Data & Analysis	DISCUSSION	

Do number neutral languages tend to have numeral classifiers instead?

JI Xiang

June 14, 2017

◆ロト ◆昼 ト ◆ 臣 ト ◆ 臣 ・ 夕 � @

INTRODUCTION Basic Concepts Hypothesis & Literature

DATA & ANALYSIS Overview WALS Autotyp

DISCUSSION Summary Potential Reasons

References

NUMERAL CLASSIFIERS (GIL 2013, WALS)

Countability:

- ► High countability: woman, dog, pencil
- ► Low countability: water, sand, smoke

Mensural numeral classifiers (common across languages):

- ► one glass of water
- two pounds of sand
- three plumes of smoke

Sortal numeral classifiers (uncommon in some languages):

- ▶ 一头牛 (one-CLF ox)
- ▶ 一条鱼 (one-CLF fish)
- ▶ 一匹马 (one-CLF horse)

Generally, "numeral classifiers" in relevant discussions exclusively refers to sortal numeral classifiers.

NOMINAL PLURALITYHASPELMATH 2013, WALS

The extent to which **plural markers** on full nouns (non-pronouns) are used in a language.

Two dimensions:

Animacy: Animate (especially human) vs. inanimate nouns

Taba:

- wang/wang=si 'child/children'
- Ilu 'leaf/leaves'
- Obligatoriness: Non/Optional/Obligatory occurrence Yoruba (No marker used if the referent is thought of collectively):
 - àwon ìwé mi [PL book my] 'my (various) books'
 - ìwé mi [book my] 'my (collection of) books' or 'my book')

▲ロト ▲ 理 ト ▲ 王 ト ▲ 王 - の Q (~

HYPOTHESIS & LITERATURE

Number neutral languages tend to have numeral classifiers. A well-known hypothesis/observation (observed by multiple authors): [Gil 1987; Thompson 1965; Link 1991; Shopen 2007]

Example: In (contemporary) Mandarin Chinese, nouns don't have plural markers, but require numeral classifiers.

- ▶ 一匹马 (one-CLF horse)
- ▶ 两匹马 (two-CLF horses)
- ▶ 两马*

000

OVERVIEW

- ► Searched through WALS, URIEL (Ethnologue), Autotyp.
- Relatively few relevant data, especially on nominal plurality (only found in WALS).
 "A large number of grammatical descriptions do not say whether plural marking is obligatory or optional" – WALS
- ► Autotyp: Data on *the number of* numeral classifiers in each language.

DISCUSSION REFERENCES R OCOORDOCOCOCOCOCOCOCO DISTRIBUTION OF NUMERAL CLASSIFIERS (GIL 2013,

WALS)



 INTRODUCTION
 DATA & ANALYSIS
 DISCUSSION
 REFERENCES
 References

 DISTRIBUTION
 OF
 NUMERAL
 CLASSIFIERS (GIL 2013,

► Mainly concentrated in East/Southeast Asia.

WALS)

 To the west: Peters out across the South Asian subcontinent. Optional again in western Asia, around Iran and Turkey.

< □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □

► To the east: Tapers out in New Caledonia & western Polynesia.

INTRODUCTION DATA & ANALYSIS DISCUSSION REFERENCES

DISTRIBUTION OF NOMINAL PLURALITY HASPELMATH 2013, WALS



<ロト < 部 > < 語 > < 語 > < 語 > < 語 < の < 0</p>

- Optional plural marking is common in Southeast and East Asia.
- Complete lack of plural marking is particularly found in New Guinea and Australia.

INTRODUCTION

ANALYSIS OF WALS DATA

- The correlation between the two maps seems quite apparent.
- Raw combination of the two sets of WALS data (114 languages):

						\$	÷	Numeral Classifiers / Occurrence of Nominal Plurality _v	Number of languages 🔅
						+	0	Optional / Only human nouns, optional	1
								Optional / Only human nouns, obligatory	0
						+	٠	Optional / No nominal plural	2
								Optional / All nouns, optional in inanimates	0
						+	•	Optional / All nouns, always optional	5
C/Dl_{r}	1	2	2	4	F	+		Optional / All nouns, always obligatory	3
C/Plu	T	Ζ	3	4	5	+		Obligatory / Only human nouns, optional	2
	8	8	12	12	8	+		Obligatory / Only human nouns, obligatory	2
	õ	1	~	_	0	+		Obligatory / No nominal plural	2
	2	T	0	5	0	+		Obligatory / All nouns, optional in inanimates	1
	2	2	2	6	1	+		Obligatory / All nouns, always optional	6
	-	-	-	0	1	+		Obligatory / All nouns, always obligatory	2
						+	٠	Absent / Only human nouns, optional	8
						+	٠	Absent / Only human nouns, obligatory	12
						+	٠	Absent / No nominal plural	8
						+	0	Absent / All nouns, optional in inanimates	8
						+		Absent / All nouns, always optional	12

Absent / All nouns, always obligatory

40

INTRODUCTION DATA & ANALYSIS DISCUSSION REFERENCES References

ANALYSIS OF WALS DATA

Simplified categories:

- ► Has numeral classifier (including optional)
- Number neutral in some way(i.e. doesn't have obligatory plural marker for all nouns)



```
DATA & ANALYSIS
```

ANALYSIS OF WALS DATA

Statistical tests show significant p-values: Pearson's Chi-squared test with Yates' continuity correction

```
data: tabulation
X-squared = 4.7314, df = 1, p-value = 0.02962
___
Fisher's Exact Test for Count Data
data: tabulation
p-value = 0.02167
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.07792435 0.88150327
sample estimates:
odds ratio
                                       <ロト < 母 ト < 豆 ト < 豆 ト < 豆 - の < 0</p>
```

0.2886215



I tried out three other different dimensions of data grouping:

- Numeral classifiers: Re-split along the dimension obligatory-optional (i.e. prevent *optional* numeral classifiers from being regarded as TRUE).
- Nominal plurality: Split along the dimension animate-inanimate (i.e. distinguish the languages where only animate nouns have plural markers but inanimate nouns don't).
- Nominal plurality: Split along the dimension obligatory-optional (i.e. distinguish the languages where plural markers are *obligatory* for at least some nouns with the languages where plural markers *are never obligatory*).



Only consider those languages where numeral classifiers are **obligatory**, i.e. exclude optionality. The p-value is still quite significant.



DATA & ANALYSIS	DISCUSSION	
00000000000000000000000000000000000000		

Pearson's Chi-squared test with Yates' continuity correction

```
data: t2
X-squared = 3.7605, df = 1, p-value = 0.05248
___
Fisher's Exact Test for Count Data
data: t2
p-value = 0.04419
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
  1.03243 47.34430
sample estimates:
odds ratio
  4.933731
                                       < ロ > < 同 > < 三 > < 三 > 、 三 < の < 0</p>
```



- Surprisingly, the split between animate/inanimate resulted in near homogeneity of the data.
- Potential explanation: Most animate words are countable, and the numeral classifiers are first applied to them anyways. i.e. the properties of inanimate words are quite irrelevant in predicting numeral classifiers.



!▶ ◀ 差 ▶ 差 の � @

	DATA & ANALYSIS	DISCUSSION	
	000000000000000000000000000000000000000		

Pearson's Chi-squared test with Yates' continuity correction

```
data: table
X-squared = 0.00085694, df = 1, p-value = 0.9766
___
Fisher's Exact Test for Count Data
data: table
p-value = 0.8143
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.3936323 3.0969354
sample estimates:
odds ratio
  1.133178
                                       < ロ > < 同 > < 三 > < 三 > 、 三 < の < 0</p>
```



- However, the split between obligatory/non-obligatory plural markers resulted in highly significant p-values.
- It strongly suggests that if a language doesn't have obligatory plural markers on any noun, it is very likely to have numeral classifiers.



●▶ 目 ∽へで

DATA & ANALYSIS	DISCUSSION	
000000000000000000000000000000000000000		

Pearson's Chi-squared test with Yates' continuity correction

```
data: table
X-squared = 10.169, df = 1, p-value = 0.001428
___
Fisher's Exact Test for Count Data
data: table
p-value = 0.001159
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
  1.719316 14.255287
sample estimates:
odds ratio
  4.748625
                                       < ロ > < 同 > < 三 > < 三 > 、 三 < の < 0</p>
```



AUTOTYP

- Autotyp doesn't offer any data on nominal plurality, but does have data about the number of numeral classifiers for 32 languages.
- The mean of the number of classifiers is about 16.9.
- However, some caution might be needed, as always. e.g. It lists Mandarin as having 50 numeral classifications, which to me really sounds like an approximation.

[Bickel et. al, 2017, The AUTOTYP typological databases]



SUMMARY

The data seem to corroborate the hypothesis clearly enough. Actually, this hypothesis was pointed out by multiple sources so the result shouldn't be that surprising.

POTENTIAL REASONS

The WALS author David Gil, who originally mentioned this phenomenon in 1987, acknowledges the correlation, but claims in the end that "... the occurrence of classifiers in classifier languages has no straightforward semantic explanation; in the absence of any better-motivated account, it must accordingly be viewed as just another arbitrary conventionalized fact about the grammars of individual languages."

However...

As a Chinese speaker, there are some functionalities of numeral classifiers, related to uncertainty reduction/disambiguation that seem very natural and plausible to me.

UNCERTAINTY REDUCTION (DISAMBIGUATION) Just as gender is used to reduce uncertainty in some Indo-European languages [Ramscar 2007], numeral classifiers serve a similar purpose as the gendered articles do.

- ► For *high frequency* nouns, co-occurrence tends to predict gender *difference*, and there are more irregularities.
- ► For *low frequency* nouns, semantic similarity tends to predict the *same gender*.

The same idea can be applied to numeral classifiers. Example: Common livestocks.

Different classifiers:

- ▶ 一头牛 (one-CLF ox)
- ▶ 一只鸡 (one-CLF chicken)
- ▶ 一匹马 (one-CLF horse)
- ▶ 一条狗 (one-CLF dog)

Same classifier:

- ▶ 一只鸡 (one-CLF chicken)
- ▶ 一只熊 (one-CLF bear)
- ▶ 一只老虎 (one-CLF tiger)
- ▶ 一只兔子 (one-CLF rabbit)

San

UNCERTAINTY REDUCTION (DISAMBIGUATION)

- On the left, none of the classifiers are the same! They are the most common livestocks that are involved in the daily lives of everybody.
- ► When somebody says "匹...", the listener already expects "horse" as the livestock to be mentioned.
- Also note the irregularity: "匹" was originally a quantifier for cloth/silk. Only horse uses this classifier among animals.

On the right, animals like "bear", "tiger", "rabbit" etc. are presumably much less common as domesticated livestock, and thus it makes more sense for them to share the same classifier.

Specification

The numeral classifier itself carries certain information about the subject. Some classifiers can also be used as a concrete noun instead of something abstract, though such usages are already rare.

e.g. "条" normally indicates an elongated shape.

Used as a classifier:

- ▶ 一条毛巾 (one-CLF towel)
- ▶ 一条河 (a-CLF river)

Used as a noun (strip):

- ▶ 切成条 (chop into strips)
- ▶ 面条 (noodle)
- ▶ 柳条 (twig)

Even the same noun can be paired with different classifiers.

- ▶ 三朵花 (Three flowers, the flowers only)
- ▶ 三枝花 (Three flowers with the stalks)

Although, normally speaking, the contrast is mostly between measural numeral classifiers and sortal numeral classifiers:

- ▶ 三篮苹果 (Three buckets of apples)
- ▶ 三个苹果 (Three-CLF apples)

▲□▶ ▲□▶ ▲目▶ ▲目▶ ▲□ ● ● ●



SUMMARY

- The correlation between the number neutrality of a language and the usage of numeral classifiers of that language seems very clear.
- An apparent explanation for the semantic functionality of the classifiers is to reduce uncertainty.
- However, a direct semantic explanation for the correlation between plural markers and numeral classifiers is less clear.
- Maybe nominal plurality is just an aspect of certain related properties of a language (e.g. isolating) that would benefit from the further uncertainty reduction offered by numeral classifiers. Further research might be needed.

INTRODUCTION DATA & ANALYSIS DISCUSSION REFERENCES **References**

References I

- David Gil. "Definiteness, Noun-Phrase Configurationality, and the Count-Mass Distinction". In: *The Representation of* (*In*)*definiteness*. Ed. by E J Reuland and A G B ter Meulen. Cambridge: MIT Press, 1987, pp. 254–269.
- David Gil. "Numeral Classifiers". In: The World Atlas of Language Structures Online. Ed. by Matthew S Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology, 2013. URL: http://wals.info/chapter/55.
- Martin Haspelmath. "Occurrence of Nominal Plurality". In: *The World Atlas of Language Structures Online*. Ed. by Matthew S Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology, 2013. URL: http://wals.info/chapter/34.

REFERENCES II

- Marian Klamer. "Typology and grammaticalization in the Papuan languages of Timor, Alor and Pantar". In: *Grammaticalization from a typological perspective.* 2016.
- Godehard Link. "Quantity and Number". In: Semantic Universals and Universal Semantics. Ed. by D Zaefferer. Berlin: Foris, 1991, pp. 133–149.
- Willard van Orman Quine. "Ontological Relativity". In: Ontological Relativity and Other Essays. Ed. by Willard van Orman Quine. New York: Columbia University Press, 1969, pp. 26-68.

INTRODUCTION DATA & ANALYSIS DISCUSSION REFERENCES References

References III

Timothy Shopen. Language Typology and Syntactic Description. Volume III: Grammatical Categories and the Lexicon. Cambridge: Cambridge University Press, 2007, p. 426. ISBN: 978-0-521-58855-3 (paperback). DOI: 10.1017/CB09780511618437. arXiv: arXiv:1011.1669v3. URL: http:

//ebooks.cambridge.org/ref/id/CB09780511618437.

- Mark J Stein. *Quantification in Thai.* 1981.
- Laurence C Thompson. *A Vietnamese Grammar*. Seattle: University of Washington Press, 1965.

Thank You!

◆ロト ◆舂 ト ◆臣 ト ◆臣 ト ○臣 - のへで