

## Gender and numeral classifiers in Modern Nepali and their Proto-Indo-European analogues

Marcin Kilarski & Piotr Gąsiorowski\*

**Abstract.** Proto-Indo-European (PIE) had two grammatical genders (common and neuter); the feminine was a shared innovation of the non-Anatolian part of the family. Using comparison with Modern Nepali, we argue that the puzzling feminine forms of the numerals ‘3’ and ‘4’ in Celtic and Indo-Iranian, and possibly also some similar constructions in Latin and Old Irish, are survivals of a system of numeral classifiers predating the full gender system. They contain the feminine element *\*-s(o)r-*, grammaticalized as a numeral classifier in PIE. A similar situation is attested in Nepali, where grammatical gender occurs alongside numeral classifiers. Analogies between numeral phrases in PIE and Nepali help elucidate the historical development in question.

**Keywords.** Nepali; Proto-Indo-European; nominal classification; grammatical gender; numeral classifiers

**1. Introduction.** Although there is extensive literature on grammatical gender, relatively little attention has been devoted to well-described languages with concurrent classification systems, i.e., either different gender systems or gender and classifiers. In this paper we illustrate four attempts in the history of Indo-European at the development of systems of numeral classifiers – either successful (in Modern Nepali) or ultimately unsuccessful (in Latin, Celtic and Indo-Iranian). These incipient systems of numeral classifiers developed at different stages with regard to the evolving complexity of Indo-European gender and in different sociolinguistic and language contact situations.

**2. Gender and numeral classifiers.** Gender and classifiers are the two commonly distinguished types of nominal classification, i.e., “[...] classification of nouns and/or extralinguistic entities to which nouns refer that is grammaticalized to some degree, and expressed in one or more syntactic contexts that relate to nouns.” (Contini-Morava & Kilarski 2013:265). Numeral classifiers appear as independent words or affixes in contexts of quantification (on numerals or quantifiers). Numeral classifiers are commonly subclassified as sortal and mensural. Sortal classifiers are obligatory in the context of enumeration and have the value ‘one’; referents are classified according to their inherent features. In contrast, mensural classifiers (measure words) provide new information about the quantity of the associated noun, which is not necessarily ‘one’ (Her 2012). In earlier work, gender and classifiers were distinguished based on such criteria as the presence of agreement, realization, applicability, assignment principles, size of inventory and degree of variability (Dixon 1982 and later typologies). In more recent work, Fedden & Corbett (2017) define a canonical gender system, where each noun has a single gender value that is based on its meaning and remains the same for all agreement targets across all domains of agreement. On this view, classifiers are classification systems which differ in various ways from this canonical system.

---

\* Authors: Marcin Kilarski, Adam Mickiewicz University, Poznań ([kilarski@amu.edu.pl](mailto:kilarski@amu.edu.pl)) & Piotr Gąsiorowski, Adam Mickiewicz University, Poznań ([gpiotr@wa.amu.edu.pl](mailto:gpiotr@wa.amu.edu.pl)).

Gender and numeral classifiers have a near-complementary distribution independent of areal and genetic factors (Sinnemäki 2019) (see Figure 1). Both types of systems appear in 22 (6%) out of 360 languages in Sinnemäki (2019), e.g. Fula, Ejagham (Niger-Congo), Pnar, Munda, Nicobarese (Austroasiatic) and Halkomelem (Salishan). The distribution of gender and numeral classifiers has been analysed in terms of a complexity trade-off (Sinnemäki 2019) as well as an effect of the complementary nature of their functions (Allasonnière-Tang & Kilarski 2020).

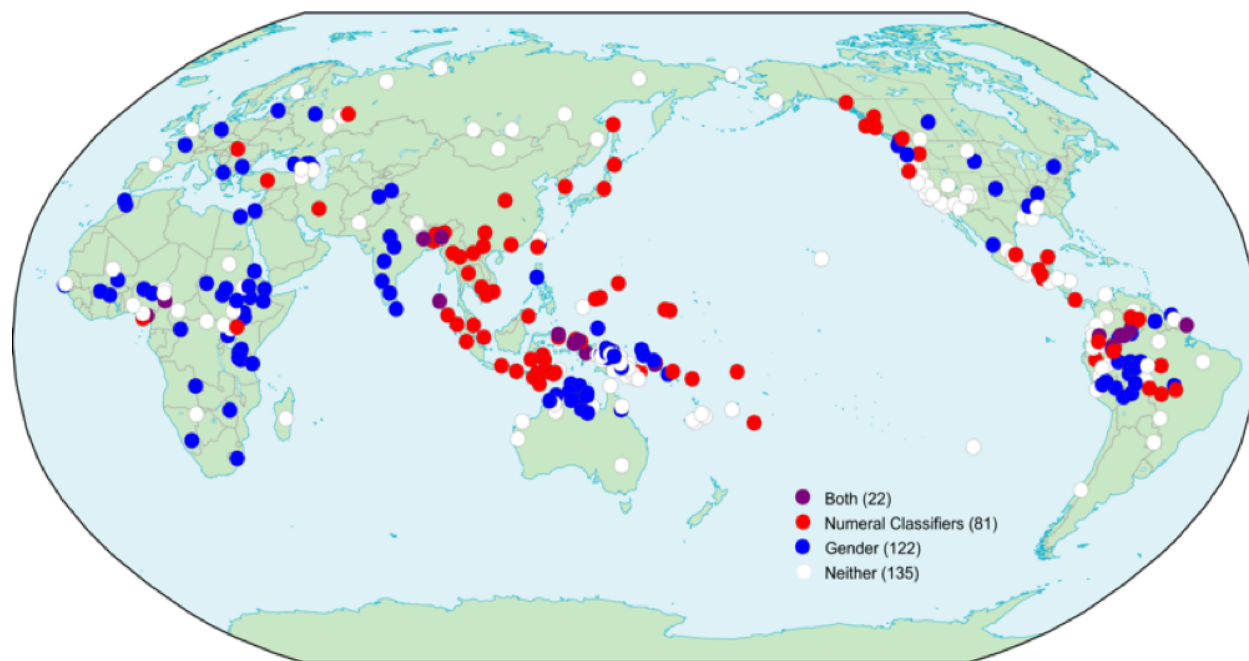


Figure 1. Distribution of gender and numeral classifiers (Sinnemäki 2019:151)

**3. Evolution of gender in Indo-European.** According to current consensus (as summarised by Luraghi 2011), Proto-Indo-European (PIE) had a two-gender system based on the distinction between animate and inanimate (“neuter”) nominals. The animate class was characterised by having different nominative and accusative forms in the singular and the count plural. No such contrast is in evidence in the dual of animate nominals or in the collective plural with the *\*-h<sub>2</sub>* ending, which could originally be formed to both animates and inanimates despite its later specialisation as neuter plural. We agree with Luraghi’s opinion that in the PIE system “... gender was in a way an epiphenomenon of case marking patterns, and it arose as a consequence of inanimate nouns not having endings for the nominative and the accusative case” (Luraghi 2011:457). We would describe it as an incipient category, not yet fully established. Its derivational function was limited at best. There were no gender-marking affixes, although animacy was to some extent correlated with accent-and-ablaut patterns and it was possible to form the animate counterpart of an inanimate noun by so-called internal derivation (Fortson 2010:122).

The rise of the feminine gender must have occurred rapidly between the separation of the Anatolian branch from the common ancestor of the rest of the Indo-European family and the subsequent split into the Tocharian lineage and the “crown group” (Inner IE) which includes all the extant Indo-European languages. Although the details of the process are still debated (cf. Kim (2014) for a recent attempt to elucidate them), it is clear that the new feminine gender had salient morphological markers in the major declensions and enforced feminine agreement between nouns and their modifiers (adjectives and pronouns) in non-Anatolian IE. Although not all

nominals carried explicitly marked gender, all former animates became polarised: some were assigned to the feminine, and the remaining ones became masculine by default. The result was the fully-fledged “classical” three-gender system (masculine, feminine, neuter) of Inner IE. It has often been reduced in the modern languages (and thus already in the Tocharian group, which had retained the feminine but lost the neuter). Occasionally, however, we can observe its further complexification – see the division of the inherited masculine into personal, animate and inanimate subgenders in Polish and a similar proliferation of subgenders in other Slavic languages (Corbett 1991:98–101, 161–168).

**4. Nominal classification in Nepali.** Nepali distinguishes both gender and numeral classifiers due to long-term contact with Tibeto-Burman languages. There is a striking disagreement in the literature concerning the number of gender and numeral classifiers in Nepali, which can be attributed to considerable areal variation, use of different definitional criteria and citations of earlier inaccurate accounts (0-11 genders; from 2 to >200 classifiers) (see, e.g. Acharya 1991, Pokharel 2010). Following Allasonnière-Tang & Kilarski (2020), we distinguish two co-existing gender systems, i.e., Gender 1 (masculine vs. feminine) and Gender 2 (human vs. non-human) as well as a numeral classifier system.

4.1. GENDER. With regard to assignment principles, Gender 1 is sex-based: female humans and female animals are feminine, with the residue assigned to the masculine gender. Gender agreement is found in adjectives, verbs, possessive adjectives and ordinal numerals as well as the general classifier (see Example (1)):

(1) Masculine vs. feminine gender in adjectives, verbs and possessive adjectives in Nepali (Allasonnière-Tang & Kilarski 2020:121–122)

- a. mer-o ramr-o keto nepali bolcha  
my-M beautiful-M boy(M) Nepali speak.PRS.3SG.M  
'My handsome boyfriend speaks Nepali.'
- b. mer-i ramr-i keti nepali bolche  
my-F beautiful-F girl(F) Nepali speak.PRS.3SG.F  
'My beautiful girlfriend speaks Nepali.'
- c. mer-o kitaab yahan cha  
my-M book(M) here be.PRS.3SG.M  
'My book is here.'

In contrast, Gender 2 is based on the human vs. non-human opposition; gender agreement is found in 3rd person personal pronouns (see Example (2)).

(2) Human vs. non-human gender in personal pronouns in Nepali (Allasonnière-Tang & Kilarski 2020:121)

- a. u ramr-o cha  
he/she beautiful-M be.PRS.3SG.M  
'He is handsome.'
- b. u ramr-i che  
he/she beautiful-F be.PRS.3SG.F  
'She is beautiful.'
- c. tyo ramr-o cha  
it beautiful-M be.PRS.3SG.M  
'It (e.g. a house) is beautiful.'

4.2. NUMERAL CLASSIFIERS. With regard to numeral classifiers, Allasonnière-Tang & Kilarski (2020) describe 10 classifiers, including a general classifier, a human classifier, together with inanimate classifiers based on shape, dimensionality and material. The classifiers are given in Table 1:

Classifier	Meaning	Example
<i>jana</i>	human	man, woman, uncle, aunt
<i>wota</i>	general	book, car, shop, telephone
<i>dana</i>	round fruits	apple, grape, orange
<i>sinka</i>	long object	noodle, bamboo
<i>ghoga</i>	long plant	maize
<i>geda</i>	grain	mustard, maize, rice
<i>pana</i>	two-dimensional	paper
<i>than</i>	two-dimensional and large	old hand-made paper
<i>koso</i>	natural capsule	banana, bean, pea
<i>khili</i>	artificial capsule	cigarette, betelnut

Table 1: Numeral classifiers in Nepali (Allasonnière-Tang & Kilarski 2020:127)

The classifiers occur in the context of quantification, i.e., with numerals and quantifiers; the noun phrase ordering is Numeral-Classifier-Noun.

As mentioned above, the general classifier *wota* shows gender agreement. Similarly to other numeral classifier languages such as Mandarin Chinese, the classifier occurs with both animate and inanimate nouns. However, the general classifier in Nepali occurs with numerals either independently, as in (3a) and (3c), or as fused with a numeral, as in (3b) and (3d). In both cases the classifier exhibits gender agreement (for further discussion see Kilarski & Tang 2018).

(3) Gender agreement on the general classifier in Nepali (Allasonnière-Tang & Kilarski 2020:130)

- a. tin wot-a keto  
three CLF.GENERAL-M boy(M)  
'three boys'
- b. tin-t-a keto  
three-CLF.GENERAL-M boy(M)  
'three boys'
- c. tin wot-i keti  
three CLF.GENERAL-F girl(F)  
'three girls'
- d. tin-t-i keti  
three-CLF.GENERAL-F girl(F)  
'three girls'

Such morphosyntactic behaviour of a classifier is typologically rare. The only other case that we are familiar with where a classifier exhibits gender agreement occurs in Ayoreo (Zamucoan; Bolivia and Paraguay), which distinguishes possessive classifiers which show agreement in gender and number (Ciucci & Bertinetto 2019).

**5. Earlier analogies.** Although the numeral classifier constructions that occur in Nepali are obviously innovations restricted to a small subset of the modern Indo-Aryan languages, there is evidence that similar developments took place in the more remote past of several linguistic lineages of Indo-European and possibly as far back as the PIE stage. Below we show how a noun meaning ‘man’ was co-opted in a function akin to that of numeral classifiers in Latin and Old Irish, and a noun presumably meaning ‘woman’ was treated in the same way in very early Indo-European.

5.1. LATIN AND GOIDELIC. As noted above, the incipient two-way gender system of PIE developed into a fully established three-gender system of Inner IE. Against this background we can observe some isolated experiments in marking natural gender for the purpose of counting with elements resembling numeral classifiers. In Latin, habitual collocations of numerals with *virī* ‘men’ show partial grammaticalization (Hackstein 2010:53–55). We find a range of examples, from fully transparent ones like *duo-virī* and *trēs-virī* to analogical backformations like *triumvirī* (from genitive plural *trium-virum*) and partly obscured univerbations like *sēvirī* < *sex-virī*. They were by no means restricted to the lower numerals, cf. *quindecim-virī* ‘fifteen men’ and *centum-virī*. Such extended numerals were often used to modify nouns, as in *trēs-virī epulōnēs* ‘three.M banquet-masters’ (see Example (4)). They referred to officials – members of boards, colleges or commissions – and their function was to emphasise the fact that the offices in question were male-only. Since Latin numerals higher than ‘two’ did not have contrastive feminine and masculine forms, the numeral by itself could not transmit this socioculturally important information.

- (4) The use of a numeral + *virī* in Latin as a modifier preceding a noun in the NOM.PL
- |                             |                          |
|-----------------------------|--------------------------|
| <i>trēs-virī</i>            | <i>epulōnēs</i>          |
| three:NOM.PL-man:NOM.PL     | banquet-master(M):NOM.PL |
| ‘the three banquet-masters’ |                          |

In Old Irish, reflexes of numerals compounded with *\*-uīh₁rom* (neuter, with a collective meaning, derived from masculine *\*uīh₁ros* ‘adult man, male human’ > OIr. *fer*), were construed with nouns in the genitive plural (see Example (5)).

- (5) The use of a reflex of *\*-uīh₁rom* fused with a numeral preceding a noun in the GEN.PL in Old Irish (after Hackstein 2010:57)
- |                   |                   |
|-------------------|-------------------|
| <i>triar</i>      | <i>churad</i>     |
| three:HUMAN.M.NOM | warrior(M):GEN.PL |
| ‘three warriors’  |                   |

They were lexicalised compounds rather than fixed collocations: *\*tri-uīh₁rom* ‘a team of three men’ > OIr. *triar* (cf. non-human masculine and neuter *tri*, feminine *teoir*, on which see below). Apart from numerals proper, OIr. *il* ‘many’ (< *\*pelh₁u-*) also formed a compound with *\*-uīh₁rom*, *ilar* ‘multitude, abundance’. The grammaticalization of such combinations was more thorough in Old Irish and its Goidelic descendants than in Latin. Since their structure became opaque very early, their etymological masculinity did not prevent them from being extended to female referents. Thus they have evolved into generalised personal rather than human-male numerals (Hackstein 2010).

It should be noted that in noun phrases involving these numeral expressions the order of elements is what one would expect in a language making use of numeral classifiers: Numeral-Masc.Marker-Noun, with the masculinity marker either following the numeral as an independent unit or fused with it.

5.2. PROTO-INDO-EUROPEAN *SOR*- ‘FEMALE’. In Gąsiorowski & Kilarski (2019) we point out that similar experiments seem to have taken place in PIE. They left telling traces in the numeral systems of two distant branches of the family, Indo-Iranian and Celtic. The notoriously enigmatic feminine forms of ‘3’ (*\*tisr-*) and ‘4’ (*\*k<sup>w</sup>etesr-*) in those two branches (see Example (6) below) can be explained as unverbated expressions *\*tri(su)-s(o)res* and *\*k<sup>w</sup>et<sup>u</sup>er-s(o)res*. The compositional element *\*-s(o)r-* is otherwise known to have formed words for female humans (also in Anatolian, which had no feminine gender). There is evidence that it was once a free-standing noun meaning ‘female human, woman, lady’ (for a discussion of its form, cf. Harðarson 2014). In our reconstruction, the first element is a de-numeral locative adverb (‘in a group of *n*’) combined with the plural of *\*sor-*. The obviously archaic form of these numerals, strongly affected by segmental attrition already in the common ancestor of Indo-Iranian and Celtic, squares well with the conjecture that they represent an attempt to mark female reference in numerals before the emergence of the feminine gender – perhaps already in PIE.

- (6) Vedic (Rigveda 9.89.5) and Modern Welsh
- a. *cáta-sr-a(s)*                      *ghrta-dúh-aḥ*  
four-CLF:F-NOM.PL      butter-yielder(F)-NOM.PL  
‘four female butter-yielders’
  - b. *pedair dyn-es*  
four:F      person-F  
‘four women’<sup>1</sup>

As in Latin and Goidelic, the marker of natural gender (here, femininity) functioned like a numeral classifier and occupied the appropriate position in a noun phrase: Numeral-Fem.Marker-Noun. In both of the two reconstructible combinations *\*sor-* was completely fused with the preceding numeral.

In conclusion, the examples provided above illustrate several analogies in the meanings and expression of categorization at different stages in the history of Indo-European. Thus, there are semantic analogies between feminine/masculine markers in PIE and modern IE in terms of the expression of feminine/masculine reference by partly grammaticalized markers in earlier stages of IE and by gender agreement markers in Nepali. In addition, in both cases we deal with a classificatory system expressed on numerals, in which animate referents are classified with respect to sex. With regard to morphosyntactic analogies, we deal with the same ordering, i.e., Numeral-Fem/Masc.Marker-Noun. In line with Greenberg’s Universal 20 (Greenberg 1972), the classifier and numeral are contiguous. In addition, there are similar patterns in the expression of the classificatory markers. The PIE feminine marker fused with the numeral by forming a compound which became morphologically opaque, while in Nepali the feminine suffix appears with the general classifier either as part of an independent word or is fused with the numeral. Latin and Goidelic exhibit a similar use of a masculine marker, either as an independent unit or its residue fused with the numeral.

**6. Contexts of emergence, maintenance and loss of numeral classifiers.** The developments discussed here illustrate what types of internal and external contexts are required for the emergence, maintenance and loss of numeral classifiers. With regard to the emergence of a classificatory system, such morphosyntactic complexification can be either spontaneous or additive. In the former case, complexification was likely to occur considering the sociolinguistic

---

<sup>1</sup> Welsh uses nouns in the sg. with numerals; *-es* is a derivational suffix forming feminine nouns.

profile of early Indo-European as a ‘society of intimates’ (Givón 1979, Trudgill 2011), as illustrated by such social and demographic features as small communities, social stability, cultural uniformity and informational homogeneity. In the latter case, we deal with diffusion of numeral classifiers in Indo-Aryan languages, e.g. Nepali, Assamese and Bengali due to stable and long-term contact with Tibeto-Burman languages.

The required contexts for maintenance of numeral classifiers are illustrated by the functions for which they are exploited as nominal classification systems. Thus in Nepali a functional trade-off allows the coexistence of different classificatory systems. As shown by Allasonnière-Tang & Kilarski (2020), gender and classifiers contribute to the lexicon and discourse in complementary ways: a function is expressed by only one system; gender and classifiers are exploited with different types of nouns (animate vs. inanimate); or gender and classifiers are exploited for the same function in the same category of nouns to convey different meanings.

In Proto-Indo-European the two-gender system and the putative sex-based numeral classifier system (both of them incipient, though the former better established than the latter) expressed complementary meanings – animate:inanimate vs. masculine:feminine. They did not compete for the same function, which made their coexistence possible. The rise of the feminine gender resulted in an overlap with the meaning expressed by the feminine classifier and – by creating a sex-related distinction – with the masculine classifier, if the latter was already used alongside *\*sor-* (as suggested by Hackstein 2010:57). The conflict was resolved in favour of the more robust category. In both Celtic and Indo-Iranian the relicts of the feminine classifier found a new use as feminine gender numerals. They stood in a suppletive relationship with the regular animate forms *\*trejes* and *\*kvetuores*, reinterpreted as contrastively masculine in terms of grammatical gender. This resembles the functional shift of the Goidelic compound numerals discussed above (from male-human to personal).

Finally, both internal and external factors may contribute to the simplification of a classificatory system. The loss of numeral classifiers in PIE was due to further elaboration of the gender system, while the interplay between gender and numeral classifiers in varieties of Nepali depends on their proximity to either Tibeto-Burman or other Indo-Aryan languages. For example, more extensive inventories of classifiers are found in the east of Nepal in proximity to Tibeto-Burman languages, while in central Nepal fewer classifiers are used, mainly the general classifier *wota* and the human classifier *jana*.

Comparative evidence throws some light on the grammaticalization path of *\*sor-*. It was first a generic word for ‘female human being’, replaced in its lexicosemantic function by *\*g<sup>w</sup>enh<sub>2</sub>-* but retained as a compound element, gradually evolving into a bound morpheme, that is, in effect, a derivational suffix (best known from the Anatolian languages). It seems that it was also employed as a feminine numeral classifier. It is difficult to tell how productive the system was during its heyday, since only its scattered remains can be identified – the use of *\*sor-* is clearly attested on numerals ‘3’ and ‘4’, with no direct attestation for other numerals, and a similar use of *\*uih<sub>1</sub>ro-* can be inferred from only two branches, widely regarded as close relatives. All that we can see is debris from a failed experiment. The nascent numeral classifier system was outcompeted by a fully-fledged feminine gender, and the feminine classifier was exapted as a marginal feminine marker in the new agreement system.

**7. Conclusions.** According to our proposal both gender and numeral classifiers were once a common Indo-European trait. They developed in parallel in the protolanguage and stayed out of conflict with one another until the equilibrium was broken by the complexification of the gender system. As the Nepali parallel shows, this is a perfectly possible but insufficiently discussed sce-

nario, made still more likely by the sociolinguistic context of early Indo-European, which seems to have favoured morphosyntactic elaboration.

## References

- Acharya, Jayaraj. 1991. *A descriptive grammar of Nepali and an analyzed corpus*. Washington, D.C.: Georgetown University Press.
- Allasonnière-Tang, Marc & Marcin Kilarski. 2020. Functions of gender and numeral classifiers in Nepali. *Poznań Studies in Contemporary Linguistics* 56(1). 113–168. <https://doi.org/10.1515/psicl-2020-0004>.
- Ciucci, Luca & Pier Marco Bertinetto. 2019. Possessive classifiers in Zamucoan. In Alexandra Y. Aikhenvald & Elena Mihás (eds.), *Genders and classifiers: A cross-linguistic typology*. 144–175. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780198842019.001.0001>.
- Contini-Morava, Ellen & Marcin Kilarski. 2013. Functions of nominal classification. *Language Sciences* 40. 263–299. <https://doi.org/10.1016/j.langsci.2013.03.002>.
- Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139166119>.
- Dixon, R.M.W. 1982. “Noun classifiers and noun classes”, in: R.M.W. Dixon, *Where have all the adjectives gone? And other essays in semantics and syntax*. Berlin: Mouton, 211–233.
- Fedden, Sebastian & Greville G. Corbett. 2017. Gender and classifiers in concurrent systems: Refining the typology of nominal classification. *Glossa: A journal of general linguistics* 2(1). 1–47. <https://doi.org/10.5334/gjgl.177>.
- Fortson, Benjamin W. IV. 2010. *Indo-European language and culture: An introduction*. (2nd edn.) Malden, MA: Wiley-Blackwell.
- Gašiorowski, Piotr & Marcin Kilarski. 2019. Ex oriente lux: How Nepali helps to understand relict numeral forms in Proto-Indo-European. In Magdalena Wrembel, Agnieszka Kiełkiewicz-Janowiak & Piotr Gašiorowski (eds.), *Approaches to the study of sound structure and speech: Interdisciplinary work in honour of Katarzyna Dziubalska-Kolaczyk*. 76–84. London: Routledge. <https://doi.org/10.4324/9780429321757>.
- Givón, Talmy. 1979. *On understanding grammar*. Orlando, FL: Academic Press.
- Greenberg, Joseph H. 1972. Numeral classifiers and substantival number: Problems in the genesis of a linguistic type. *Working Papers in Language Universals* 9. 1–39.
- Hackstein, Olav. 2010. *Apposition and nominal classification in Indo-European and beyond*. Vienna: Verlag der Österreichischen Akademie der Wissenschaften.
- Harðarson, Jón Axel. 2014. Das andere Wort für ‘Frau’ im Urindogermanischen. In Sergio Neri & Roland Schuhmann (eds.), *Studies on the collective and feminine in Indo-European from a diachronic and typological perspective*. 23–54. Leiden: Brill. [https://doi.org/10.1163/9789004264953\\_004](https://doi.org/10.1163/9789004264953_004).
- Her, One-Soon. 2012. Distinguishing classifiers and measure words: A mathematical perspective and implications. *Lingua* 122(14). 1668–1691. <https://doi.org/10.1016/j.lingua.2012.08.012>.
- Kilarski, Marcin & Marc Tang. 2018. The coalescence of grammatical gender and numeral classifiers in the general classifier *wota* in Nepali. *Proceedings of the Linguistic Society of America (PLSA)* 3(1). 56:1–10. <https://doi.org/10.3765/plsa.v3i1.4352>.
- Kim, Ronald I. 2014. A tale of two suffixes: \*-h<sub>2</sub>, \*-ih<sub>2</sub>, and the evolution of feminine gender in Indo-European. In Sergio Neri & Roland Schuhmann (eds.), *Studies on the collective and*

- feminine in Indo-European from a diachronic and typological perspective*. 115–136. Leiden: Brill. [https://doi.org/10.1163/9789004264953\\_006](https://doi.org/10.1163/9789004264953_006).
- Luraghi, Silvia. 2011. The origin of the Proto-Indo-European gender system: Typological considerations. *Folia Linguistica* 45(2). 435–463. <https://doi.org/10.1515/flin.2011.016>.
- Pokharel, Madhav P. 2010. Noun class agreement in Nepali. *Kobe Papers in Linguistics* 7. 40–59.
- Sinnemäki, Kaius. 2019. On the distribution and complexity of gender and numeral classifiers. In Francesca Di Garbo, Bruno Olsson & Bernhard Wälchli (eds.), *Grammatical gender and linguistic complexity*. Vol. 2. *World-wide comparative studies*. 133–200. Berlin: Language Science Press. <https://doi.org/10.5281/zenodo.3462782>.
- Trudgill, Peter. 2011. *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford: Oxford University Press.