

WHY MIGHT SOV BE INITIALLY PREFERRED AND THEN LOST OR RECOVERED? A THEORETICAL FRAMEWORK

RAMON FERRER-I-CANCHO

Complexity and Quantitative Linguistics Lab, TALP Research Center, Departament de Llinguàtiques i Sistemes Informàtics, Universitat Politècnica de Catalunya, Campus Nord, Edifici Omega, Jordi Girona Salgado 1-3, Barcelona, 08034, Catalonia (Spain)
rferrericancho@lsi.upc.edu

Little is known about why SOV order is initially preferred and then discarded or recovered. Here we present a framework for understanding these and many related word order phenomena: the diversity of dominant orders, the existence of free words orders, the need of alternative word orders and word order reversions and cycles in evolution. Under that framework, word order is regarded as a multiconstraint satisfaction problem in which at least two constraints are in conflict: online memory minimization and maximum predictability.

1. Introduction

There is converging evidence that SOV (subject-object-verb) or its semantic correlate (agent-patient-action) is a word order emerging at the very origins of language (Dryer, 2005; Pagel, 2009; Gell-Mann & Ruhlen, 2011; Goldin-Meadow, So, Özyürek, & Mylander, 2008). However, the reasons why SOV is initially preferred, and later abandoned or readopted are not well understood. This requires, in our opinion, introducing a theoretical framework.

The ordering of O, S, and V is a particular case of the ordering of a head (V) with two dependents (O and S). Let us consider a general case where a head and its n dependents (complements or modifiers)^a, must be arranged linearly (the 1st element has position 1, the 2nd element has position 2 and so on). Then, D_l , the online memory cost of placing the head in position l ($1 \leq l \leq n + 1$) can be defined as the sum of the cost of dependencies before and after the head, i.e. (Ferrer-i-Cancho, 2013b)

$$D_l = \sum_{d=1}^{l-1} g(d) + \sum_{d=1}^{n+1-l} g(d), \quad (1)$$

^aAlthough we are blindly borrowing the concept and head and dependent from syntactic theory, link direction or hierarchy are not relevant for our theoretical arguments.

where $g(d)$ is the cognitive cost of a syntactic dependency of length d , which is assumed to be a strictly monotonically increasing function of d . D_l is minimized when the head is placed at the center (Ferrer-i-Cancho, 2013b). The optimal placement of the head would change if one wished to maximize the predictability of certain elements. To maximize the predictability of the head (e.g., V), the head should be placed last while to maximize the predictability of the dependents (e.g., S and O) the head should be put first (Ferrer-i-Cancho, 2013a). Therefore, there is a conflict between minimum online memory and maximum predictability provided that $n > 1$.

2. Word order phenomena as a multiconstraint engineering problem.

2.1. *The diversity of word orders*

According to the conflicts above, it is expected that there is not a single winner in the world-wide statistics of the dominant orderings for O, S and V. N_1 , N_2 and N_3 are defined, respectively, as the number of languages whose dominant word order has the verb first, second or last. In a sample of 1377 languages, it is found $N_1 = 120$, $N_2 = 499$ and $N_3 = 569$ (Dryer, 2011), suggesting that two possible strategies for maximizing predictability and the strategy for minimizing online memory expenditure are all three represented in world languages. Verb initial orders suggest a strategy of maximizing the predictability of the subject and the object; verb final orders suggest a strategy of maximizing the predictability of the verb; central verbs suggest a strategy of minimizing online memory expenditure. Here it is not intended to establish whether the counts above indicate that a certain strategy is better than another in absolute terms for real languages (this is left for future work).

Following a similar argument it has been suggested that word order diversity could emerge from the struggle between two cognitive domains (the computational system of grammar and the direct interaction between the sensory-motor and the conceptual system) trying to impose their preferred structure on human language (Langus & Nespors, 2010). Our approach emphasizes the conflict between universal abstract principles of sequential processing and does not need to recur to any specific cognitive system or cognitive domain.

2.2. *Languages lacking a dominant word order*

The conflict between online memory and predictability implies that there is no unquestionable placement for the verb and might explain the existence of a 14% of languages lacking a dominant word order (Dryer, 2011) and why a lack of dominant order is an intermediate stage between SOV, i.e. maximum predictability of the verb and SVO, i.e. minimum online memory (Pagel, 2009).

2.3. *Word order reversions in evolution*

The conflict between online memory and predictability may shed light on word order reversions and cycles in word order evolution. A typical example is the reversion from SVO to SOV (Pagel, 2009; Gell-Mann & Ruhlen, 2011). For instance, Mandarin Chinese was originally an SOV language and became SVO; it is currently in the processes of moving back to SOV and thus displays both orders (Li & Thompson, 1981; Goldin-Meadow et al., 2008). Cycling between SOV and SVO could be interpreted as cycling between two incompatible attractors: maximum predictability of the head and minimum online memory. This cycling could be the manifestation of bistable system (Strogatz, 1994). Another example of reversion is the transition from *SVO* to *VSO/VOS* and then back SVO occasionally (Gell-Mann & Ruhlen, 2011).

We do not mean that reversions or cycles (repeated reversions) are a necessity of the conflict between online memory and predictability, rather, that this conflict is a relevant factor underlying word order back and forth changes. We are providing a hypothesis rather than a complete explanation.

2.4. *Alternative orders with a head at the center*

It is well-known that SVO is an alternative word order in languages where SVO is not the dominant word order (Greenberg, 1963). This suggests that having SVO as an alternative is a natural consequence of the conflict between maximizing predictability and minimizing online memory: if a language does not result in a dominant order with the verb at the center then it should have it as alternative to compensate the choice of a dominant order that does not comply with all the constraints.

2.5. *Verb last in computer prediction experiments.*

Computer simulation is a powerful tool for word order research (Reali & Christiansen, 2009; Gong, Minnet, & Wang, 2009). Recently, SOV has been obtained in two-stage computer simulation experiments with recurrent neural networks (learners) that have addressed the problem of the emergence of word order from the point of view of sequential learning (Reali & Christiansen, 2009). During the first stage, networks learned to predict the next element of number sequences and the best learners were selected. During the second stage, language was introduced and coevolved with the learners. The best language learners and the best learned language were selected. The best language learners had to comply with the additional constraint of maintaining the performance on number prediction of the first stage. Notice that predictability is an explicit selective pressure for the neural networks in both stages and that the languages that were selected in the second stage must have been strongly influenced by the pressure to maximize predictability. This suggests that SOV surfaces in these experiments because postponing the

verb is the optimal strategy when maximizing its predictability (Ferrer-i-Cancho, 2013a).

2.6. *The preference for head last in simple sequences and its loss in complex sequences.*

When a head has at least two arguments and the sender maximizes the predictability of the last element, the last item has to be the head (Ferrer-i-Cancho, 2013a). Recent gestural communication experiments with only one head, i.e. a verb or an action, and two modifiers, i.e. a subject or actor and an object or patient, show a preference for placing the head at the end in simple sequences (Goldin-Meadow et al., 2008; Langus & Nespors, 2010) which suggests that the predictability of the head is being maximized. Notice that the null hypothesis that there is no prior preference for head placement cannot explain this phenomenon.

A crucial finding is that this head last preference is lost in complex sequences (Langus & Nespors, 2010). The preference for head last in experiments with simple sequences and its loss in longer sequences suggests that (a) maximizing predictability dominates in short sequences (no memory-predictability conflict) while it competes with online memory for longer sequences and (b) sequence length is a critical parameter in word order phenomena. This means that predictability maximization would be the principle dominating in early stages of language evolution. To support hypotheses (a) and (b), it is needed to explain why online memory minimization would have been neglected for short sequences and recovered for longer sequences. To see this, consider that

- $n = 2$ is the minimum number of complements or modifiers needed to observe a preference for the verb to appear as the first or the last item of sequence. This is the number of complements or modifiers in the gestural communication experiments where a preference of the actor (subject) and the patient (object) to precede that action (verb) has been found (Goldin-Meadow et al., 2008; Langus & Nespors, 2010).
- In all these experiments with short sequences, the dominant order emerging is head last (more precisely the sequence actor-patient-action). It has been demonstrated that this placement of the head maximizes the dependency lengths (Ferrer-i-Cancho, 2013a).
- But as there are only $n = 2$ complements/modifiers, the cost of head last is simply $D_3 = g(1) + g(2)$, the smallest among all head last configurations with $n \geq 2$ (D_3 is indeed the minimum value of D_{n+1} for $n \geq 2$). The results of the gestural experiments suggest that the online memory cost of a head last configuration can indeed be neglected and thus only predictability matters. As expected, the preference for head last in gestural experiments involving longer sequences (for instance a main clause and a

subordinate clause) disappears, suggesting that the conflict between online memory minimization and predictability concerns specially long or complex enough sequences

- However, $n = 2$ does not seem to be enough to warrant that the online memory cost can be neglected since it has been shown above that about one third of word languages follow the *SVO* or the *OVS* order, which is a case of $n = 2$. The key is the fact that, in the gestural experiments where a preference for head last is found, elements are atomic (i.e. made of a single "word" or unit), which gives, as it will be demonstrated, a further online memory advantage with regard to the case of the ordering of subject, verb and object in word languages. So far, distance has been measured in elements (constituents) for simplicity certain elements can be made, for instance, of a subordinate clause, which happens in real languages and certain experiments (Langus & Nespors, 2010). Thus, online memory cost can be estimated more accurately if distance is measured in words (Ferrer-i-Cancho, 2008). In that case, the minimum online memory cost is achieved when elements are atomic (Appendix A). Thus, the experiments in (Goldin-Meadow et al., 2008; Langus & Nespors, 2010) where elements are atomic, follow the setup where online memory cost can be most easily neglected, and thus, not surprisingly, head last surfaces. In contrast, subjects, objects and verbs are not necessarily atomic, which may explain why central verb orders are found in about one third of world languages (including languages lacking a dominant word order) (Dryer, 2011) or in the gestural experiments with complex sequences (Langus & Nespors, 2010), despite their *a priori* disadvantages in terms of predictability of the verb. Notice that the abundance of verb last orders (N_3) does not contradict the principle of online memory cost minimization. Indeed, the relative position of adjectives and verbal auxiliaries in verb last orders can be explained in terms of online memory cost minimization (Ferrer-i-Cancho, 2008). Thus, the fact that a language has *SOV* as dominant does not mean that online memory cost minimization is inactive. Langus and Nespors (2010) attribute the preference for head last order in simple gestural experiments to a dissociation between communication and language but such dissociation is not necessary. Here we simply argue that a principle of language, i.e. online memory minimization, can be neglected in those cases and thus one is able to explain the results of these experiments through another principle of language, i.e. maximum predictability. None of this principles is specific to language.

The hypothesis of a correlation between sequence complexity and the pressure for minimizing online memory suggests the following questions:

- Why does *SVO* not appear more frequently in world languages? The statistical evidence indicating that *SVO* is the second most frequent order after

SOV has already been reviewed. This raises two questions. First, do all languages have a high sequential complexity? The stability of SOV might be higher in languages producing syntactically simpler sentences. The issue of whether world languages have the same complexity is a matter of debate in linguistics (Sampson, 2009). Second, what is the best way of measuring the abundance of a word order? It turns out that the most frequent word order, if frequency is measured in number of speakers and not in number of languages (Dryer, 2011), is SVO by far (Bentz & Christiansen, 2010).

- Why does SVO (or the symmetric OVS) not appear more clearly in the gestural experiments in (Langus & Nespors, 2010) with complex sentences? Evolving towards SVO from SOV may need (1) more time (2) more interaction between individuals and (3) more individuals than in the bounded experiments in (Langus & Nespors, 2010). We suggest that the speed of the evolution towards SVO or its accessibility may depend on at least these three factors. The need of (1) is supported by the fact the lacking a dominant word order is a transient configuration between SOV and SVO (Pagel, 2009). As for (3), notice it has been argued that the degree optimization of a language may depend on its number of speakers (Sampson, 2009; McWhorther, 2001).

Our interpretation of the preference for head last in simple sequences and its loss in complex sequences given here clearly differs from that of Langus and Nespors (2010). While they argue that the computational system of grammar has been bypassed in simple gestural experiments showing a preference for head last, our interpretation is simply that in this case sequences are short enough to make the online memory cost negligible (Appendix). In our interpretation there is no flip of systems. The principles of sequential processing are constant but the strength of a particular principle depends on the experimental conditions.

3. Final remarks

It should not be interpreted that our theoretical conflict between principles predicts that all verb placements (verb first, verb last or central verb) should have about the same abundance after discarding languages lacking a dominant word order and controlling for sentence length, number of constituents, their size and other factors. The point is that a ring backbone defines the most likely transitions between orderings of *S*, *V* and *O* (Ferrer-i-Cancho, 2013b) and thus some configurations (e.g., verb initial orders) are more difficult to reach, despite their optimality, due to the initial preference for SOV and the attraction towards SVO (Ferrer-i-Cancho, 2013b). Word orders cannot be fully explained with the individual cognitive biases considered above (Dunn, Greenhill, Levinson, & Gray, 2011; Ferrer-i-Cancho, 2013b).

Acknowledgements

The essence of the ideas above started circulating in January 2009 and was presented in the Kickoff Meeting "Linguistic Networks" which was held in Bielefeld University (Germany) in June 5, 2009. We thank the participants, specially G. Heyer and A. Mehler for valuable discussions. We are also grateful to E. Santacreu-Vasut, Simon Kirby and reviewers for comments on the current version. This work was supported by the grant BASMATI (TIN2011-27479-C04-03) from the Spanish Ministry of Science and Innovation.

Appendix: Online memory cost function

D'_l , a more accurate online memory cost function than D_l , is introduced next. Imagine that there are $n + 1$ constituents that can be made of more than one word and thus the sequence length in words is $m \geq n + 1$. The term main head or root is used to refer to the head of the head constituent, which is the head of V for the particular case of the ordering of S, V, O . Now dependencies are formed between the head word of each complement/modifier and the head of the root constituent, following the same conventions of dependency grammar (Mel'čuk, 1988). We assume that $g(d)$ is defined for $d \in [1, m - 1]$. If the main head or root belongs to the l -th constituent of the sequence ($1 \leq l \leq n + 1$), the total online memory cost of the dependencies between the root and the heads of its n complements/modifiers may be defined as $D'_l = \sum_{i=1}^{l-1} g(d_{i,l}) + \sum_{i=l+1}^{n+1} g(d_{i,l})$, where $d_{i,j}$ (with $i, j \in [1, n + 1] \subset \mathbb{N}$) is the distance in words between the head word of the i -th constituent and that of the j -th constituent. If constituents are made of a single word, one has $d_{i,j} = |i - j|$ and thus $D'_l = D_l$ (recall Eq. 1). The point is that given a sequence of constituents, D'_l is minimum when constituents are atomic (i.e. constituents are made of a single word; in that case, the only word of the constituent is the head). To see it, assume $i \neq l$ and notice that $d_{i,l} \geq |i - l|$ (with equality if and only if the i -th constituent, the l -th and the intermediate constituents are atomic) and $g(d_{i,l}) \geq g(|i - l|)$ (since $g(d)$ is a monotonically increasing function of d).

References

- Bentz, C., & Christiansen, M. H. (2010). Linguistic adaptation at work? the change of word order and case system from Latin to the Romance languages. In A. Smith, M. Schouwstra, B. de Boer, & K. Smith (Eds.), *Proceedings of the eight international conference on the evolution of language* (p. 26-33). London: World Scientific.
- Dryer, M. (2005). Order of subject, object and verb. In M. Haspelmath, M. S. Dryer, D. Gil, & B. Comrie (Eds.), *The world atlas of language structures*. Oxford: Oxford University Press.
- Dryer, M. (2011). Order of subject, object and verb. In M. Dryer & M. Haspelmath

- (Eds.), *The world atlas of language structures online*. Munich: Max Planck Digital Library. (Available online at <http://wals.info/chapter/81>. Accessed on 2013-04-23.)
- Dunn, M., Greenhill, S. J., Levinson, S. C., & Gray, R. D. (2011). Evolved structure of language shows lineage-specific trends in word-order universals. *Nature*, *473*, 7982.
- Ferrer-i-Cancho, R. (2008). Some word order biases from limited brain resources. A mathematical approach. *Advances in Complex Systems*, *11*(3), 393-414.
- Ferrer-i-Cancho, R. (2013a). The placement of the head that maximizes predictability: an information theoretic approach. *submitted*.
- Ferrer-i-Cancho, R. (2013b). The placement of the head that minimizes online memory: a complex systems approach. *Language Dynamics and Change*, in press. (<http://arxiv.org/abs/1309.1939>)
- Gell-Mann, M., & Ruhlen, M. (2011). The origin and evolution of word order. *Proceedings of the National Academy of Sciences USA*, *108*(42), 17290-17295.
- Goldin-Meadow, S., So, W. C., Özyürek, A., & Mylander, C. (2008). The natural order of events: how speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences*, *105*(27), 9163-9168.
- Gong, T., Minnet, J. W., & Wang, W. S.-Y. (2009). A simulation study of word order bias. *Interacion Studies*, *10*(1), 51-76.
- Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. H. Greenberg (Ed.), *Universals of language* (p. 73-113). London: MIT Press.
- Langus, A., & Nespors, M. (2010). Cognitive systems struggling for word order. *Cognitive Psychology*, *60*(4), 291-318.
- Li, C. N., & Thompson, S. A. (1981). *Mandarin Chinese: a functional reference grammar*. Berkeley: University of California Press.
- McWhorther, J. H. (2001). *The power of Babel*. New York: Times Books.
- Mel'čuk, I. (1988). *Dependency syntax: theory and practice*. Albany: State of New York University Press.
- Pagel, M. (2009). Human language as a culturally transmitted replicator. *Nature Reviews Genetics*, *10*(6), 405-415.
- Reali, F., & Christiansen, M. H. (2009). Sequential learning and the interaction between biological and linguistic adaptation in language evolution. *Interacion Studies*, *10*, 5-30.
- Sampson, G. (2009). A linguistic axiom challenged. In G. Sampson, D. Gil, & P. Trudgill (Eds.), *Language complexity as an evolving variable* (p. 1-18). Oxford: Oxford University Press.
- Strogatz, S. H. (1994). *Nonlinear dynamics and chaos: With applications to physics, biology, chemistry and engineering*. Reading, MA: Perseus Books.