Rapid information gain explains cross-linguistic tendencies in numeral ordering

Emmy Liu (me.liu@mail.utoronto.ca)
Yang Xu (yangxu@cs.toronto.edu)
Department of Computer Science
Cognitive Science Program
University of Toronto

Abstract

One previously unexplained observation about numeral systems is the shared tendency in numeral expressions: Numerals greater than 20 often have the larger constituent number expressed before the smaller constituent number (e.g., *twenty-four* as opposed to *four-twenty* in English), and systems that originally adopt the reverse order of expression (e.g., *four-and-twenty* in Old English) tend to switch order over time. To explore these phenomena, we propose the view of Rapid Information Gain and contrast it with the established theory of Uniform Information Density. We compare the two theories in their ability to explain the shared tendency in the ordering of numeral expressions around 20. We find that Rapid Information Gain accounts for empirical patterns better than the alternative theory, suggesting that there is an emphasis on information front-loading as opposed to information smoothing in the design of large compound numerals. Our work shows that fine-grained generalizations about numeral systems can be understood in information-theoretic terms and offers an opportunity to characterize the design principles of lexical compounds through the lens of informative communication.

Keywords: language universals; numeral system; lexical compound; information theory; informative communication

Number is a fundamental domain of human cognition (Spelke & Kinzler, 2007), but numeral systems vary substantially across cultures (Comrie, 2013). For instance, some cultures in the Amazon lack exact numerals for expressing numbers beyond 5 (Gordon, 2004; Pica, Lemer, Izard, & Dehaene, 2018). Some languages use body parts to describe numbers (Comrie, 2013). However, the majority of languages in the world define numbers precisely and over a large range through recursive numeral systems (Comrie, 2013). Recent work has suggested that the diversity of numeral systems is constrained by the need for efficient communication (Xu & Regier, 2014; Kemp, Xu, & Regier, 2017). By this account, numeral systems are designed to facilitate highly informative communication of numbers, despite their differences in complexity.

The proposal of informative communication helps to explain why numeral systems vary the way they do, but it does not directly account for fine-grained generalizations about numeral expressions. In particular, many languages express compound numerals by specifying the larger constituent number first (e.g., *twenty-four* in English or Mandarin), and fewer languages express these in the reverse order (e.g., *vierentwintig* in Dutch, interpreted as "four twenty"). Moreover, numeral systems that originally use the reverse order of expression (e.g., Old English expresses 24 as *four-and-twenty*) tend to switch order over time (Berg & Neubauer, 2014). This preference of having the larger constituent number expressed before the smaller constituent number is prevalent in numerals for the range above 20 but less prominent for smaller numbers (Calude & Verkerk, 2016). Here we ask what principles might account for this shared tendency in numeral ordering.

This problem has been discussed by Greenberg in his cross-linguistic generalization about the design of recursive numerals (Greenberg, 1978). Recursive numeral systems represent numbers based on the canonical expression $x_1 n^k + ... + x_k n + y$. Here n is called the base and the values of x_i's and y are in the range of 1 to the base (Comrie, 2013). For numbers in the range $1 - 100$ in a base-10 system such as English or Mandarin, xn will be considered the *base* term (i.e., 10, 20, ..., 90) and y (i.e., 1, 2, ..., 9) will be considered the *atom* term. Greenberg observed that if a numeral system has both atom-base (e.g., *fifteen*) and base-atom (e.g., *twenty-four*) orderings in its numeral expressions, the system will always begin with atom-base, and then switch to base-atom at some number on the number line (Greenberg, 1978). In English and many other languages, this switch takes place at 20.

Independent work from Hurford has sought to address this phenomenon in light of the "packing strategy" (Hurford, 2007). According to this proposal, numeral expressions should allow one to go as far as possible along the number-line with a given set of terms (Hurford, 2007). This would imply that terms should be arranged in decreasing order, with the larger constituents coming first, and it confirms that the base-atom order should be preferred over the atom-base order. Although this work provides an intuitive theory for the ordering preference in large numerals, it leaves open two important questions: 1) why the base-atom order is preferred across languages for numbers above 20, but this preference is substantially less for smaller numerals (e.g., 11 to 19), and similarly, 2) why ordering switch should typically take place in numerals above 20 and in particular, why it occurs only in one direction (atom-base→base-atom) but not in the other (base-atom→atom-base).

We examine the problem of numeral ordering through the lens of informative communication. Consistent with the growing literature on this topic, we suggest that language design is driven by the basic need for efficient communication (Gibson et al., 2013; Kemp et al., 2017). Extending this line of research, we propose the view of *Rapid Information*

Gain (RIG) that focuses on explaining the design of compound numerals, particularly the ordering of constituent expressions in terms of the need to optimize information flow. We hypothesize that lexical ordering of a compound numeral expression should maximize information gain for the listener in the process of reconstructing the speaker's intended referent. We contrast this view with the established theory of Uniform Information Density (UID) postulating that information smoothing should be preferred (instead of information front-loading) in word ordering in sentences, online (Levy & Jaeger, 2007) or offline (Maurits, Navarro, & Perfors, 2010). We show that RIG explains empirical patterns better than UID in the domain of numerals, and we believe this work has the potential for developing a domain-general account of the design principle of lexical compounds.

Two theories of informative communication

We present the numeral ordering problem in a simple communicative scenario, illustrated in Figure 1a. Here the speaker has the target number 85 in mind and wishes to convey that number to the listener. We consider two possibilities in the ordering of constituent expressions of that numeral, using English as an example: 1) "Eighty-five", which is the *attested order* or base-atom; 2) "Five-eighty", which is the *alternate order*, or atom-base in this case. The problem is to determine which order should be generally preferred in natural languages and in what range of the number line this preference might be most prominent.

We postulate that the preferred numeral order should tend to minimize the listener's uncertainty in reconstructing the target number as the speaker's utterance is processed. We consider how uncertainty arises over time in the listener's mind as the constituent expressions are uttered sequentially by the speaker. Based on the ordering of "eighty-five", upon hearing the first constituent "eighty", the listener would consider numbers in the range 80-89 as possible candidates for the target, because numerals for numbers within that range all begin with the same constituent. In this case, uncertainty depends on the probability ratio between the actual target and the candidate set. Based on the ordering of "five-eighty", upon hearing the alternative first constituent "five", the listener would instead consider numerals that begin with "five" (e.g., 5, 15, ..., 85, 95) as the candidate set for the target. We illustrate these alternative candidate sets in Figure 1a.

We consider two alternative theories that quantify uncertainty given choices of numeral ordering based on Shannon's information theory (Shannon, 1948). The first view is based on Uniform Information Density (Levy & Jaeger, 2007), which predicts that uncertainty incurred should be as smooth as possible. This view suggests that the listener would experience a uniform information flow as a compound expression is uttered. We propose a second view, Rapid Information Gain, that makes the alternative prediction. We hypothesize that the preferred order in compound numerals should tend to front-load information as opposed to smoothing information, such that uncertainty in the listener can be reduced as quickly as possible. We illustrate the predicted uncertainty profile from each theory in Figure 1b. As we show later, the property of information front-loading is more salient in the ordering of larger numbers (>20) than in the case of smaller numbers, which explains why the cross-linguistic preference and the ordering switch toward base-atom expressions are stronger for larger numbers. We now describe the details of each theory.

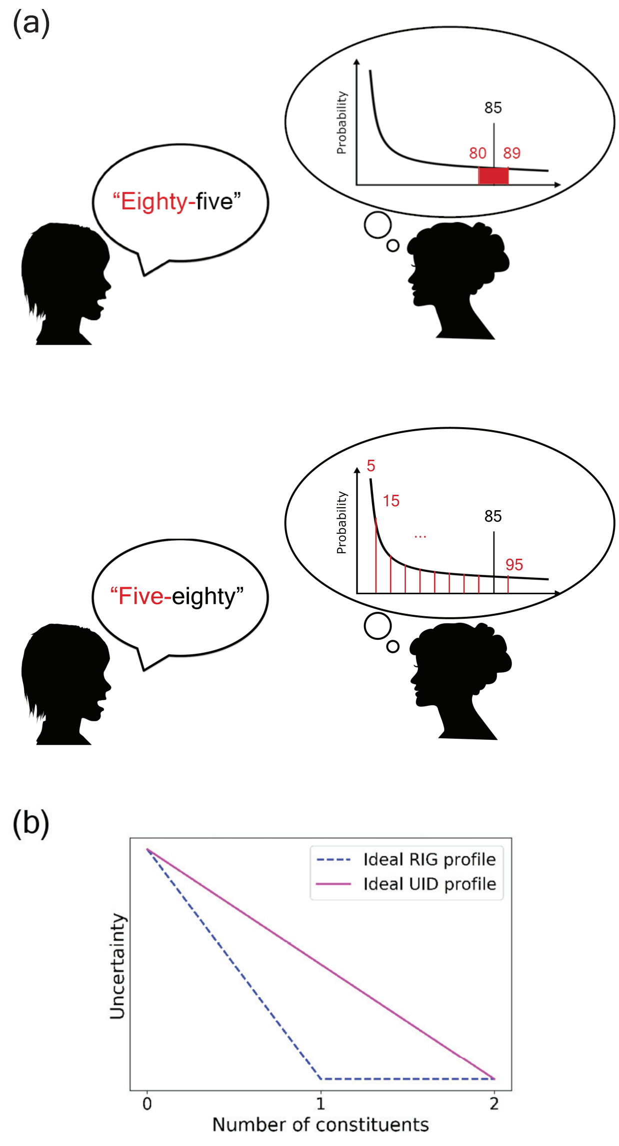

Figure 1: Illustration of the numeral ordering problem and the two theoretical proposals of informative communication.

Uniform Information Density (UID). Following Shannon (1948), we define uncertainty by surprisal or negative log probability $-log_2(p(\cdot)) = log_2(\frac{1}{p(\cdot)})$. We define the information content of a compound linguistic expression by the sum of surprisals from its sequential constituents, following the formulation of UID (Levy & Jaeger, 2007). The cumulative information conveyed by an expression U with n constituents $w_1...w_n$ in reference to a target t is the following:

$$log_2 \frac{1}{p(U)} = log_2 \frac{1}{p(t)} + log_2 \frac{1}{p(t|w_1)} + ... + log_2 \frac{1}{p(t|w_1...w_n)} \quad (1)$$

In the case of two-constituent numeral expressions such as *twenty-four* (i.e., base constituent and atom constituent), this formulation effectively captures the information flow of a compound numeral as it is processed incrementally in terms of its constituent expressions:

$$log_2 \frac{1}{p(t)} \rightarrow log_2 \frac{1}{p(t|w_1)} \rightarrow log_2 \frac{1}{p(t|w_1 w_2)} \quad (2)$$

Cumulative surprisal defined in Equation 1 can thus be simplified to

$$log_2 \frac{1}{p(U)} = log_2 \frac{1}{p(t)} + log_2 \frac{1}{p(t|w_1)} + log_2 \frac{1}{p(t|w_1 w_2)} \quad (3)$$

As such, the cumulative surprisal of hearing "twenty-four" would be $log_2 \frac{1}{p(\text{“twenty-four”})} = log_2 \frac{1}{p(24)} + log_2 \frac{1}{p(24|\text{“twenty”})} + log_2 \frac{1}{p(24|\text{“twenty-four”})}$.

Empirical studies of UID typically focus on speaker information modulation given the predictability of different units. This would involve measuring information-theoretic entropy rather than surprisal formulated here. However, the UID principle implies that the flow of information to follow a uniform trajectory in cumulative surprisal, and we test the applicability of this proposal in the case of numeral ordering.

More specifically, UID suggests an even distribution of information (in the design of compound numerals), such that the amount of information conveyed in the sequence of constituents should be identical. This predicts that if the speaker has alternative ways of ordering a numeral expression, she should choose the order in which information is distributed more evenly. Here we are interested in the cost of a numeral order versus its reverse order, and we quantify cost by measuring how a numeral order deviates from the theoretical UID information flow. Prior work has taken a similar approach to examine whether UID predicts preferred word orders (e.g., subject-verb-object) across languages (Maurits et al., 2010). In that work, deviation from UID is defined by the percentage deviation from the theoretical UID information flow. Abbreviating the components of the information flow in Equation 2 by $I_0 = log_2(\frac{1}{p(t)})$, $I_1 = log_2(\frac{1}{p(t|w_1)})$ ..., we measure the deviation from UID following Maurits et al. (2010):

$$d = \frac{n}{2(n-1)} \sum_{i=1}^{n} \left| \frac{I_{i-1} - I_i}{I_0} - \frac{1}{n} \right| \quad (4)$$

Here n is the phrase length of an expression (Maurits et al., 2010). In our work, we use the same formula to quantify how the design of a numeral expression deviates from UID. Concretely, we consider $n = 2$ because each compound numeral expression that we use for analyses has two constituents. We also know that $I_2 = 0$ since full certainty is obtained after the second (or last) constituent of a numeral is uttered. UID predicts a linear relationship between information content and number of constituents. If UID explains the shared tendency in numeral ordering across languages, we should expect the attested numeral order to yield a smaller deviation from the linear information profile than the alternate order, more so for the numerical range above 20 than the range under 20.

Rapid Information Gain (RIG). We propose an alternative theory for numeral ordering based on rapid information gain. We postulate that the ordering of numerals should facilitate quick delivery of information to the listener, such that the constituent expression that contains more information should be arranged prior to the constituent that contains less information. This notion of rapid information gain is related to work on optimal data selection. For instance, when performing a series of tasks, optimal data selection implies that people should order the tasks so that they gain the most information possible at each step (Oaksford & Chater, 2003). We believe that similar principles apply to the design of numerals. Our proposal is not equivalent to the claim that the larger numeral should always precede the smaller numeral in a compound (Hurford, 2007; Berg & Neubauer, 2014). Instead, it suggests that the ordering of constituent numerals depends on the amount of information they convey, as opposed to their magnitudes per se. We demonstrate later that our proposal correctly predicts information front-loading to be more critical for high-order numbers than low-order numbers, an aspect that could not be explained fully by a magnitude account that always predicts the larger numeral to be expressed first in a compound numeral.

We evaluate our proposal by measuring the cumulative surprisal of a numeral expression over its constituents:

$$c = \sum_{i=0}^{n} I_i \quad (5)$$

This formulation is the same as Equation 3, and we consider $n = 2$ and since $I_2 = 0$, $c = I_0 + I_1$. The RIG theory predicts an elbow-like information profile which differs from the linear profile predicted by UID (see illustrations of the two theoretical information flows in Figure 1b). We expect that a lower cumulative surprisal should generally be preferred as a consequence of rapid information gain. More specifically, the attested numeral order should yield a lower cumulative surprisal than the alternate order when there is a strong preference toward the attested order (e.g., for numbers >20), but the two possible orders might yield similar cumulative surprisals when there is greater flexibility in the ordering conventions of numerals across languages (e.g., for numbers <20).

Materials and methods

To facilitate the information analyses and evaluation of the two theories, we collected numeral frequencies for estimating surprisals along with cross-linguistic numeral data.

Numeral frequencies. We estimated probabilities of the number terms for the range 1-100 (following Xu & Regier, 2014) in 8 different languages: English, French, German, Hebrew, Italian, Mandarin, Russian, and Spanish. We collected these frequency data from the Google Ngrams corpora (Michel et al., 2011) by averaging numeral frequencies from 1900 to 2000. We used part-of-speech tags for numerals in the corpus if those were available for a given language. For each language, we queried frequencies of numeral terms from a standard set of numeral expressions (data from `www.sf.airnet.ne.jp/ts/language/number.html`). When multiple expressions were available for a numeral, we took the most frequent expression. The frequencies of the numerals for each of the languages were normalized to probabilities so that they sum to 1.

Calculation of surprisals. To calculate surprisals, we decomposed a numeral expression into two separate constituents, atom and base, while ignoring connectives such as hyphens, e.g., "twenty-one" → ["twenty", "one"]. Although it is possible to split some terms into multiple constituents, e.g. "quatre-vingts huit" (4x20+8=88) → ["quatre", "vingts", "huit"] ([4, 20, 8]), we chose to split only along additive terms for consistency. We did not choose to treat suffixes as separate constituents. We calculated the surprisal based on each constituent expression, where surprisal is the negative log probability of the target number being correctly inferred from the set of candidate targets. Finally, for each numeral expression we computed the deviation from UID according to Equation 4 and the cumulative surprisal for RIG according to Equation 5.

Cross-linguistic numeral data. We tested the theories against numeral data collected from 334 languages in 53 listed language families sampled from *Numeral Systems of the World's Languages* (Comrie & Chan, 2018). We sampled languages evenly from each family whenever possible, taking 10 from each family, or if 10 were not available, taking the maximum number possible. This was so that language families with a large number of languages such as Indo-European or Sino-Tibetian did not bias the sample. For each language, we recorded the attested orders in the numeral expressions, atom-base or base-atom, for the numerical ranges of 11-19 and 21-29 (chosen to be symmetric about 20 where order switch most commonly takes place). If a language did not have sufficient data for the numerical ranges, we would exclude that language and sample other languages from the family until 10 or the maximum possible number were collected.

Results

Empirical patterns in the ordering of numerals. We first present cross-linguistic tendencies and switches in "atom-base" and "base-atom" ordering of numeral expressions in the sample of 334 languages that we considered. Table 1 summarizes the cross-linguistic occurrences for these orders in the numerical ranges 11-19 and 21-29. If the atom-base ordering was used for at least one term in 11-19 in a language, we considered that language as having an atom-base ordering in that range. We observed that the base-atom order is attested in more than 96% of the languages for the range 21-29, whereas this order is attested much less commonly in about 76% of the languages for the lower range 11-19. This finding confirms descriptive generalizations from previous work (e.g., Greenberg, 1978) and indicates an asymmetric preference toward base-atom ordering in larger numerals, and more flexibility in the ordering of smaller numerals.

Table 2 confirms that the same asymmetric preference applies to switches in the ordering of numerals. In particular, out of all languages that were attested to have switched order in numeral expressions, switch took place exclusively in the direction atom-base → base-atom but not in the opposite direction. Moreover, out of the 63 languages that use the atom-base order for expressing the numerical range 11-19, 52 (or ∼83%) switch the order to base-atom but only for numerals expressing the range 21-29. Together, these empirical data suggest that preference toward the base-atom order is more prominent in larger but not smaller numerals.

Numeral frequencies across languages. Figure 2 summarizes the meta-mean and language-specific probabilities of numerals, estimated from the corpus-based frequencies over the past 200 years. These probability profiles show a consistent near-logarithmic decay that confirms previous findings in cross-linguistic numeral and digit-based frequencies (Greenberg, 1978; Calude & Verkerk, 2016): Numerals in the lower numerical range tend to be referred to more frequently than numerals in the higher range. We used these probabilities for surprisal calculations for the two theories.

Table 1: Ordering conventions in numerals across languages.

Number of languages	Range 11-19	Range 21-29
atom-base ordering	63	11
base-atom ordering	271	323

Table 2: Switch in numeral ordering conventions. For each language, the original numeral order is the same as that in the lower range 11-19, and ordering switch is attested in numerals for the upper range 21-29.

Number of languages	No switch	Switched
atom-base → base-atom	11	52
base-atom → atom-base	271	0

Evaluation of the two theories. We evaluated UID and RIG by first considering a "template" language that reflects the cross-linguistic tendency in numeral ordering we and other scholars have observed: Atom-base order in numerals

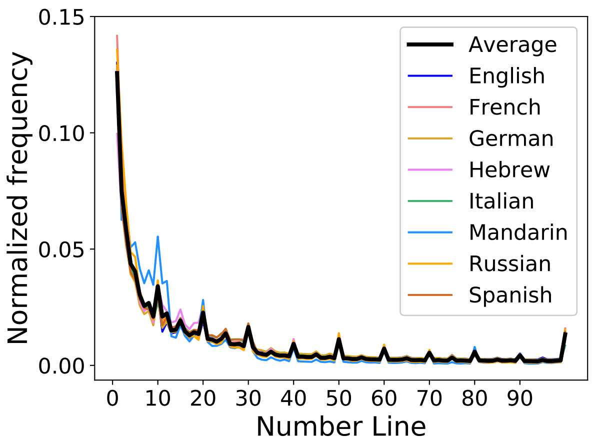

Figure 2: Numeral frequencies across 8 languages.

for the range 11-19, and base-atom order in numerals for the range above 20. An ideal theory should explain 1) why there is a strong preference for the attested base-atom order over the alternate atom-base order in the upper numerical range; 2) why this preference between the attested and alternate orders is much weaker in the lower range. As such, we expected a greater discrepancy in the attested and alternate orders for the theoretically predicted information profile we described (e.g., under UID or RIG), and a substantially smaller discrepancy in these orders for the same measure of information. To test these ideas, we calculated the information profile for each of the numerals within the range 1-100 based on the mean numeral probabilities we had obtained. We performed these calculations for both the attested order and the alternate order, resulting in two sets of measures for UID deviation and two sets of measures for RIG cumulative surprisal.

Figure 3 (a) and (d) summarize the results. At the broad level, both UID and RIG identify the attested order to be closer to their theoretical information profiles than the alternate order. However, a closer examination of these results reveals variation in the precision of these theories. For the numerical range beyond 20, UID shows an ambivalent preference toward the base-atom order over the atom-base order, manifested in the noisy deviation scores between the two orderings. In contrast, RIG provides a clearer advantage of the base-atom order over the atom-base order for numerals in the same range, indicating that there is a dominance toward the first order as predicted by this theory. Moreover, for numerals in the range 11-19, UID shows a strong support for the base-atom order, but RIG shows that both orderings render roughly equal cumulative surprisals—this suggests that information front-loading is less relevant to ordering variation in this lower numerical range.

To further examine the precision of the two theories, we examined their predictions for two sample languages, English and Mandarin. For these cases, we used language-specific numeral probabilities for calculations of UID deviation and RIG cumulative surprisals. Figure 3 shows that the results for these individual languages are consistent with our findings with the template language, such that RIG provides a more precise explanation for the asymmetric preference in ordering of larger and smaller numerals. Figure 4 illustrates the information profiles in the attested and alternate orders with two example numerals, *fifteen* and *twenty-four* in English, along with the theoretical predictions from UID. In both cases, the attested order shows an elbow-like information profile that deviates from the ideal linear profile of UID, providing evidence against the idea that numerals are designed under the criterion of information smoothing. Importantly, the information profile under the alternate order for *fifteen*—a low-order numeral—is almost identical to the elbow-like profile under the attested order, reflecting the fact that information front-loading is insensitive to ordering of numerals in this range. It is worth noting that both alternate and attested profiles deviate from the UID prediction. In addition, for *twenty-four*, the alternate order produces an information profile that approaches the UID prediction. This profile yields a cumulative surprisal higher than the attested order, suggesting information front-loading is desirable for larger numerals in English.

As a final analysis, we examined whether the preferred ordering switch from atom-base to base-atom can be explained away by the theory of RIG. In particular, we performed a focused analysis that compares cumulative surprisal between these two orders for the numerical ranges 11-19 and 21-29 respectively. We expected that the cumulative surprisal might be comparable under the two orders for the smaller range, but substantially discrepant for the larger range, which would explain why switching of order tends to occur beyond 20 and only in the atom-base $\rightarrow$ base-atom direction.

For each of the numerical range in question, we conducted a permutation test that shuffles the numeral expressions between the base-atom and atom-base orders. We then repeated the shuffle 100,000 times and for each repetition, calculated the mean difference in cumulative surprisal between the two orders. This effectively helped construct the null hypothesis that there should be no between-order difference in cumulative surprisal. We also calculated the same quantities for the unshuffled data, and compared those against the null distributions for the two numerical ranges of interest. Figure 5 shows that there is no statistical significance ($p = 0.56$) to reject the null for the range 11-19, but there is high statistical significance ($p < 0.004$) in rejecting the null for the range 21-29. These results provide evidence for the idea that information front-loading is equally prominent under atom-base or base-atom orderings for smaller numerals, but it is more prominently represented in the base-atom order as opposed to the atom-base for larger numerals. Possibly due to this reason, historical changes in ordering convention of numerals tend to occur uni-directionally beyond but not below 20.

Discussion

We investigated two theories for explaining the shared tendency in the ordering of numeral expressions. We found that

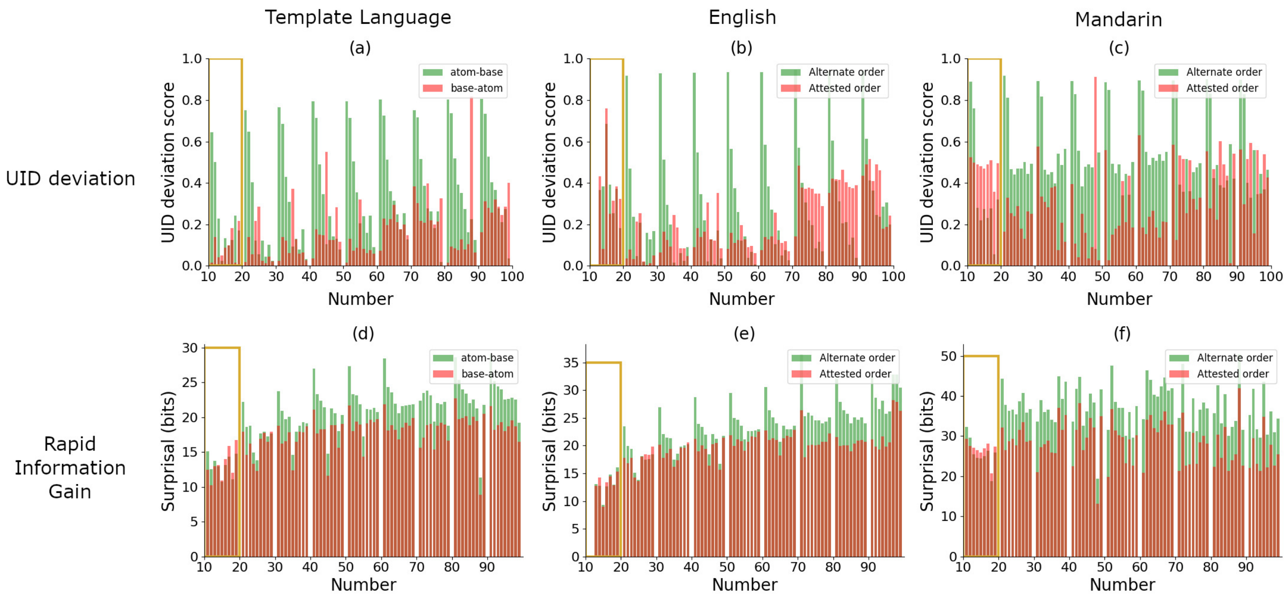

Figure 3: UID deviation (top row) and cumulative surprisal (top row) for template language, English, and Mandarin.

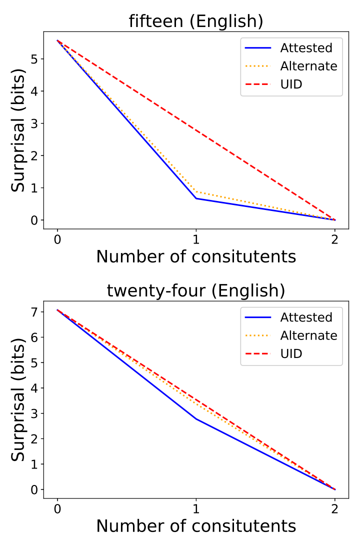

Figure 4: Information flows under alternative orders of expression for English numerals 15 and 24. The attested order for 15 is atom-base ("fifteen"), and the alternate order is base-atom ("teenfif"). The attested order for 24 is base-atom ("twenty-four"), and the alternate order is atom-base ("four-twenty"). "UID" refers to the UID theoretical prediction.

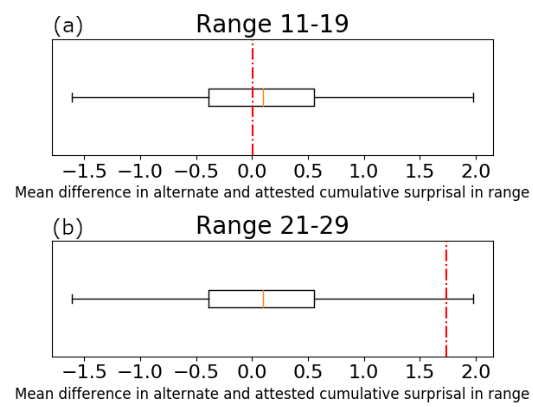

Figure 5: Permuted surprisals with attested ones marked.

the proposal of rapid information gain provides a better account for the empirical data across languages than the existing theory of uniform information density. Our findings suggest that the dominant preference toward the base-atom ordering in larger numerals reflects the need for information front-loading as opposed to information smoothing, and greater flexibility in the ordering of smaller numerals is explained partly by the fact that information flow is less affected by ordering conventions in numerals for the lower range. Our study differs from existing research in UID that focuses on information processing at the sentence level. Our emphasis is to characterize the design principles of complex lexical items, particularly compounds. This difference in the level of analysis might provide one explanation as to why UID does not predict as well in the current study. An alternative possibility is that the domain of numerals has characteristics that make a uniform information flow less desirable than information front-loading. Future research should delineate when UID might apply and when alternative principles such as RIG are more appropriate. It is also worth exploring whether the RIG principle can be applied to compounds in other domains.

Acknowledgements

We would like to thank Blair Armstrong and Suzanne Stevenson from the University of Toronto and Terry Regier from the University of California, Berkeley for their constructive comments on the manuscript. This research is supported by an NSERC DG grant and a Connaught New Researcher Award to YX.

References

Berg, T., & Neubauer, M. (2014). From unit-and-ten to ten-before-unit order in the history of English numerals. *Language Variation and Change*, *26*, 21–43.

Calude, A. S., & Verkerk, A. (2016). The typology and diachrony of higher numerals in Indo-European: A phylogenetic comparative study. *J Lang Evol*, *1*, 91-108.

Comrie, B. (2013). Numeral bases. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online.* Leipzig: Max Planck Institute for Evolutionary Anthropology.

Comrie, B., & Chan, E. (2018, Oct 30). *Numeral systems of the world's languages.* Retrieved from `https://mpi-lingweb.shh.mpg.de/numeral/`

Gibson, E., Piantadosi, S. T., Brink, K. A., Bergen, L., Lim, E., & Saxe, R. (2013). A noisy-channel account of crosslinguistic word-order variation. *Psychol Sci*, *24*, 1079-88.

Gordon, P. (2004). Numerical cognition without words: Evidence from amazonia. *Science*, *306*, 496–499.

Greenberg, J. H. (1978). Generalizations about numeral systems. In C. A. F. Joseph H. Greenberg & E. A. Moravcsik (Eds.), *Universals of human language, volume 3: Word structure* (Vol. 3, p. 249295). Stanford University Press.

Hurford, J. R. (2007). A performed practice explains a linguistic universal: Counting gives the packing strategy. *Lingua*, *117*, 773–783.

Kemp, C., Xu, Y., & Regier, T. (2017). Semantic typology and efficient communication. *Annual Review of Linguistics*, *4*, 109–128.

Levy, R. P., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In *Advances in neural information processing systems 19* (pp. 849–856).

Maurits, L., Navarro, D., & Perfors, A. (2010). Why are some word orders more common than others? A uniform information density account. In *NIPS 23.*

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., , ... Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, *331*, 176–182.

Oaksford, M., & Chater, N. (2003). Optimal data selection: Revision, review, and reevaluation. *Psychonomic Bulletin & Review*, *10*, 289–318.

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2018, 10). Exact and approximate arithmetic in an Amazonian indigene group with a reduced number lexicon. , 499-503.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*, 379–423.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*, 89-96.

Xu, Y., & Regier, T. (2014). Numeral systems across languages support efficient communication: From approximate numerosity to recursion. In *CogSci 36.*