), pp. 1213–1235. https://doi.org/10.1515/ling-2017-0026.

Hawkins, John A. (2004): Efficiency and complexity in grammars. Oxford: Oxford University Press.

Talmy, Leonard (1988): The relation of grammar to cognition: A synopsis. In: Rudzka-Ostyn, Brygida (ed.): Topics in cognitive linguistics. Amsterdam: John Benjamins, pp. 166–205. https://doi.org/10.1075/cilt.50.08tal.

Zipf, George Kingsley (1935): The psycho-biology of language: An introduction to dynamic philology. Cambridge, MA: MIT Press.

## Contact information

**Shigeki Yoshida**

The University of Tokyo

shige.mountain.linguistics@gmail.com


**Mai Hayashi**

The University of Tokyo

maimaimaingay@gmail.com

**Sakura Ishikawa**

Tokyo University of Foreign Studies

sakura.ishikawa.di2@gmail.com

**Yuko Morokuma**

National Institute for Japanese Language and Linguistics

y.ukumari322@gmail.com

**Yuta Sakon**

The University of Tokyo

y.sakon322@gmail.com

**Yui Suzuki**

The University of Tokyo

suzuki.yui.s.y@gmail.com

**Mizuki Tanigawa**

The University of Tokyo

mizuki.tanigawa.ling@gmail.com

**Naonori Nagaya**

The University of Tokyo

nagaya@l.u-tokyo.ac.jp