**Review Article**

**Lexical numeral systems and relative language universals**

**Abstract**

In this paper, we attempt to investigate universality in lexical numerals across a number of languages and/or cultures. Cardinal numerals from one (1) to hundred (100) of various languages have been lexically, morphologically and arithmetically analyzed in connection with and in the explanation of their universality. The analysis is based on Greenberg’s idea of Language Universals, which has been further developed and termed *Relative Language Universals* in this paper. Following Justus, we contend that the most prominent type of lexical numeral system in the majority of languages across the world is base-10 (i.e. decimal system) irrespective of cultural differences and even diversities in areal morphology in the lexical numerals construction, a position which is supported by Greenberg’s generalized prediction of base-10 numeral system as a language universal. Furthermore, through lexical diffusion and in a constraint-based analysis, we explain why base-10 is the preferred and dominant system and/or why it is assuming prominence among counting systems regardless of typological and/or genetic differences.

**Keywords:** language universals, lexical numbers, numeral systems, typology

**Introduction**

This paper discusses the universality of lexical numerals across a number of languages and/or cultures. Particularly, cardinal (or lexical) numerals in a sizeable number of languages are observed and lexically, morphologically and arithmetically analyzed. Scholars like Flegg, Cinque, and Donohue have looked into the typology of lexical numerals across a number of languages and expressed different views on Greenberg’s proposal of universals of lexical numerals. In this paper, we propose alternative universals to Greenberg’s language universals on number based on the position that absolute universality is far-fetched. Based on our idea of relative language universals, we argue for a relative universality in the system of counting (with lexical numerals) across languages.

**Hypothesis and data analysis**

Flegg observes that there is a tendency to count numerals in many languages on base-10 (i.e. decimal system) and the database of this study, some of which are given in (1) and (2) below, seems to support this observation. Although these languages are typologically and/or genetically unrelated, it is interesting to note that many of them adopt or use the base-10 numeral system. This immediately suggests that, indeed, base-10 is or could be the dominant numeral system. The Proto-Indo-European family of languages is generally assumed as the source of base-10. Regardless of inheritance or borrowing through language contact, however, one wonders why the majority of languages among the over six thousand nine hundred known and living languages in the world adopt a base-10 numeral system. In other words, are there reason(s) why many languages use (or are tilting towards the use) of base-10? Also, with many languages using base-10, could we claim that the base-10 numeral system constitutes a case of (near) language universal?

**Numeral calculations**

Based on our database of lexical numeral from several languages, we illustrate the various numeral systems in this section. For the sake of commonality and ease of comprehending each system, the number ‘one hundred and twenty three’ (123) is mostly illustrated in most languages. With these illustrations, it could be observed that a wide range of languages seems to support Greenberg’s generalized prediction of base-10 numeral system as a language universal.

Following this observation, we will continue to delve into the universality or otherwise of base-10 in the rest of the paper.

### I. Base-10 (decimal system)

#### i. Chinese (for standard Putonghua and all regional dialects):

→ 1 100 2 10 3

→ 1 x 100 + 2 x 10 + 3 (literal calculation)

(1 x 100) + (20 + 3) = 123 (underlying calculation)

#### ii. Thai:

→ 1 100 2 10 3

→ 1 x 100 + 2 x 10 + 3 (literal calculation)

(1 x 100) + (20 + 3) = 123 (underlying calculation)

#### iii. Zhuang:

→ 1 100 2 10 3

→ 1 x 100 + 2 x 10 + 3 (literal calculation)

(1 x 100) + (20 + 3) = 123 (underlying calculation)
iv. **German:**

\[
ein \quad \text{hundert dreundzwanzig}
\]

\[
\rightarrow 1 \quad 100 \quad 3 \quad \text{and} \quad 20 \\
\rightarrow 1 \quad x \quad 100 + \quad 3 + \quad 20 \quad \text{(literal calculation)}
\]

\[
(1 \times 100) + (20 + 3) = 123 \quad \text{(underlying calculation)}
\]

v. **Tagalog:**

\[
\text{sandaan dalawa-mpu’t} \quad \text{tatto}
\]

\[
\rightarrow 100 \quad 2 \quad 10 \quad ‘t \quad 3 \\
\rightarrow 100 + 2 \quad x \quad 10 + \quad 3 \quad \text{(literal calculation)}
\]

\[
(1 \times 100) + (20 + 3) = 123 \quad \text{(underlying calculation)}
\]

Observe that ‘-mpu’ is the short form of ‘sampu’, meaning ‘ten’ and ‘-’t’ is the short form of ‘at’, meaning ‘and’.

vi. **Turkish:**

\[
yüz yirmi üç
\]

\[
\rightarrow 100 \quad 20 \quad 3 \\
\rightarrow 100+20+3 \quad \text{(literal calculation)}
\]

\[
(1 \times 100) + (20 + 3) = 123 \quad \text{(underlying calculation)}
\]

vii. **Akan (Twi):**

\[
ha \quad ne \quad adaunu-mmensaa
\]

\[
\rightarrow 100 \quad \text{and} \quad 10 \quad x \quad 23 \quad \text{(literal calculation)}
\]

\[
\rightarrow 100 + (10 \times 2) + 3 = 123 \quad \text{(underlying calculation)}
\]

II. **Base-10 + base-20 (decimal-vigesimal system)**

viii. **Dagaare:**

\[
k \quad ne \quad lezare \quad ne \quad bata
\]

\[
100 \quad \text{and} \quad 20 \quad \text{and} \quad 3
\]

<table>
<thead>
<tr>
<th>Numeral</th>
<th>Gumulgal language version 1</th>
<th>Gumulgal language version 2</th>
<th>Kauralgal language</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Urapon</td>
<td>Urapon</td>
<td>Neecha</td>
</tr>
<tr>
<td>2</td>
<td>Ulkasar</td>
<td>Okasa</td>
<td>Boolla</td>
</tr>
<tr>
<td>3</td>
<td>Ulkasar-urapon (2+1)</td>
<td>Okasa-okura (2+1)</td>
<td>Boolla-neecha (2+1)</td>
</tr>
<tr>
<td>4</td>
<td>Ulkasar-ulkasar (2+2)</td>
<td>Okasa-okura (2+2)</td>
<td>Boolla-booala (2+2)</td>
</tr>
<tr>
<td>5</td>
<td>Ulkasar-ulkasar-ulkasar (2+2)</td>
<td>Okasa-okura-okura (2+2)</td>
<td>Data not available (N.A.)</td>
</tr>
<tr>
<td>7</td>
<td>Ras (all numerals beyond 6 are called ‘ras’ collectively)</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>8</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>9</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>10</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
</tbody>
</table>
VI. Base-3, 4, 6, 7, 8, 9, 11, 12, 13, 15, 16, 40, etc.

English and French and other Romance languages (base-12): This has to do with a historical remnant of counting and it is restricted to the counting of limited items. For example, in English, we have ‘one dozen of eggs, two dozens of eggs, etc. and in French, we say ‘une douzaine de stylos (i.e., one dozen of pens); deux douzaines de stylos.

xiii. Hawaiian (base-40): This is an old counting style:

elua kanaha me kaiwakalua (numeral “100"
\[2 \times 40 \, \text{and} \, 20 \quad (\text{literal calculation})\]
\[2 \times 40 \, + \, 20 \, = \, 100 \quad (\text{underlying formula})\]

Russian, Ukrainian and Byelorussian: Numeral 40 “sórok”.

Wiese\textsuperscript{14} suggests that “the Russian word for forty, sórok, is believed to originate from the old Nordic word for furs, sekr, which were traded in bundles of forty.”

VII. Multiple mixed bases

As noted earlier, these bases are due to millennium-long inter-cultural contacts and interactions.

xiv. French (base-10 + base-20 + base-60): Table 3.


The Welsh numeral, ddeg, changes into -theg in some cases due to the consonantal mutation in the preceding numeral pym-. The numeral deu also changes into dau in the numeral dau ar bymtheg.
Lexical numeral systems and relative language universals

Numeral systems

With his proposal of forty-five (45) language universals, Greenberg attempts to predict what typological and/or the morphological structures of some constructions should pertain in different languages, including structures that involve number. Numbers eighteen (18) and twenty (20) of his proposed universals are the ones that most relate to number and its relation to other constituents. These are presented as (8a) and (8b) below.

a. Universal number 18: When the demonstrative adjective precedes the noun, the demonstrative and the numeral, with overwhelmingly more than chance frequency, do likewise.

b. Universal number 20: When any or all of the items (demonstrative, numeral, and descriptive adjective) precede the noun, they are always found in that order. If they follow, the order is either the same or its exact opposite.

c. While Greenberg's proposals do not include a particular universal numeral or counting system, his assumption of linguistic universals (whether in terms of absolute universality or statistical universality) cannot be overlooked in any attempt to understand the counting systems across languages. So, the question one could ask and expect to get an answer is, is there a universal numeral system across languages? And, this is where Greenberg's idea of language universals comes to bear on this paper.

As noted earlier, our research database hugely supports the fact that base-10 numeral system is the most prominent. We have shown some representative languages of it as the Sino-Tibetan, the Tai-Kadai, Altaic-Turkic, Austr-Asiatic, Polynesian, majority of Indo-European and Semitic language families. Thus, following Justus, we contend that the most prominent type of lexical numeral system in a great majority of languages is base-10 – i.e. the decimal system – despite cultural differences and diversities in areal morphology in the lexical numerals construction. Our database also shows base-10 is often involved in cases where other base counts seem to dominate. For instance, Basque and the West African languages of Dagaare and Yoruba adopt base-10 plus base-20, hence a mixed numeral systems. Two or three languages, including Cambodian, adopt base-5 (quinary system) plus base-10, making a mixed numeral system. The rest, such as base-2, 4, 12, 15, 40, 60, are uncommon and are adopted in some exceptionally rare and highly restricted cultures. In particular, Justus notes that base-60 (i.e., sexagesimal) system is hardly found in modern languages. It is believed that it was used during civilizations like ancient Babylonian and ancient Middle Eastern empires. It was also used in ancient China, i.e. during the Shang dynasty between 1600–1100BC, in the counting of 60-year traditional calendar cycle. We explain in section 4 that these numeral systems are rare because they violate three basic principles that establishes an optimal counting system and these explain why the base-10 system is dominant and relatively universal. The last type of the numeral systems, multiple mixed bases, is quite common in many languages and is due to millennium-long inter-cultural contacts and interactions. In this system, three or more different numeral bases are used to express a wide range of numerals. Considering a system of base-5, base-12, base-20 and base-60 in particular, Justus refers to multiple mixed bases as ‘non-pure’ counting; where ‘pure’ counting refers to single bases. We will continue to use the term multiple in order to distinguish between numeral systems of three or more different bases and those with two different bases (i.e. dual-based), which could also be described as ‘non-pure’ counting.

Sapir-Whorf hypothesis and language universals

Named after Edward Sapir and Benjamin Lee Whorf, the Sapir-Whorf hypothesis is a mould theory of language; i.e., the position that language is ‘a mould in terms of which thought categories are cast’. Specifically, it suggests that each language has its own worldview based on cultural relativity and that language affects the way in which a linguistic group (of people) perceives the world. Cognitively, therefore, differences in worldview may exist between speakers of different languages. The core of their suggestion is emphasized in Whorf as follows:

“We dissect nature along lines laid down by our native languages. The categories and types that we isolate from the world of phenomena we do not find there because they stare every observer in the face; on the contrary, the world is presented in a kaleidoscopic flux of impressions which has to be organized by our minds – and this means largely by the linguistic systems in our minds. We cut nature up, organize it into concepts, and ascribe significances as we do, largely because we are parties to an agreement that holds throughout our speech community and is coddled in the patterns of our language. The agreement is, of course, an implicit and unstated one, but its terms are absolutely obligatory; we cannot talk at all except by subscribing to the organization and classification of data which the agreement decrees” Whorf.

To some extent, Sapir and Whorf were right in pointing out that each linguistic group of people tends to dissect the world differently with individual native languages as cutting tool. Justus view in the...
analysis of numerals in particular is in line with the Sapir-Whorf hypothesis in the sense that, as she suggests, counting strategies of lexical numeral systems vary from one language to another. However, we believe that some linguists and experts of other disciplines might have over-emphasized the Sapir-Whorf hypothesis with regards to the fact that it predicts all kinds of cultural relativities, including numeral systems. Supposing Sapir-Whorf hypothesis (with particular reference to numerals) is entirely right, some questions (that really have to be answered) could be raised, some of which we state below.

i. If each linguistic group is cognitively pre-determined to perceive the world differently and to develop or adopt its own numeral system, why is it that base-10 numeral system is used in a lot of genetically distant and typologically unrelated language families?

ii. Why is it that many languages that use numeral systems other than base-10 still tend to have cognitive free will to borrow base-10 numerals from other languages?

iii. Is the base-10 numeral system a mere coincidence or is there an explanation for its worldwide usage?

iv. Is there something relatively universal behind the use or adoption of base-10 by different languages (and the diversity of numeral systems across languages)?

Towards the universality of base-10 numeral system

It is not far-fetched to take as a fact that the dominance of base-10 numeral system has to do with influences due to cultural or language contacts. In particular, Greenberg (1978) explains the impact of intercultural influences or language contacts on the lexical numeral systems as follows:

“It is a well-known phenomenon that higher numerals are more commonly borrowed than lower ones, usually in a sequence starting at a certain number … In other instances of contact, the result is the replacement of an old system by a new one, with or without borrowing … This seems to show that, for the speakers, there is a certain psychological reality attached to the notion base. There are instances in which related or borrowed words for one base are used with the value of a different base.”

Our data supports Greenberg’s observation that higher numerals are commonly borrowed from other languages (with a stronger linguistic and cultural influence). Indeed, the so-called ‘old systems’ are hard to count because of their clumsiness, hence the desire to replace them with an efficient numeral system(s). In other words, users prefer a user-friendly numeral system(s) to cope with increasingly complicated counting issues. This certainly constitutes an argument for borrowing from a source language with a more convenient and efficient counting system. It is a fact that various linguistic groups attempt to preserve their own traditional ways of counting. However, generally, people tend to replace clumsy numeral systems when it comes to seeking optimal efficiency and productivity. In some cases, the old numeral system of the receiving language and the new numeral system of the source language may co-exist, as in languages using mixed or multiple numeral systems, as in Welsh, French, and some ethnic minority languages. In our attempt to explain why base-10 numeral system seems to be universal across languages, we propose the notion of relative language universal (RLU) on numeral systems, which is stated as follows.

Relative language universal (RLU): A (counting) system is relatively universal among other systems where it is the most common system between different and genetically unrelated languages and to which languages of other systems seem to tilt to.

As stated in (10), RLU diverges in argument from Greenberg’s language universals, specifically in the sense of absoluteness and relativity. One could however say that RLU is a modification of Greenberg’s language universals. That is, RLU explains that several numeral systems may be in existence, as has been noted in our illustrations above. However, depending on the number of languages that uses a particular numeral system and/or the high degree of adoption of this system by different and unrelated languages or linguistic groups (which traditionally use other systems), the system could be described as relatively universal. As will be observed in the following sub-sections, therefore, the concept of relativism imputed in RLU defines the high number of languages that use base-10 and the increase in the adoption of it among languages.

Our proposal of RLU involves two sets of ideas; namely, internal relative language universals (IRLUs) and external relative language universals (ERLUs). IRLUs have to do with factors that relate to the structuring of numerals (e.g., ethno-linguistic, cognitive and/or arithmetic factors) while ERLUs are due to human-related factors affecting numeral usage (e.g., social, cultural, historical, economic, political factors).

Internal relative language universals (IRLUs)

The term ‘internal’ (in IRLUs) particularly refers to the fundamental blocks of numeral systems (e.g. counting reference to base), without which any numeral system would be hardly functional. Three basic principles constitute IRLUs and provide reasons as to why base-10 numeral system is relatively universal among languages. Motivated on mankind’s desire for counting efficiency and ease of counting, these three principles are optimal regularity, optimal utilization and optimal convenience. With the notion of optimality in these principles, we emphasize the desire for the best (or relatively universal) system of counting among a variety of possibilities. The principles are individually explained as follows.

Optimal regularity suggests that an efficient numeral system should not involve structurally complicated ways of counting such as back-counting, over-counting and half-counting. Thus, complications are seen as irregular (and, for that matter, worst). As noted earlier, Welsh, for instance, used to employ a complicated multiple-based numeral system. A base-10 numeral system was later introduced into the language, which is preferred and is presently being used by Welsh speakers. This explains the desire of a linguistic group for an optimally regular and simplest system of counting.

Optimal utilization also explains that effective counting should involve the use of a proximate or inalienable natural tool. Accordingly, the base-10 numeral system is a common (and the preferred) system across many linguistic and/or cultural groups because it involves the utilization of a proximate or readily available natural tool, our ten fingers. In other words, base-10 may have originated from the use of our ten fingers as the tool for keeping count. Such a view could be supported by the fact that children learn or are taught how to count with the aid of the fingers in most traditional societies. This may include pointing to or touching of the objects they are counting.

Citation: Marfo C, Kam SS, Bodomo A. Lexical numeral systems and relative language universals. Art Human Open Acc J. 2020;4(1):5–11.
DOI: 10.15406/ahaoj.2020.04.00145
Optimal convenience explains that a number system must be halfway between small numeral bases and large numeral bases, hence the convenience of base-10. We note that a large numeral base (e.g., base-20, base-40 or base-60) may make counting of small amounts less efficient, whereas a small numeral base (e.g., base-2, base-4 or base-5) makes the counting of large amounts more difficult and arithmetically unachievable. Base-10 on the other hand tends to capture the counting of small and large amounts conveniently. What is more, it effectively keeps inter-numeral system association; i.e., it works well with the other numeral systems. As could be seen in Table 5 below, most languages use this association (specifically, dual-based mixed numeral system); e.g., Cambodian uses base-10 and base-5 and Babylonian uses base-10 and base-60.

| Table 5 Combinations of major types of dual-based mixed numeral systems |
|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|
| base-2                      | base-5                      | base-10                     | base-15                     | base-20                     | base-60                     |
| Default                     | Default                     | Default                     | Data not available yet      | Default                     | Default                     |
| (Austronesian)              | (Australian)                | (Austronesian)              | (e.g., Cambodian)           | (e.g., Maenge)              | (e.g., Dagaare)             |
| base-5                      | ✓                           | ✓                           | ✓                           | ✓                           | ✓                           |
| (Austronesian)              |                             |                             |                             |                             |                             |
| base-10                     | ✓                           | ✓                           | ✓                           | ✓                           | ✓                           |
| (Australian)                |                             |                             |                             |                             |                             |
| base-15                     | Data not available yet      | Data not available yet      | Default                     | Default                     |                             |
| (Austronesian)              |                             |                             |                             |                             |                             |
| base-20                     | x                           | ✓                           | ✓                           | ✓                           | Default                     |
| (e.g., Maenge)              |                             |                             |                             |                             |                             |
| base-60                     | x                           | x                           | ✓                           | x                           | Default                     |
| (Babylonian)                |                             |                             |                             |                             |                             |

External relative language universals (ERLUs)

As noted earlier, ERLUs have to do with human-related factors that contribute to the adoption of particular numeral system by geographically and/or linguistically unrelated languages. The term ‘external’ in ERLUs classifies these factors as meta-linguistic and some of them are colonization and migration that ultimately result in century-long language contacts and interferences (from a source language to a target language) over a long period. Indeed, language contacts and interferences due to one reason or another are what initiate series of cultural and linguistic changes through outright borrowing of foreign lexicon, modification of borrowed lexicon and concurrent use of both borrowed lexicons and native lexicons, including lexical numerals. As1 notes, ‘[numerals] may reflect specific cultural phenomena’. Thus, language contacts may lead to a partial or total shift in lexical numeral systems of some languages, particularly less dominant languages or ethnic minority languages. Now, assuming the base-10 numeral system originated from one linguistic region, the spread of it to neighboring linguistic regions or its dominance could be explained by language contacts and interferences.

We suggest that the adoption of base-10 is completed in a language through lexical diffusion;4,21-24 i.e. the gradual spread of a sound change through the relevant items in the vocabulary. Wang4 in particular suggests that lexical diffusion starts from a small area of lexical words and the affected area spreads gradually to other areas. He specifically proposes lexical diffusion to explain the gradual replacement of old lexicons of Chinese by new lexicons, but it has been resorted to in lexicology and morphology of other languages.4,21 Wang and Cheng4 in particular notes, ‘we hold that words change ... by discrete, perceptible increments (i.e. phonetically abrupt), but severally at a time (i.e. lexically gradual) ...’. Following and extending the thought of Wang and Cheng,4 we assume that, in theory, lexical diffusion is also applicable to the explanation of the gradual replacement of old and arithmetically complicated numeral system of some languages by a new and efficient numeral system from another language, hence its invocation in this paper.

While lexical diffusion could explain the dominance/spread of base-10, it leaves the issue of age of the numeral systems unexplained. However, the spread of base-10 could be explained through archeological and ecological insights and inspirations. Specifically, the effects of meta-linguistic factors pin down old and clumsy numeral systems to the bottommost layer of a ‘sedimentary of numeral systems’ and ultimately become fossilized, having been adversely influenced. The borrowed optimal numeral system, base-10, then constitute the top layer of the ‘sedimentary of numeral systems’.

Conclusion

In this paper, we have discussed lexical numeral systems and have particularly delved into why, among all numeral systems, base-10 is the dominant one. We have attempted to explain the optimal status of base-10 from the perspective of two different, but intertwined ideas, namely internal and external relative language universals. In proposing these two ideas, we have identified some possible reasons and/or causes for the worldwide preference and adoption of base-10 numeral system among some languages, whether or not they are typologically and/or genetically related. In particular, with the three identified principles of optimal regularity, utilization and convenience, we have explained why base-10 is dominant among the languages in...
our research database. We hope that our proposal of relative language universals could explain the gradual trend of dominance of the base-10 numeral system.

Acknowledgments

None.

Conflicts of interest

The authors declare that there is no conflict of interest.

Funding

None.

References