On the Putative Classifier Languages in Europe: A Case Study of Russian
[論歐洲的疑似量詞語言：俄語的研究]

Chi-Pin Hsu [徐啟斌][1], Vladimir Kurdyumov [耿華][2], and One-soon Her [何萬順][1, 3]

[1] Tunghai University [東海大學] | [2] Moscow City University [莫斯科市立大學中文系] | [3] National Chengchi University [國立政治大學]

In Eurasia, the hotbed of classifier languages as an areal feature is in East Asia, Southeast Asia, part of eastern India, and a less concentrated cluster in the Middle East and Central Asia. In Europe, only a few languages have been claimed to employ numeral classifiers. Russian, located in Eastern Europe, has a tripartite numeral construction [Num X N] where X's are considered sortal classifiers by some researchers, e.g., Sussex (1976) and Goto (2012), and is listed as a classifier language in the World Atlas of Classifier Languages (WACL). However, the status of Russian as a classifier language is controversial. In this paper, we first adopt the definition of classifier languages proposed by Her et al. (2022). We then apply explicit syntactic criteria to evaluate Russian numeral constructions and examine the tripartite [Num X N] structure. Ultimately, our analysis demonstrates that the elements previously identified as sortal classifiers in Russian are in fact measure nouns, and we reject their classification after examining the reasons behind this misidentification. The findings suggest that it would be prudent to reexamine all putative classifier languages in Europe.

Keywords: numeral classifier, sortal classifier, classifier language, Russian

關鍵詞： 量詞、分類詞、分類詞語言、俄語

1. Introduction

Various lexical or morphological strategies are employed in languages to categorize nouns by highlighting certain inherent semantic aspects of the noun. The mass/count distinction is one such aspect. Among such mechanisms, numeral classifiers have received much attention in recent years (e.g., Chierchia 1998, Rothstein 2010, Khrizman 2016).

In this paper, we follow Her et al. (2022) in defining numeral classifiers as grammatical elements that serve as the multiplicand in numeral constructions, where the numeral acts as the multiplier. These classifiers quantify nouns by specifying inherent or measurable properties and form a unified syntactic category (Her et al. 2022, Sera et al. 2023). Before discussing the geographical distribution of numeral classifier languages in the world, it is essential to first establish a definition of what constitutes a numeral classifier and, thus, a classifier language. Similar and related to the count/mass distinction of nouns, numeral classifiers as an independent syntactic category in a classifier language also consist of two subcategories distinguished on semantic grounds: SORTAL CLASSIFIERS (C) and MENSURAL CLASSIFIERS (M). Sortal classifiers (C) are distinct in that their inherent value is numerically fixed at *1*, while mensural classifiers (M) can take values other than *1*, and the value can be fixed or variable[1].

Chinese, along with other East and Southeast Asian languages such as Japanese, Korean, Vietnamese, and Thai, is an example of a language with numeral classifiers, which refer to the unmarked counting or measuring units in a nominal phrase that serve to facilitate the quantification of the noun by a numeral. In Chinese, a canonical numeral classifier construction involves a numeral (Num), a C/M, and a noun (N), hence [Num C/M N]. Examples with sortal classifier and mensural classifier in Chinese are given in (1a) and (1b), respectively.

(1) a. 五 匹/隻 馬
wu *pi/zhi* *ma*
five $C_{pack\text{-}animal}/C_{animal}$ horse
'five horses'

b. 五 對/群 馬
wu *dui/qun* *ma*
five M_{pair}/M_{group} horse
'five pairs/groups of horses'

Example (1) demonstrates the key distinction between sortal and mensural classifiers. The sortal classifier in (1a) inherently denotes a fixed value of *1*, whereas the mensural classifier in (1b) can take values other than *1*, with the quantity either fixed or variable. Moreover, in both cases, the classifier directly quantifies the noun without requiring an adposition, in contrast to the adposition *of* in the English translation of Example (1b), 'five pairs/groups of horses'. However, this does not mean that all classifier-like words that

[1] It should be noted that different studies may employ varying definitions of the concept 'classifier language.' In this study, we adopt the definition and criteria of Her et al. (2022), which explicitly distinguish classifier languages based on the presence of numeral classifiers as a syntactically distinct category. This definition provides a clear structural criterion for differentiating classifiers from other quantificational elements, making it particularly suitable for our analysis.

directly quantify a noun without an adposition are true classifiers. The grammatical system of a given language must be examined to determine whether such classifier-like words constitute a distinct lexical category. We will examine the status of mensural classifiers in Russian in Section 4 and that of sortal classifiers in Section 5.

It has been noted that few European languages have numeral classifiers; some even claim that sortal classifiers are absent in European languages (e.g., Rijkhoff 1998:328). Aikhenvald (2000, 2011) observes that the use of various noun categorization devices ranges from the numeral classifier systems in the languages of East and Southeast Asia to the largely grammaticalized systems of genders in European languages, and to noun classes in African languages. Numeral classifiers thus appear to be largely in complementary distribution with genders and noun classes in different language groups or families (e.g., Allassonnière-Tang et al. 2021). Indeed, few Asian languages with numeral classifiers have grammatical gender agreement, and most languages tend to be conservative in that they rarely adopt more than one strategy (e.g., Aikhenvald 2000, Blench 2012, Allassonnière-Tang et al. 2021).

However, in the vast literature on numeral classifiers and the various grammars on individual languages, and, more importantly, in the recently released database World Atlas of Classifier Languages (WACL), a small number of European languages such as Russian, Polish, and Bulgarian have been claimed to employ numeral classifiers (Her et al. 2022). After all, like genders and noun classes, numeral classifiers are one of the common grammatical devices for noun categorization. It would thus be surprising if they are completely absent in the European languages. On the other hand, caution needs to be exercised when claiming that a language has numeral classifiers, as some scholars have proposed for Russian based on its numeral constructions (cf. Sussex 1976, Goto 2012). We believe, however, that the alleged sortal classifiers in Russian warrant serious reconsideration. See the two Russian examples in (2)[2] with putative sortal classifiers and the two Chinese examples in (3) with sortal classifiers.

(2) a. два человека охраны
dva čeloveka oxrany
two.M.NOM person.M.SG.GEN guard.F.SG.GEN
'two guards'[3]

b. две штуки помидоров
dve štuki pomidorov
two.F.ACC piece.F.SG.GEN tomato.M.PL.GEN
'two pieces of tomatoes'[4]

(3) a. 两 名 守衛
liang ming shouwei
two Chuman guard
'two guards'

[2] The Russian examples in this paper, unless otherwise noted, were provided and confirmed by a native Russian speaker.
[3] Anatoly Rybakov. Heavy Sand (1975-1977) from RNC from Russian National Corpus, hereafter RNC.
[4] https://studopedya.ru/1-3148.html

b. 兩 顆 番茄
liang *ke* *fanqie*
two C_{round} tomato
'two pieces of tomatoes'

This Russian tripartite [Num X N] construction in (2) does seem to resemble the Chinese [Num C/M N] classifier construction in (3). It is thus tempting to view *čeloveka* and *štuki* occupying the position of X as sortal classifiers like *ming* and *kuai* in Chinese. Our goal in this paper is to demonstrate that despite the superficial similarity, these putative sortal classifiers in Russian are in fact nouns and Russian does not have the syntactic category of numeral classifiers at all. Having introduced a working definition of numeral classifiers in this section, we now turn to an examination of their geographical distribution. This step allows us to contextualize Russian within the broader typology of classifier languages before proceeding to a detailed syntactic analysis in later sections.

This paper is organized as follows. In section 2, we first offer an overview of the distribution of classifier languages in the world and show where Russian is situated geographically in the global context. Section 3 then describes and characterizes the syntax and semantics of numeral classifiers as a foundation for the deliberation on Russian as a putative classifier language. In section 4, we examine the numeral constructions in Russian and demonstrate that in the construction [Num X N], if X is a word of measure, it is a noun, not a mensural classifier. We then follow up this argumentation in section 5 and show that the several putative Russian sortal classifiers in in the [Num X N] construction are likewise nouns, not sortal classifiers. Section 6 consists of some concluding remarks.

2. Russian and Classifier Languages in the World

As discussed in Section 1, we adopt the definition of classifier languages from Her et al. distribution. Many languages of East and Southeast Asia employ numeral classifiers; nevertheless, the presence of numeral classifiers has been proposed in some European languages (e.g., Hurford 2003, Dékány & Hegedűs 2021). In Greenberg's (1990) seminal work, three Indo-European languages in Europe are included in the list of 103 classifier languages: Breton, Irish Gaelic, and Bulgarian. Since then, the presence of numeral classifiers has been proposed in some Germanic languages (Sussex & Cubberley 2006), such as German (Sussex & Cubberley 2006:315) and Low German (Zimmermann 2011:226), Slavic languages, e.g., Russian (Sussex 1976), Polish (Sussex & Cubberley 2006), and Bulgarian (Cinque & Krapova 2007), and Celtic languages , e.g., Breton and Irish Gaelic (Greenberg 1990). However, the formal status of the putative classifier construction in some of these languages remains controversial.

According to Gil's (2013) survey in the World Atlas of Language Structures (WALS), while most of the 140 classifier languages recorded cluster in Southeast Asia, the use of classifiers extends eastward into the Pacific and also westward through the Middle East and into Europe. A much more comprehensive database on classifier languages has been released recently. The World Atlas of Classifier Languages (WACL), documented by Her et al. (2022), consists of a survey of 3,338 languages of the world and has identified 723 classifier languages, as shown in Figure 1. The foremost hotbed of classifier languages is

quite visibly in East and Southeast Asia, where the classifier languages overwhelm the non-classifier languages, while in the rest of the world, the scenes are largely the opposite.

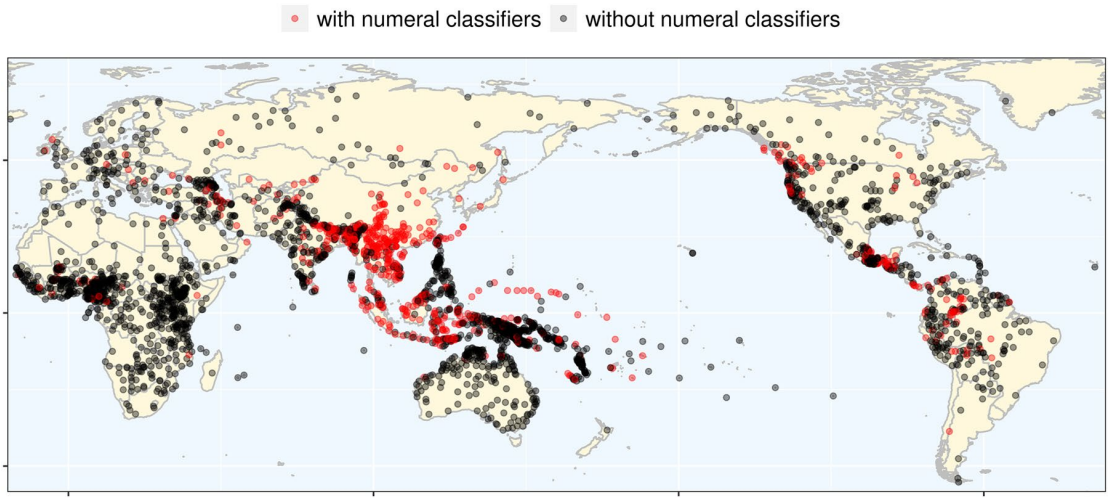

Figure 1. Distribution of 723 classifier languages among the 3,338 languages of the world recorded in WACL (Her et al. 2022: 8)

To offer a clearer view of the global distribution of classifier languages, Figure 2 filters out non-classifier languages and thus better reveals classifier languages identified even in areas overwhelmed by non-classifier languages. Note that Russian is singled out and marked by a red dot on the map and the classifier language immediately underneath is Tatar, a Turkic language.

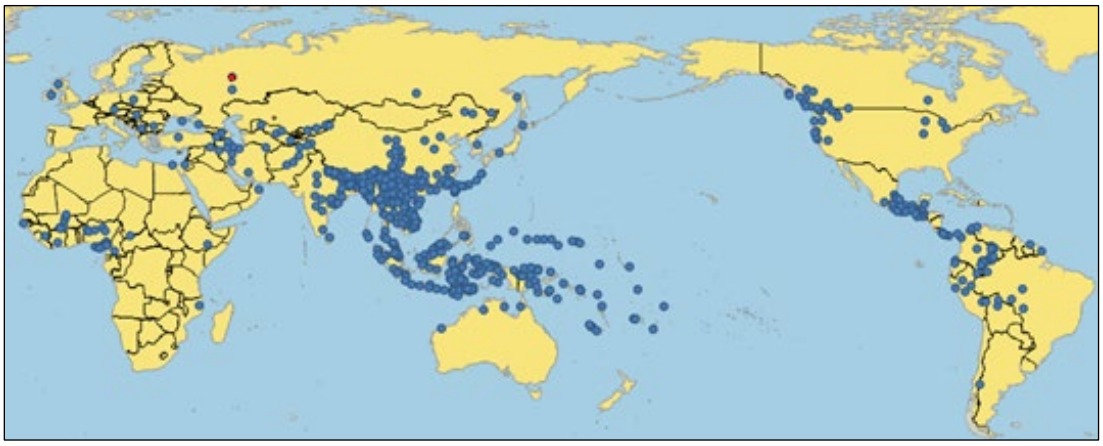

Figure 2. Distribution of the world's 723 classifier languages in WACL
(red dot: Russian, a putative classifier language; right underneath it is Tatar)

Figure 2 also reveals a less concentrated cluster of classifier languages in the Middle East and Central Asia. To account for the apparent diffusion pattern of classifier languages in Asia and the Pacific, Her & Li (2023) propose a hypothesis of a single origin, suggesting that numeral classifiers initially developed indigenously in Sinitic and subsequently spread to neighboring languages in the region. Note that Europe is expressly excluded from this

hypothesis pending further research, precisely due to the uncertainty of the putative classifier languages in Europe. As shown in Figure 3, there are 10 European languages recorded in WACL as classifier languages: Russian, Tatar, Crimean Tatar, Bulgarian, Balkan Romani, Hungarian, Polish, Standard German, Scottish Gaelic, and Irish Gaelic. English and French, however, have been analyzed by WACL as lacking numeral classifiers.

Figure 3. Distribution of the 10 putative European classifier languages in WACL

In the literature, the putative classifiers in these European languages show various properties and evidence of emergence and development different from the classifier languages in other areas of Eurasia (e.g., Emeneau 1956, Barz & Diller 1985, Haspelmath 2001, Hurford 2003, Stilo 2018). Following Her & Li (2023), we take the conservative view that further research is needed on each of these European languages to verify or reject their alleged classifier constructions. The focus of this paper is on Russian.

Originating in Eastern Europe, the Russian language belongs to the Slavic group of the Indo-European family and has spread into an extensive area of Eurasia following the expansion of the Russian regime in recent history. In some literature, a small number of words in Russian have been identified as sortal classifiers (e.g., Sussex 1976, Goto 2012), and Russian is identified accordingly as a classifier language in WACL.

As shown in Figure 2, if Russian is indeed a classifier language, it would be notably the northmost one in the world. However, the remote geographical location alone should not be taken as an indication against Russian as a classifier language, because, after all, there is a classifier language located immediately underneath Russian, i.e., Tatar, a Turkic language. Recent studies, including Hölzl & Cathcart (2019), Allassonnière-Tang et al. (2023), and Chen et al. (to appear), have provided evidence suggesting that some members of the Turkic languages, as well as those in the Mongolic and Tungusic families of the Altaic region, and the Koreanic and Japonic languages of East Asia, exhibit the use of sortal classifiers. This typological distribution is particularly relevant given the presence of classifier languages in the Altaic region, including Tatar, a Turkic language located immediately south of Russian. The fact that classifier languages are attested in this area suggests that the existence of classifiers in Russian should not be dismissed based on geography alone.

Figure 4 from Allassonnière-Tang et al. (2023:301) is a GIS map showing the distribution of 55 Mongolic, Turkic, and Tungusic languages in the Altaic region. Note that 20 are marked as having sortal classifiers. A full list of the 55 languages and their metadata is available in the supplementary materials of Allassonnière-Tang et al. (2023).[5]

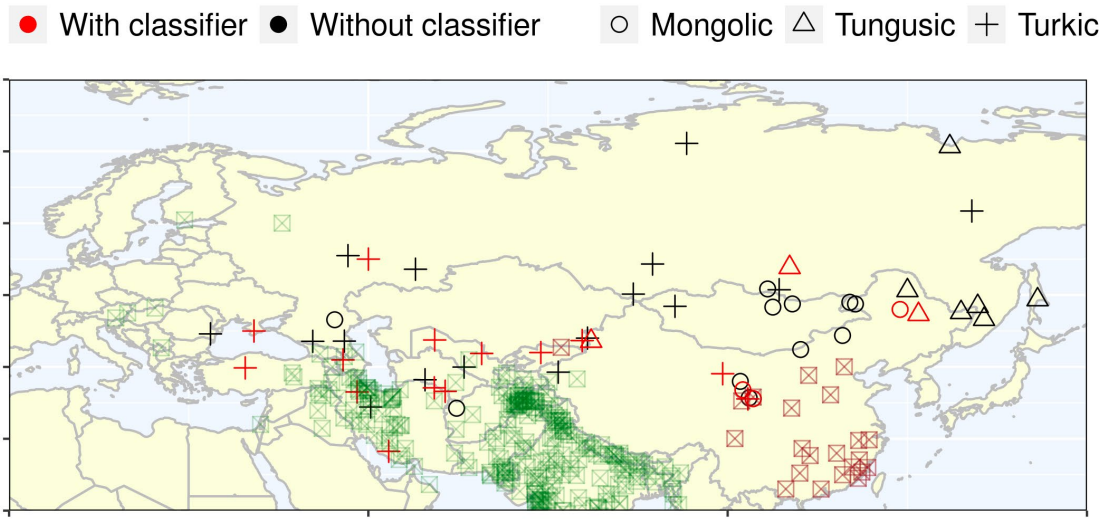

Figure 4. A geographical overview of the 55 Altaic languages. Neighboring languages are shown in shaded colors (Indo-Iranian: green, Sinitic: brown).

As indicated by the + signs in Figure 4, there are 31 Turkic languages in total, and 14 (45%, 14/31) are classifier languages. Tatar, the northmost red + sign, is between Chuvash on the left and Bashkir on the right, both non-classifier languages. The westmost Turkic language is Gagauz, spoken primarily in Moldova, is a non-classifier language, while Crimean Tatar and Turkish on the opposite sides of the Black Sea are both classifier languages. Slavic languages, on the other hand, are primarily located in Europe, and out of the 20 Slavic languages documented in WACL, only three allegedly have sortal classifiers, as indicated by the red dots in Figure 5.[6]

[5] See: https://osf.io/hdur5?view_only=e7bc178f204746c5ad76973f53fb6bca

[6] As pointed out by an anonymous reviewer, while there is research on CL in Russian, there is none on Ukrainian or Belarusian, and the status of Rusyn/Ruthenian as a dialect vs language is contested. These maps should thus be treated critically, but further discussions are beyond the scope of this paper.

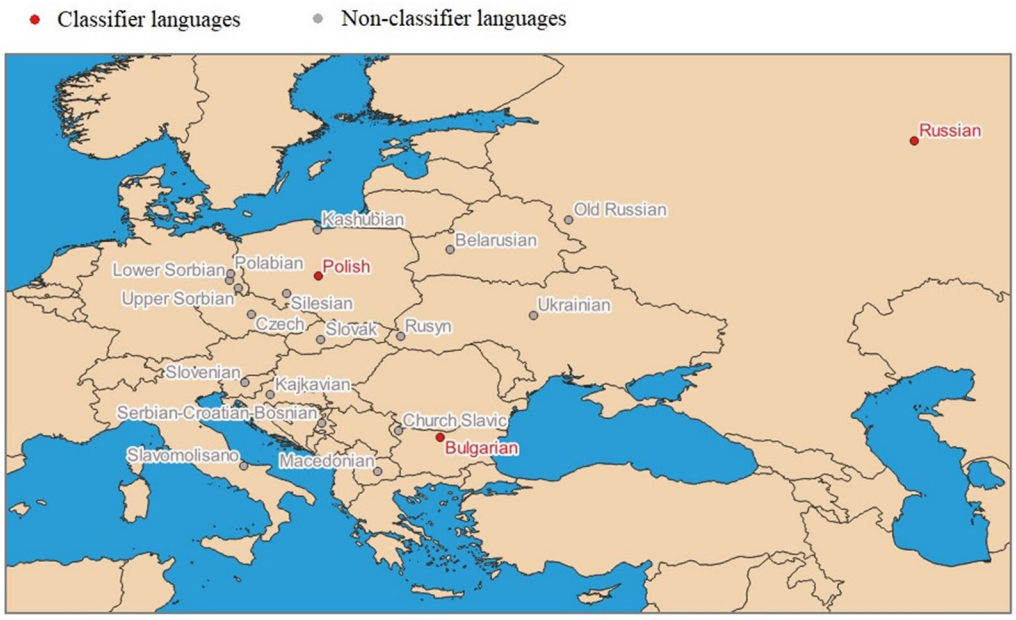

Figure 5. An overview of 20 Slavic languages. Neighboring languages are shown in shaded colors (Putative classifier languages: red, non-classifier languages: grey)

Among the 20 Slavic languages in WACL, Russian, Polish, and Bulgarian have been identified as having putative numeral classifiers. While some Turkic languages have developed classifier-like elements possibly due to language contact (Hölzl & Cathcart 2019), there is no clear evidence that similar processes have influenced Slavic languages, either from non-Slavic languages to Slavic languages or among Slavic languages. In particular, Russian exhibits classifier-like patterns in constructions involving *čelovek* 'person' and *duša* 'soul', but whether these patterns constitute a true classifier system remains an open question. While the WACL database provides a useful starting point for mapping classifier languages, it is important to critically evaluate the criteria used in classifying languages as having or lacking numeral classifiers. For example, including Russian, Polish, and Bulgarian as classifier languages remains controversial (cf. Sussex 1976, Goto 2012). Additionally, as the status of languages such as Ukrainian and Belarusian has not been extensively studied, their classification in WACL should be interpreted with caution. In the following sections, we shall first establish the criteria for numeral classifiers and then reexamine the putative sortal classifiers in Russian accordingly.

3. Numeral Classifiers and the Two Subcategories

Her (2012a) proposed a formal criterion for a dichotomy of C/M, namely into C and M. The proposal is that the relation between C/M and Num is multiplication. The quantity of the head noun is the product of Num and the number that the C/M represents. The distinction between C and M is that C always represents the number *1*, while M represents the number that is not necessarily *1*. Therefore, the number of horses in (1a) is $5\times C_{pack\text{-}animal}$ or $5\times C_{animal}$ and they are both $5\times 1 = 5$. The number of horses in (1b) is then either $5\times M_{pair} = 5\times 2 = 10$ or $5\times M_{herd} = 5$ *herds*. M provides additional information on the

quantity and changes the quantity of the referent. On the other hand, C does not change the number of the referent referred to by the noun phrase. Therefore, C can sometimes be omitted in Mandarin Chinese for prosodic or stylistic reasons or when used with large round numbers, as the following examples illustrate.

(4) 五(張)餅二(條)魚餵飽五千個人
Wu (zhang) bing liang (tiao) yu
Five C_{flat} bread two C_{long} fish
'five loaves of bread and two fish'

All of the three examples in Mandarin Chinese listed in (4) show the redundancy of C in expressing quantities. The distinction between sortal classifiers and mensural classifiers can thus be distinguished by the slightly modified formal rule proposed by Her (2012b) as

(5) C/M distinction in Set-theoretic Terms:
Given two well-formed NumP containing words or suffixes Num, K, and N, X the set of semantic attributes denoted by K, and Y the set of semantic attributes denoted by N, K is C *iff* X⊂Y; otherwise, K is M.

One may then think of the English construction like *five head of cabbage*, where the noun *cabbage* is numerated through the use of *head*, which barely adds extra meaning to the canonical noun phrase with number. One can also simply express the same idea without using head, i.e., *five cabbages*. This is very much like the omission of Chinese classifiers in (4). On the other hand, as for M, one may likewise think *head* as the counting words in the English equivalent in (1)b., i.e., *pairs* and *herds*, or the mensural classifiers like *cups* or *buckets* in *three cups/buckets of water*. From the surface, these words have the same function as the C/M in canonical classifier languages like Chinese. However, it should be noted that syntactically, Chinese C/M is an independent category from nouns. It is positioned after the numeral or the quantifier and before the head noun, a place where other nouns are forbidden; thus *wu zhi/pi/dui/qun* ma (five C_{animal}/$C_{pack\text{-}animal}$/M_{pair}/M_{herd} horse) are grammatical but not **wu qiuchang ma* (five field horse; intended: 'horses that fill five ball fields'). Non-classifier languages of course have words of measure, e.g., box, group, kilo, etc. in English, as shown in (6).

(6) a. *five boxes of chocolate*
b. *five boxes of excellent quality*
c. *five boxes of the kitchen*
d. *five boxes of his*

However, since there is no overt structural position for a C/M word, the putative C/M is actually the head noun of an attributive genitive construction as shown in (6b). As a result, words of measure are nouns instead of C/M words in classifier languages. Consequently, for a word X in the construction [Num X N] to be a genuine classifier, N cannot be restricted to be a genitive attributive.

4. Dismissing Mensural Classifiers in Russian

In this section, Russian numeral constructions with numerals, modifiers, and nouns will be described. The numerals discussed in this section are limited, as certain numeral constructions, such as those involving large numerals like "millions," whose syntactic behavior is already well-documented (Corbett 1978, Timberlake 2004, Hikita 2016) and do not directly contribute to the focus of this article, i.e., whether Russian has numeral classifiers.

Russian declension system, with its two numbers and six cases, is already complex. The interaction between numerals and case inflection further complicates this system. An explanation for the declensional patterns in Russian numeral phrases (NumPs) is found in Corbett (1978), who uses a matrix to describe how numerals like один/odin '1', два/dva '2', три/tri '3', пять/pjatʹ '5', сто/sto '100', тысяча/tysyača '1000', and миллион/million '1000000' interact with case, number, and gender (Corbett 1978:46). The complexity of Russian NumP is shaped by various factors, including countability, the historical loss of the dual number, and contextual influences. These aspects have been examined in detail by scholars such as Mel'čuk (1985), Andersen (2006), Stepanov and Stateva (2018), Letuchiy (2020), and Nesset (2020), among others.

Different numerals show different patterns of agreement with the component words in NumPs, and Corbett (1978, p. 43) invokes the theory of squishies (Ross, 1972) to explain such complexity. The numeral *один*/*odin* '1' agrees with the number, case, gender, and animacy of N in its NumP, which makes *один*/*odin* '1' similar an adjective (Hikita, 2016, p. 345). Applying the theory of squishes, Corbett (1978) comments that the numerals 1, 2, 3, 5, 100, 1000, and 1000000 listed in the matrix of Corbett (1978:46) form a continuum, the larger the numeral in the matrix is, the more "nouniness" the numeral possesses. Based on his cross-linguistic research, Corbett (1978:47) observed that in many languages, the "nouniness" increases with the values of numerals. Hikita (2016:353-362) also discusses the incremental increase of the "nouniness" of the Russian numerals with increasing values in more detail.

The numerals 2, 3, and 4 are paucal numbers, and those from 5 to 9, teens, tens, and hundreds are considered as general numerals (Timberlake 2004). In the interest of our discussion of putative Russian classifiers, we limit our description of the rules of the numbers and cases in N's in numeral constructions from 1 to 30. The following is an algorithm of the number and case in a numeral construction adopted from Timberlake (2004:189) and Hikita (2016):

(7) a. As aforementioned, *один*/*odin* '1' agrees with the number, case, gender, and animacy of N in its NumP, which makes один/odin '1' similar to an adjective (Hikita 2016:345).

b. When the numeral phrase is in an oblique case, the numeral, modifiers, and noun all appear in the oblique case, and modifiers and the noun are in the plural form.

c. When the numeral phrase is in the nominative or accusative case:
c_1: If the numeral is a paucal:

$c_{1.1}$: If the numeral phrase is in the accusative case and the quantified noun is animate, the numeral takes the genitive case, and modifiers and the noun are in the genitive plural form.

$c_{1.2}$: In other contexts, the numeral typically appears in the nominative=accusative form. The accompanying noun, however, is rendered in the genitive singular. Modifiers generally take on plural forms, but there is an exception for feminine nouns in non-existential contexts, where modifiers align with the numeral and appear in the nominative=accusative case. In all other cases, the modifiers adopt the genitive form.

c_2: For general numerals, the numeral remains in the nominative=accusative form, irrespective of animacy, and the noun and modifiers are consistently in the genitive plural form.

d. For complex numerals, Timberlake (2004:191-194) makes a rather long remark. highlights a distinction in complex numerals between formal and informal registers. In the formal written register, all elements of a complex numeral phrase (NumP), including the noun and modifiers, match the case and number of the phrase. In the oblique case, they are plural; in the direct case, they align with the direct case. The final numeral determines the noun's form, that is, it is not a paucal, the noun and modifiers are genitive plural; otherwise, the noun is genitive singular, while adjectives are plural and vary based on gender. Paucals do not inherently mark animacy, but when they do, nouns and modifiers take the genitive plural. For numerals ending in *один/odin* '1', the noun remains singular, and the numeral agrees in number. In informal contexts, the declension of complex numerals is often simplified: only peripheral elements may decline, while central elements remain undeclined.

From (7), it can be seen that X in [Num X N] has complicated number marking. However, it should be noted that, as pointed out by Dayal (2004), this does not guarantee that expressing kind or individualization can choose only one of the number marking methods available in a given language. Consequently, a language with number marking may still resort to other linguistic mechanisms to express individualization.

Having described the rules of the numbers and cases in N's in numeral constructions from 1 to 30, we are now in a better position to examine whether putative classifiers X's in a tripartite NumP constructions [Num X N] such as (10) are real classifiers.

(8)

три	ящика	яблок
tri	*jaščika*	*jablok*
three.NOM	box.M.SG.GEN	apple.N.PL.GEN

'three boxes of apples'

(9)

три	ноги	гусениц
tri	*nogi*	*gusenic*

three.NOM leg.F.SG.GEN caterpillar.F.PL.GEN
'three legs of caterpillars'

(10) три ноги гусеницы
tri *nogi* *gusenicy*
three.NOM leg.F.SG.GEN caterpillar.F.SG.GEN
'three legs of the caterpillar'

Jaščika 'box' in (8) is a nominal measure word for the quantity of apples. Since Num is three, *jaščika* 'box' is in its genitive singular form, and by (7), it means that *jaščika* 'box' is not a modifier but rather a noun. The word *jablok* 'apples' is a postponed attributive, it is in its genitive plural form. Hence, the internal structure of (8) is [[*tri jaščika*] *jablok*], or [[three box] apples].

Nogi 'leg' in (9) and (10) is not a measure word for the quantity of caterpillars. However, since Num is three, *nogi* 'leg' is in its genitive singular form as described by (9). Since *gusenic* 'caterpillars' is a postponed attributive, it is in its genitive plural form. Hence, the internal structure of (9) is [[three legs] (of) caterpillars].

Semantically, *jaščika* 'box' in (8) is a measure word, and *nogi 'leg'* is not. However, syntactically, the component words of (9) and (10) have similar grammatical numbers. In addition, *gusenicy* 'caterpillar' in (10) is in its genitive singular form, and gusenic 'caterpillars' in (9) is in its genitive plural form. This shows that the grammatical number of N in [Num X N] in (9) and (10) is decided semantically rather than decided by Num. It is a piece of evidence that in (8), N in a tripartite construction [Num X N] in is a postponed attributive, and X is a measure noun rather than a mensural classifier.

5. Putative Russian Sortal Classifiers as Nouns

Several studies claim that Russian is a classifier language and list *человек*/*čelovek* 'person', *штука*/*štuka* 'piece', and *голова*/*golova* 'head' as examples of Russian sortal classifiers (Sussex, 1976; Comtet, 1993; Goto 2012; Paperno 2012), and the putative classifiers in the examples given often occur in the place of X in the [Num X N] construction. Following Goto (2012, p.4) *čelovek* 'person', *štuka* 'piece', *edinica* 'unit', *duša* 'soul', and *golova* 'head' are identified as classifiers. To further illustrate their usage, we use examples (11)–(15) to demonstrate their behavior in the [Num X N] constructions.

(11) a. два человека охраны
dva *čeloveka* *oxrany*
two.M.NOM person.M.SG.GEN guard.F.SG.GEN
'two guards'[7]

b. шесть человек охраны
šest' *čelovek* *oxrany*
six.NOM person.M.PL.GEN guard.F.SG.GEN
'six guards'

[7] Anatoly Rybakov. Heavy Sand (1975-1977) from RNC.

(12) a. две штуки помидоров
dve *štuki* *pomidorov*
two.F.ACC piece.F.SG.GEN tomato.M.PL.GEN
'two pieces of tomatoes'[8]

b. шесть штук помидоров
šest' *štuk* *pomidorov*
six.ACC piece.F.PL.GEN tomato.M.PL.GEN
'six pieces of tomatoes'[9]

(13) a. две единицы товара
dve *edinicy* *tovara*
two.F.ACC unit.F.SG.GEN goods.M.SG.GEN
'two units of goods' (very formal)[10]

b. шесть единиц товара
šest' *edinic* *tovara*
six.ACC unit.F.PL.GEN goods.M.SG.GEN
'six items of goods' (very formal)[11]

(14) a. две души крестьян
dve *duši* *krest'jan*
two.F.NOM soul.F.SG.GEN peasant.M.PL.GEN
'two peasants'

b. шесть душ крестьян
šest' *duš* *krest'jan*
six.NOM soul.F.PL.GEN peasant.M.PL.GEN
'six peasants'

(15) a. две головы ослов
dve *golovy* *oslov*
two.F.NOM head.F.SG.GEN donkey.M.PL.GEN
'two donkeys' (very formal)[12]

b. шесть голов ослов
šest' *golov* *oslov*
six.NOM head.F.PL.GEN donkey.M.PL.GEN
'six donkeys' (very formal)

Note that Num's in (11)–(15) are all *two*, and X's in the tripartite constructions [Num X N] are all in their genitive singular forms. In addition, the genders of *two* and X's are in agreement. However, the genders and numbers between Num's and N's in (11)–(15) do not follow (7). This suggests that the internal structure of (12a) *[dve štuki pomidorov]* '[two pieces of tomatoes]' is more like *[[dve štuki] pomidorov]* '[[two pieces] of tomatoes]'

[8] https://studopedya.ru/1-3148.html
[9] https://u-f.ru/Article/u202/2012/12/15/649071
[10] https://moskvichmag.ru/gorod/на-какие-продукты-мы-тратили-больше-вс/
[11] https://u-f.ru/Article/u202/2012/12/15/649071
[12] https://uprveter32.ru/index.php/podvedomstvennye-organizatsii-2/2312-leptospiroz-leptospirosis

rather than *[dve [štuki pomidorov]]* '[two [pieces of tomatoes]]'. Thus, it follows that N's in (11)–(15) are postponed attributives. That is, the internal structure of the behavior of putative sortal classifiers also holds for complex numbers like what is shown in (16).

(16) двадцать две единицы принтеров
dvadcat' *dve* *edinicy* *printerov*
twenty.NOM two.F.NOM unit.F.SG.GEN printer.M.PL.GEN
'twenty-two printers' (very formal)

It can be observed that the tripartite construction [Num X N] in (16) follows a similar pattern in (13). Furthermore, to express approximate numbers of items, there is a structure called approximative inversion. The usual word order in a NumP in Russian is [Num N] as in (17). However, in the construction of approximative inversion, the word order becomes [N Num] as in (18).

(17) *pjat'* *knig*
five.NOM book.F.PL.GEN
'five books'

(18) *knig* *pjat'*
book.F.PL.GEN five.NOM
'about five books' (Khrizman 2016: 31)

Goto (2012, p. 23) observed that word orders involving what she referred to as classifiers preceding numerals convey the meaning of approximate quantity. She represented these structures as [[Cl Num]N] and [N[Cl Num]].[13] An illustrative example provided by Goto (2012:16) is reproduced below as (19). Compared with (18), (19) provides further evidence that elements traditionally regarded as sortal classifiers are, in fact, nouns.

(19) *V* *jaščike* *bylo* *štuk* *40* *xorošix* *jablok.*
in box were item.F.PL.GEN. 40 good.PL.GEN apple.N.PL.GEN
'There were around 40 good apples in a box.'[14]

By (7) and examples (11)–(15), X's taken as classifiers in the tripartite construction [Num X N] should be nouns. That is, the putative sortal classifiers given by Goto (2012) should be nouns. N's in examples (11)–(16) are actually postponed attributives. Consequently, we refute that words like *čelovek* 'person', *štuka* 'piece', *edinica* 'unit', *duša* 'soul', and *golova* 'head' are sortal classifiers. Instead, while these words functionally act as measure words (Khrizman 2016) associated with specific semantic

[13] P. N. Filonov. Diaries (1930-1939) from RNC.

[14] Whether in the constituency of the Russian [Num X N] construction is [[Num X] N] or [Num [X N]] does not affect our analysis that X is a noun, not a classifier. However, for the question *Skol'ko jablok ty kupil?* 'How many apples did you buy?', the answer *tri jaščika* 'three boxes' seems to favor [Num X]. For the question *Ty kupil te tri jaščika jablok?* 'Did you buy those three boxes of apples?', the anaphoric response *Da, ja kupil te jabloki* 'Yes, I bought those apples' also seems to favor *tri jaščika* as a constituent. However, a further in-depth investigation is needed to settle the issue.

features (e.g., *čelovek* for [human]), their syntactic behavior aligns with that of nouns, suggesting that they should be classified as nouns. Previous studies that classified Russian as a classifier language (e.g., Sussex 1976, Goto 2012) primarily relied on the presence of words such as *čelovek* 'person' and štuka *'piece'* in [Num X N] constructions. However, these analyses did not account for the fact that X in these constructions exhibits clear nominal properties, functioning syntactically as nouns rather than classifiers. As demonstrated in this study, X is not a grammaticalized classifier but rather a measure noun within an attributive genitive structure. This oversight has led to a misclassification of Russian.

6. Conclusions

The World Atlas of Classifier Languages (WACL), documented by Her et al. (2022), is an open-access database where among the 3,338 languages surveyed, 723 are identified as classifier languages. This is the largest database of numeral classifier languages since Gil (2013) and is a significant contribution to the study of classifiers. However, it is inevitable that in such large-scale databases there are controversial cases and even errors. Recognizing this, Her et al. (2022: 11) explicitly 'welcome comments and suggestions from the linguistic community to correct and/or expand the content of WACL.' It is in the spirit of recognizing the achievement of WACL and contributing to its improvement, that we have conducted this study to reexamine the putative classifiers in Russian and thus to reevaluate its status as a classifier language in WACL.

We have thus first situated Russian as a putative numeral classifier language in the context of the global distribution of numeral classifier languages and characterized the syntax and semantics of numeral classifiers as a foundation for the deliberation of the five putative Russian sortal classifiers as X in the tripartite construction [Num X N], i.e., *čelovek* 'person', *štuka* 'piece', *edinica* 'unit', *duša* 'soul', and *golova* 'head'.

As typical numeral classifier languages employ both sortal and mensural classifiers, we first demonstrated that there are no mensural classifiers in Russian. We then argued that the five X's are nouns quantified by Num with N serving as a postponed attributive. Note that X can be in the plural form, as in the naturally occurring example (20), and have two possible functions, either referring to kinds or predicates denoting sets of individuals (Dayal 2004, Khrizman 2016, p. 18). Either way, X is a noun, not a sortal classifier.

(20) …12 человек студентов…
…12 čelovek studentov…
…12 person.PL.GEN
'…12 students…'[15]

The reason that *čelovek* 'person', *štuka* 'piece', *edinica* 'unit', *duša* 'soul', and *golova* 'head' are treated as sortal classifiers in Goto (2012) and earlier works (e.g., Sussex 1976) is likely because while X's are nouns quantified by Num's, N's, as postponed attributives, provide information about the referent, and the latter are semantically heavier than X's. This may have given the false impression that X's are classifiers. In conclusion, Russian is

[15] https://www.muiv.ru/nnov/about/news/igra-po-izbiratelnomu-zakonoda/

not a classifier language. While the current study focuses on Russian, similar classification issues may exist for other European languages listed in WACL. Further research is needed to reassess these cases, but such an investigation is beyond the scope of this paper. The findings in this paper about Russian indicate that it would be prudent to reexamine all putative classifier languages in Europe listed in the WACL database.

Funding

This research was supported by the National Science and Technology Council, Taiwan, R.O.C., under Grant No. 108-2410-H-029-062-MY3 and 111-2410-H-029-009-MY3, awarded to O.-S. Her.

Acknowledgments

The authors would like to thank the editors and anonymous reviewers for their constructive comments and suggestions, which helped to improve the paper greatly.

List of abbreviations

ACC	accusative	N	noun
C	sortal classifier	n	neutral
Cl	classifier	NOM	nominative
f	feminine	Num	numeral
GEN	genitive	NumP	numeral phrase
M	mensural classifier	PL	plural
m	masculine	SG	singular

References

Aikhenvald, Alexandra Y. 2000. *Classifiers: A Typology of Noun Categorization Devices*. Oxford: Oxford University Press.

Alexandra Y. Aikhenvald and David Green. 2011. Palikur and the Typology of Classifiers. *Language at Large: Essays on Syntax and Semantics*, ed. by Alexandra Y. Aikhenvald and R. M. W. Dixon, 394-450. Leiden: Brill.

Allassonnière-Tang, Marc, Olof Lundgren, Maja Robbers, Sandra Cronhamn, Filip Larsson, One-Soon Her, Harald Hammarström, and Gerd Carling. 2021. Expansion by migration and diffusion by contact is a source to the global diversity of linguistic nominal categorization systems. *Humanities and Social Sciences Communications* 8:331. Retrieved April 7, 2025, from https://doi.org/10.1057/s41599-021-01003-5

Allassonnière-Tang, Marc, Gao Zhong-liang, Chen Shen-An, and Her One-soon. 2023. Phylogenetic analyses for the origin of numeral classifiers in Mongolic, Tungusic, and Turkic languages. *Concentric: Studies in Linguistics* 49.2:295-315.

Andersen, Henning. 2006. SomeThoughts on the History of Russian Numeral Syntax. *Harvard Ukrainian Studies* 28.1-4:57-67.

Barz, R. K. and A. V. N. Diller. 1985. Classifiers and standardisation: Some South and South-East Asian comparison. *Papers in South-East Asian linguistics No.9:*

Language policy, language planning and sociolinguistics in South-East Asia, ed. by David Bradley, 155-184. Canberra: Pacific Linguistics.

Blench, Roger. 2012. The role of agriculture in explaining the diversity of Amerindian languages. *The past ahead: language, culture and identity in the Neotropics,* ed. by C. Icendahl, 13-37. Uppsala, Sweden: Acta Universitatis Upsaliensis.

Chao, Yuen Ren. 1968. *A Grammar of Spoken Chinese* Berkeley, CA: University of California Press.

Chen, Shen-An, Marc Allassonnière-Tang, Yung-Ping Liang, and One-soon Her. (to appear). On the distribution and origin of sortal classifiers in Altaic languages. *Journal of Chinese Linguistics*.

Chierchia, Gennaro. 1998. Reference to Kinds across Language. *Natural Language Semantics* 6:339-405. Retrieved April 7, 2025, from https://doi.org/10.1023/A:1008324218506

Cinque, Guglielmo and Iliyana Krapova. 2007. A note on Bulgarian numeral classifiers. *Pitar Moş: A Building With a View. Papers in Honour of Alexandra Cornilescu,* ed. by G. Alboiu, A. A. Avram, L. Avram, and D. Isac, 45-51. Bucharest: Editura Universităţii din Bucureşti.

Comtet, Roger. 1993. L'emploi de *čelovek* comme proforme en russe contemporain [The employment of *čelovek* as pro-form in contemporary Russian]. *Revue des Études Slaves Année* 65.1:73-80.

Corbett, Greville G. 1978. Numerous Squishes and Squishy Numerals in Slavonic. *International Review of Slavic Linguistics* 3:43-73.

Dayal, Veneeta. 2004. Number marking and (in)Definiteness in Kind Terms. *Linguistics and Philosophy* 27:393-450.

Dékány, Éva and Veronika Hegedűs. 2021. Postpositions: formal and semantic classification. *Syntax of Hungarian: Postpositions and Postpositional Phrases,* ed. by Katalin É. Kiss and Veronika Hegedűs, 11-191. Amsterdam: Amsterdam University Press.

Dickey, Stephen McCartney and Laura A. Janda. 2015. Slavic aspectual prefixes and numeral classifiers: Two kinds of lexico-grammatical unitizers. *Lingua* 168:57-84.

Emeneau, Murray Barson. 1956. India as a linguistic area. *Language*, 32.1:3-16.

Gil, David. 2013. Numeral classifiers. *The World Atlas of Language Structures Online,* ed. by Matthew S. Dryer and Martin Maspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved April 7, 2025 from https://wals.info/feature/55A#2/28.0/149.9

Goto, Ksenia. 2012. On Russian Numeral Classifiers. *論叢現代語・現代文化 [Collection of essay of modern language and culture]*, 8, 13-27.

Greenberg, Joseph Harold. 1990. Numeral classifiers and substantival number: Problems in the genesis of a linguistic type. *On language: Selected writings of Joseph H. Greenberg,* ed. by Keith Denning and Suzanne Kemmer, 166-193. Stanford: Stanford University Press.

Haspelmath, Martin. 2001. The European linguistic area: Standard Average European. *Language Typology and Language Universals,* ed. by Martin Haspelmath, 1492-1510. Berlin: de Gruyter.

Her, One-Soon. 2012. Distinguishing classifiers and measure words: A mathematical perspective and implications. *Lingua* 122.14:1668-1691.

Her, One-Soon. 2012. Structure of classifiers and measure words: A lexical functional account. *Language and Linguistics* 13.6:1211-1251.

Her, One-Soon, Harald Hammarström, and Marc Allassonnière-Tang. 2022. Defining numeral classifiers and identifying classifier languages of the world. *Linguistics Vanguard* 8.1:151-164. Retrieved April 7, 2025, from https://wacl.clld.org/

Her, One-Soon and Hui-Chin Tsai. 2020. Left is right, right is not: On the constituency of the classifier phrase in Chinese. *Language and Linguistics* 21.1:1-32.

Her, One-Soon and Li Bing-Tsiong. 2023. A single origin of numeral classifiers in Asia and the Pacific: A hypothesis. *Nominal Classification in Asia and Oceania: Functional and Diachronic Perspectives,* ed. by Marc Allassonnière-Tang and Marcin Kilarski, 113-160. Amsterdam: John Benjamins.

Hikita, Go. 2018. *俄語文法大全[Complete Russian Grammar]* (Oriniral work published 2016 and translated by Huang Q.). New Taipei: Taiwan Mansion International Group.

Hurford, Jim R. 2003. The interaction between numerals and nouns. *Noun Phrase Structure in the Languages of Europe,* ed. by Frans Plank, 561-620. Berlin: Mouton.

Hölzl, Andreas and Chundra Cathcart. 2019. Sortal numeral classifiers in Central Asia. Paper presented at the 13th Conference of the Association for Linguistic Typology, University of Pavia, Pavia.

Khrizman, Keren. 2016. Functional Unit Classifiers in (Non)-Classifier Russian. *Baltic International Yearbook of Cognition, Logic and Communication vol. 11: Number: Cognitive Semantics and Crosslinguistic Approaches*, ed. by Susan Rothstein and Jurģis Šķilters. Retrieved April 7, 2025, from https://doi.org/10.4148/1944-3676.1115

Kurdyumov, Vladimir A. 2005. *Курс китайского яазыка. Теоретическая грамматика. [漢語理論語法教程(耿氏語法)][Chinese language course. Theoretical grammar.]*. Moscow: Citadel-Trade & Lada/Veche Publishing.

Lewis, Paul M., Gary F. Simons, and Charles D. Fennig. 2009. *Ethnologue: Languages of the World* (16th ed.). Dallas, TX: SIL International.

Letuchiy, Alexander B. 2020. Russian constructions with numerals and quantifiers: Their non-standard use in the subject and object position. *Rhema* 1:70-89. Retrieved April 7, 2025, from https://doi.org/10.31862/2500-2953-2020-1-70-89

Mel'čuk, Igor A. 1985. *Поверхностный синтаксис русских числовых выражений [Surface syntax of Russian numeral expressions]*. Vienna, Institut für Slawistik der Universität Wien.

Nesset, Tore 2020. A long birth: The development of gender-specific paucal constructions in Russian. *Diachronica* 37.1-4:1-26. Retrieved April 7, 2025, from https://doi.org/10.1075/dia.18057.nes

Paperno, Denis. 2012. Qiantification in Standard Russian. *Handbook of Quantifiers in Natural Language,* ed. by Edward L. Keenan and Denis Paperno, 729-779. Dortrecht, Netherlands: Springer.

Rijkhoff, Jan. 1998. Order in the noun phrase of the languages of Europe. *Constituent order in the languages of Europe*, ed. by Anna Siewierska, 321-382. Berlin: Mouton de Gruyter.

Ross, John Robert. 1972. The Category Squish: Endstation Hauptwort. *Papers from the Eighth Regional Meeting, Chicago Linguistic Society,* ed. by Paul. M. Peranteau, Judith. N. Levi, and Gloria. C. Phares, 316-328. Chicago, IL: Chicago Linguistic Society.

Rothstein, Susan. 2010. Counting and the Mass/Count Distinction. *Journal of Semantics* 27.3:343-397. Retrieved April 7, 2025 from https://doi.org/10.1093/jos/ffq007

Stepanov, Arthur and Penka Stateva. 2018. Countability, agreement and the loss of the dual in Russian. *Journal of Linguistics* 54.4:779-821. Retrieved April 7, 2025, from https://doi.org/10.1017/S0022226718000130

Stilo, Don. 2018. Numeral classifier systems in the Araxes-Iran linguistic area. *The Diachrony of Classification Systems,* ed. by William B. McGregor and Søren Wichmann, 135-164. Amsterdam: John Benjamins.

Sera, Maria D., Hooi Ling Soh, Drew Brinker., Jenny Yi-Chun Kuo, Judith W. Fuller, Christopher Hammerly, James Stevens, Ruxue Shao, Chris Batteen., Dingcheng Li, Nobuko Davis, Wenting Cai. 2023. A cross-cultural study of language and cognition: Numeral classifiers and solid object categorization, *Memory & Cognition* 51:601-662. Retrieved April 7, 2025, from https://doi.org/10.3758/s13421-022-01376-1

Sussex, Roland. 1976. The Numeral Classifiers of Russian. *Russian Linguistics* 3.2:145-155.

Sussex, Roland and Paul Cubberley. 2006. *The Slavic languages*. Cambridge: Cambridge University Press.

Timberlake, Alan. 2004. *A Reference Grammar of Russian.* Cambridge: Cambridge University Press.

Zhang, Niia Ning. 2011. The constituency of classifier constructions in Mandarin Chinese. *Taiwan Journal of Linguistics* 9.1:1-50.

Zimmermann, Malte. 2011. On the functional architecture of DP and the feature content of pronominal quantifiers in Low German. *The Journal of Comparative Germanic Linguistics* 14.3:203-240.

Address for correspondence

One-Soon Her
Department of Foreign Languages and Literature
Tunghai University
Taichung, TAIWAN
hero@thu.edu.tw
Graduate Institute of Linguistics
National Chengchi University
Taipei, Taiwan
onesoon@gmail.com

Co-author information

Chi-Pin Hsu

Anthony Chi-Pin Hsu [徐啟斌], Vladimir Kurdyumov [耿華], and One-Soon Her [何萬順]

Department of Foreign Languages and Literature
Tunghai University
Taichung, TAIWAN
hgacph2014@gmail.com

Vladimir Kurdyumov
Department of Chinese Language
Moscow City University
Moscow, Russia
vkplans@gmail.com