Computational constraints on compositional interpretation: Refocusing the debate on language universals

Elsevier Editorial System(tm) for Lingua
Manuscript Draft

Manuscript Number: LINGUA-D-10-00018R1

Title: Computational constraints on compositional interpretation: refocusing the debate on language universals

Article Type: Full Length Article

Corresponding Author: Dr. Susan Rothstein,

Corresponding Author's Institution:

First Author: Susan Rothstein

Order of Authors: Susan Rothstein

Suggested Reviewers:

***Detailed Response to Reviewers**

We have corrected the typos and made the explanations required.

1) Title Page (WITH Author Details)

Computational constraints on compositional interpretation: refocusing the debate on language universals

Susan Rothstein[a*] and Alessandro Treves[b]

Gonda Brain Research Center[a]
Bar-Ilan University
Ramat Gan 52900
Israel
rothss@mail.biu.ac.il

SISSA – Cognitive Neuroscience Sector[b]
via Beirut 2,
34151 Trieste,
Italy
ale@sissa.it

*Corresponding author

Computational constraints on compositional interpretation: refocusing the debate on language universals

Abstract: We argue that the debate on language universals should be directed away from the discussion as to whether typological diversity is or is not an argument against the existence of language universals. Instead, given our growing awareness of the fact that the neural mechanisms underlying language use are the same as those underlying other cognitive functions in both humans and mammals, the central question for cognitive neuroscientists and linguists is what neural mechanisms can facilitate compositional interactions, and how the range of grammatical structures and the ability to use language creatively emerges from a much narrower range of neural mechanisms. We suggest two complementary methods for investigating these issues, one more linguistically oriented and one more computationally oriented, and present preliminary results from investigations concerning the expression of the mass-count distinction crosslinguistically, using both methodologies. These results suggest that universality in language might express itself at the deeper level of the computational operations involved in the processing of language, rather than in the results of those computations.
Keywords: linguistic universals, compositionality, neural computation, mass-count distinction, information measures

1. The vicious circle of unanalysed data

Evans and Levinson's (2009) paper makes the important point that there is immense structural diversity in terms of syntax, phonology and morphology in the world's languages. Freidin (2009) and Pesetsky (2009) in their commentaries in BBS make the equally important point that taking structural differences at face value may miss underlying structural similarities. This reflects the way that the debate over universals is overshadowed by the tendency for "universalists" to propose universals, typologists to produce counterexamples to refute them, and universalists then to reject the counterexamples as "unanalysed data". Since many of the world's 6000-8000 languages are undocumented or unanalysed, and each of them takes decades, if not a lifetime, to study thoroughly, the chances of all "unanalysed data" turning into analysed data are minimal, and there is no way out of the circle of the debate with the rules set as above.

2. A neural computation oriented approach to the issue of language universals

Basic observations in comparative neuroscience suggest a way to reorient the discussion on universals. Evidence accumulated over the past century, but little noticed by linguists, indicates that there is no dedicated neuronal machinery to subserve complex cognitive capacities, least of all machinery specialized for language. Our abilities to learn, produce and interpret language are supported by the same neural mechanisms, at the cortical microcircuitry level, as any other cognitive function, including those (most) that we share with other mammals. But while something is known about the neural mechanisms underlying, for example, visual and auditory processes, we know next to nothing about how the nervous system processes information about the relation between sounds and symbols and how complex meanings are represented compositionally in the brain. We can infer the elementary cortical transactions to be the same as in better understood functions, but how they sum up to produce the faculty of language, and why they should do so only in the human species, remains unclear. The mystery may stem precisely from the nature of language as an emergent property, its being more than the sum of its parts, arising, that is, by the system-level combination of elementary cortical operations. From such a perspective, the central question for cognitive neuroscientists is what neural mechanisms can facilitate compositional interactions, and how the range of grammatical structures emerges from a much narrower range of neural mechanisms. Further, what is the nature of the plasticity which allows children to acquire a native-speaker linguistic competence in a finite time; that is, how do children learn to manipulate a finite set of symbols in such way that they can produce and interpret an infinite number of novel strings and thus convey and comprehend new information (Chomsky 1965, 1981). Evans and Levinson's (2009) paper does not (as they explicitly state) say anything about what has become known as "the logical problem of language acquisition".

3. Compositionality at the level of syntax/semantic/conceptual representation

From the perspective just described, it seems to us that universal properties of language are to be found precisely among those characteristics which Evans and Levinson (following Greenberg et al. 1963) discount as features that are universal by definition (page 20), including 'discreteness, arbitrariness, productivity' and the ability to combine arbitrary phonological symbols into meaningful elements. These features may well be implicit in the conventional definition of what a language is, but they are not trivially reducible to neuronal operations. If they were, how come animals with a nearly identical nervous system do not speak? The property of language which appears to pose the most serious computational challenge to the nervous system is in fact missing from Evans and Levinson's list: it is what we may call *hierarchical compositionality at the level of the syntactic/semantic interface*. It is the property of combining abstract elements, each associated with a meaning, into a complex symbol whose meaning is dependent on the meaning of its parts, and then combining these complex symbols into more complex entities whose meaning is yet again predictable on the basis of its immediate constituents. As far as we can see, assigning meaning to sentences in this way is a feature of all languages, and even if a language displays no evidence of syntactic constituent structure and is truly non-configurational, semantic construal requires combinatorics with respect to meaning. Hierarchical compositionality of this kind is a "universal by definition"; it is a property which may well set aside language and language manipulation from a range of other cognitive capacities. This is in essence the problem of recursion at a more abstract (and pre-linguistic) level, and while there may be languages which do not involve sentence embedding and thus do not involve that peculiar type of recursion, it seems that all languages require some form of compositionality in this general sense. Even languages cited like Straits Salish (Jelinek 1995), analysed as consisting of n-place categorially neutral predicates which can combine with subject and object clitics and with arguments derived from predicates via a form of relative clause construction, use hierarchical compositionality and limited recursion as the basis of their grammar.

4. Compositionality can only be analysed in sufficiently large systems

A limit on many approaches that have been proposed may be their relying on toy cases of extremely reduced complexity, and on the assumption that any neural mechanism can be later scaled up to address real-life tasks. Reducing the complexity of natural behavior to laboratory manageability is common practice in cognitive science, but it may prevent us from revealing properties that emerge only above a critical size or complexity, out of a phase transition (a sudden qualitative change in behaviour at the slight quantitative change of a parameter describing the system). Compositionality may be related to cortical latching dynamics, the ability of a large cortical network to hop indefinitely from state to state, spontaneously, when its activity is not clamped by an incoming external input. Latching dynamics indeed appears to emerge only beyond a phase transition (Treves, 2005). Yet, the neural underpinnings of compositionality remain an object of rarefied speculation. In contrast, the range of linguistic solutions that have evolved in order to compose meaningful utterances is truly impressive, as the target article illustrates, but simply observing clusters of solutions, the *attractors* that Evans and Levinson refer to (in a wider sense, cp. Treves, 2005), does not advance our understanding of the computational question, how can compositional structures be processed at the neuronal level. Various exciting proposals have been put forward (Pulvermüller 2002; van der Velde & de Kamps, 2006; Gayler, 2003, 2006; Eliasmith and Thagard, 2001; Hashimoto, 2008), often as neuronal mechanisms for identifying syntactic structure, but they do not yet satisfactorily solve the problem of matching semantic with syntactic representations, and thus leave open the question of how to match meanings with neural representations. Again, the core issue is what neural mechanisms can facilitate compositional interpretation, and it appears necessary to address it using model systems of sufficient complexity, where solutions can be imagined to scale up to real-life problems. What is the mapping between syntactic structure (i.e. grammatical expressibility) and semantic or conceptual structure? A further question is how the grammaticalisation of real-world conceptual distinctions such as singularity/plurality, mass/count, and distinctions in gender, takes place. These distinctions are at their root prelinguistic and thus universally available, but their grammaticalisation emerges differently across languages, and raises the question of exactly how the language/world mapping is neurally encoded, where the variation comes in and just how wide the variation can be.

There seem to us to be two complementary ways to address the issue of the possibility of common structure underlying crosslinguistic variation. Assume that different grammatical structures can be seen as reflecting different solutions to the problem of the neural mechanics of composing elements of meaning into 'bigger' chunks of meaning. The more linguistically oriented approach is to identify crosslinguistically a range of grammatical solutions to the compositional problem posed by the mapping of a particular semantic construction or operation into a syntactic structure, and to try and identify what computational properties these solutions have in common. If there are neurally-based computational constraints on semantic composition, semantic operations underlying parallel grammatical phenomena may show similar properties, even if the grammatical expression of these computational processes is very different from language to language. These similarities would give insight into how the language computation actually works. The more computationally oriented way is to identify biologically plausible ways of computing the information necessary for linguistic performance, and to construct models of how these computational mechanisms could interact with naturally available data, both in situations of acquisition and of ‘normal language use’. If neural mechanisms prove to constrain grammatical devices, then this may lead to evidence for universals of language in a cognitively plausible way.

5. Examples

In this section we present an example of each of the proposed methodologies for approaching the issue of language universals. Together, they address the grammatical realisation of the contrast between counting and measuring:

(i) grammatical expression of counting vs measuring:
Semantic research (Carlson 1977, Chierchia 1998, Landman 2004, Rothstein 2009) has clarified that there are two distinct grammatical operations involving number. The one is **counting,** putting individuable entities in one-to-one correspondence with the natural numbers as in *three girls*, and the other is **measuring**, ascribing a value on a dimensional scale to a quantity. In languages with a distinction between mass and count nouns, counting only requires a number and a sortal nominal. Depending on the theory, the sortal either presupposes the atomicity of the entities in its denotation (Landman 2004, Chierchia 1998) or grammatically encodes it (Krifka 1989, Rothstein 2010). Measuring requires a number and a unit or measurement as in *two kilos of flour*, *200 kilos of furniture* and ignores any atomic structure that there may be in the denotation of the predicate. Grammatically, in a language with a mass/count distinction, counting is an operation which applies to count nouns, and measuring is an operation which applies to mass nouns. In complex nominals, the distinction between counting and measuring phrases is normally blurred. *Two glasses of water* and *three boxes of books* are ambiguous between a counting reading illustrated in (1), in which the containers, filled with water or books, are physically present, and a measuring reading in (2), in which the nominals denote water to the quantity of two glasses and books to the quantity of two boxes:

(1) a. She brought two glasses of water for our guests
b. The post office complained about the weight of three boxes of books.

(2) a. Add two glasses of water to the soup!
b. Three boxes of books won’t fit on the shelf.

Rothstein (2009) shows that classifier expressions expressing measuring and counting have different syntactic analyses in English and Modern Hebrew. In counting expressions, the classifier is analysed as a head taking a complement and the complement is a referential expression. The numeral directly modifies the complex NP in a structure directly parallel to the simple expression *three girls*.

(3) a. [three [girls]]
b. [three [boxes of books]]

In measuring expressions, the classifier combines first with the numeral to form a complex predicate which modifies the complement NP, itself analysed as a predicate nominal. This is analogous to the structure proposed for the explicit measure expressions in (4b).

(4) a. [three boxes] of books
b. [three kilos] of flour

This contrast is made explicit in Modern Hebrew which has two standard complex nominal forms, an absolute form (5a) where the nominal head is in root form and takes a PP complement comprised of P and a referential nominal (allowing modification, quantifiers and etc). The second (5b) is a construct state nominal, in which the head is reduced phonologically, and the complement is a bare NP (Borer 1999, 2009)

(5) a. šaloš kosot šel mayim
three cups of water

b. šaloš kosot mayim
three cups water

(5a) has only the counting interpretation, while (5b) is ambiguous between counting and measure reading. Rothstein (2009) argues that this is because (5b) is an underspecified syntactic structure, which allows analysis analogous either to (3) or to (4), while (5a) can only have the structure in (3). Definite numerical construct states must be analysed for independent reasons as in (6) and do not allow numeral and classifier to form a constituent. In these constructions, predictably, the measure reading is impossible.

(6) [šlošet] [bakbukey ha-mayim]
three bottles DEF-water
"The three bottles of water" (definite: multiple embedded construct state forms)

In constructions analogous to (6), where the classifier is a measure expression such as *kilo,* the compositional demands of the syntax and semantics conflict: definite construct state syntax requires the classifier to be construed as a constituent with the complement, but the semantics of the measure head *kilo* requires the classifier to be construed as a constituent together with the number. The result is that the computation is blocked and the construction is ungrammatical. There is no way to express directly in Hebrew e.g. *the three litres of wine that I drank/the three kilos of flour that I ordered* although the indefinite is perfectly acceptable. This is evidence that the grammatical expression of measuring involves number and classifier to be construed as a complex predicate, while in counting constructions, the head is individuating. Preliminary investigations into Russian and French suggest that these languages also show grammatical contrasts between counting and measuring constructions, although the grammatical indications differ in each case. In languages without a mass/count contrast, classifier systems are used to count as well as to measure as in (7) (from Mandarin):

(7) a. sān gè rén
three CL man
“three men”

b. liăng gōngjīn píngguŏ

two kilo apple
"two kilos apple"

However, preliminary results (XP Li, in progress) shows that in counting contexts such as (8a), Cl + N are construed as a constituent, while in measuring expressions such as (8b) Num + Cl are construed as a constituent, indicating syntactic contrasts between counting and measuring expressions which parallel the results for English and Hebrew.

These results indicate some common representation of the measuring and counting computations at the level of modifier/modificand relations, although the grammatical expression of the relation may differ greatly from language to language depending on the morphosyntactic structure of the language. This suggests that crosslinguistic commonalities in counting and measuring systems will be found at a more abstract level of computational representation than linguists usually use to describe morphosyntactic representations. It points towards a level at which neural computations and semantic compositions might be matched, and suggests a level of abstraction at which we should look to model the neural representation of grammatical structure.

(ii) Crosslinguistic diversity of the mass count distinction

Crosslinguistic variation in the expression of the mass/count distinction has been a puzzle for cognitive science and linguistics. On the one hand, there is a clear link between cognitive mechanisms of individuation and linguistic mechanisms for counting. Count nouns naturally denote individuable or bounded entities with stable spatial properties across time (e.g. *cat, boy, table, book*), while mass nouns are associated with substances which take their spatial dimensions from containers (e.g. *water, mud*), or whose physical boundedness varies over time or depends on the artefact constructed from it (e.g. *wood, gold*). (See also Soja, Carey and Spelke 1991, Prasada, Ferenz and Haskell 1994, Gathercole 1986). On the other hand, there is no direct association between count terms and individuable objects or mass terms and substances, and it has always been claimed that there is wide crosslinguistic variation in how the mass/count distinction is played out grammatically. This raises two questions:

(i) how different is the expression of the distinction crosslinguistically, i.e. to what degree is the mass/count distinction an arbitrary grammaticalisation, rather than a straightforward expression of a conceptual distinction

(ii) how is the grammatical distinction derived computationally from a conceptual distinction (if it is not a direct reflection of it).

We hypothesize that, in languages that do make a distinction between mass and count nouns, the exact way in which a noun should be used syntactically (whether it can be pluralized, whether it admits exact or indeterminate quantifiers, etc.) should be related to basic semantic features of the noun – with some variability across languages arising from differences among the features that determine syntactic usage – apart from a few exceptions, borrowings from other languages and other special cases. Of course, in no language do we expect the mass/count distinction to be simply a distinction between two groups of nouns, those that are countable and those that are not. Still, it should be possible to specify the correct syntactic usage of each noun via a sufficient number Q_L of binary questions, which can be designed in each language L to lay out how individual nouns behave in the mass/count domain. If we represent the yes/no answers for a particular noun as components of a vector, each noun is associated with a binary usage vector with Q_L [0,1] components. Distinct usage vectors define equivalence classes, that is, nouns associated with the same vector behave identically in language L, and mass/count syntactic categories in that language are described as a partition into a maximum of 2^{Q_L} (2 to the power Q_L) classes. A neurally plausible mechanism for using a particular language correctly, then, posits that the usage vector for each noun can be derived from a universal semantic vector describing the features of the noun, via a simple feedforward network with language-specific synaptic weights.

For this hypothesis to be plausible, the nouns in different languages should cluster into usage vectors in roughly similar ways. We naively expect 2 main classes that largely correspond across languages, the class of the prototypical mass nouns and that of the prototypical count nouns – the sharp edges of the continuum – and several smaller classes that would show considerable variability from language to language. Further, assuming prototypical mass and count nouns to occur with roughly similar frequency in our sample, we expect about one *bit* of entropy, in the partition of nouns across usage classes, to reflect the basic mass/count distinction, and to be shared across languages; and additional fractions of a *bit* to reflect smaller usage classes around the middle of the continuum, and to be shared only among syntactically related languages. This is simply a quantitative way to state that most of the mass/count-related syntax should just express common semantic cognition, with some additional language specific "decorations". If this naïve hypothesis were to be correct, different weighting of possibly 10-20 semantic features might account for language specific instantiations of the mass/count distinction. Then acquiring those weights, e.g. by associative learning, would be equivalent to acquiring the correct mass/count usage in a given language, starting from an effectively universal neuronal representation of the meaning of nouns encoded as distributed cell assemblies, articulated in terms of their features.

Our empirical assessment of this naïve hypothesis has led us to discard it. In work that will be described in a separate publication (Kulkarni, Rothstein and Treves, in preparation), we find that the correspondence among the mass/count syntactic usage vectors across a sample of 5 languages is limited – when focusing on pairwise correspondence. When considering triplets of language the correspondence, as measured by 3-way Shannon information, (a somewhat technical measure of the similarity of the classes across three languages) it is virtually null: the grammaticalisation of the mass/count distinction, when analysed statistically, appears to diverge, even among historically-related languages.

6. Conclusions

We see that two complementary approaches, one linguistic the other statistical, apparently lead to somewhat incongruent assessments of the universality of a specific language feature, the mass/count distinction. However, they both indicate that universality in language might express itself at the deeper level of the computational operations involved in the processing of language, rather than in the results of those computations. What the two approaches also have in common is their pointing to the need to go beyond the sterile philosophical debate about whether a particular item should be regarded as universal or not, framed in the binaries argumentation based on examples and counterfactuals, and to start assessing quantitatively the degree of universality of specific aspects of language, doing justice to the infinite levels of complexity of this fascinating cognitive capacity.

Bibliography

Borer, H., 1999. Deconstructing the construct. In: Johnson K., Roberts, I., (Eds.), Beyond Principles and Parameters, Kluwer publications (Springer), Dordrecht, pp. 43-89.

Borer, H., 2009. Compounds: the view from Hebrew. In: Lieber R., Stekauer, P. (Eds.), The Oxford Handbook of Compounding, Oxford, Oxford University Press, pp491-511.

Carlson, G., 1977, Amount relatives. Language 53, 520-542.

Chierchia, G., 1998. Plurality of mass nouns and the notion of 'semantic parameter'. In: Rothstein S., (Ed.), Events and Grammar, Kluwer publications (Springer), Dordrecht, pp.53-104.

Chomsky, N., 1965. Aspects of the Theory of Syntax. The MIT Press, Cambridge, MA.

Chomsky, N., 1981. Principles and parameters in syntactic theory. In: Hornstein, N., Lightfoot, D., (Eds.), Explanations in Linguistics: The Logical Problem of Language *Acquisition.* Longman, London and New York, pp. 32-75.

Eliasmith, C., Thagard, P., 2001. Integrating structure and meaning: A distributed model of analogical mapping. Cognitive Science 25, 245–286.

Evans, N., Levinson, S., 2009. The myth of language universals: Language diversity and its importance for cognitive science. Behavioral and Brain Sciences 32, 429-448.

Freidin, R., 2009. A note on methodology in linguistics. Commentary on Evans and Levinson. Behavioral and Brain Sciences 32, 454-455

Gathercole, V., 1986. Evaluating competing linguistic theories with child language data: The case of the mass-count distinction. Linguistics and Philosophy, 151-190.

Gathercole, V., 1997. The linguistic mass/count distinction as an indicator of referent categorization in monolingual and bilingual children. Child Development 68, 832-842.

Gayler, R. W., 2003. Vector symbolic architectures answer Jackendoff's challenges for cognitive neuroscience. Paper presented at the Joint International Conference on Cognitive Science, University of New South Wales, Sydney, Australia

Gayler, R. W., 2006.Vector symbolic architectures are a viable alternative for Jackendoff's challenges. Behavioral and Brain Sciences. 29, 78-79.

Greenberg, J. H., Osgood C. E., Jenkins J. J., 1963. Memorandum concerning language universals. In: Greenberg J. H. , (Ed.), Universals of Language. The MIT Press, Cambridge, MA. ppxv-xxvii.

Hashimoto, T., 2008. Consideration of language evolution through cognitive modeling of grammaticalization (in Japanese). Collaboration Report of *The Institute of Statistical Mechanics*, No.217 "Information Theory of Dynamical Systems 7 - Dynamism of Natural Language", pp. 13-21, 2008

Jelinek, E., 1995. Quantification in Straits Salish. In: Bach, E., Jelinek, E., Kratzer A., Partee, B., (Eds.) Quantification in Natural Languages, Kluwer publications (Springer), Dordrecht, pp. 487-540.

Krifka M., 1989. Nominal reference, temporal constitution and quantification in event semantics. In: Bartsch, R., van Bentham, J., van Emde Boas, P., (Eds.) Semantics and Contextual Expressions, Foris, Dordrecht, pp 75-155.

Landman, F., 2004. Indefinites and the Type of Sets. Blackwell, Oxford.

Li, XP., in progress: The Semantics of Classifiers in Chinese. Ph.D.dissertation. Bar-Ilan University.
Pesetsky, D., 2009. Against taking linguistic diversity at “face value”. Commentary on Evans and Levinson. Behavioral and Brain Sciences 32, 464-465.
Prasada, S., Ferenz, K., Haskell, T., 2002, Conceiving of entities as objects and as stuff. Cognition, 83, 141-165.
Pulvermüller, F., 2002. A brain perspective on language mechanisms: from discrete neuronal ensembles to serial order. Progress in Neurobiology 67, 85-111
Rothstein, S., 2009. Individuating and measure readings of classifier constructions: evidence from Modern Hebrew. Brill Annual of Afroasiatic Languages and Linguistics 1, 106-145.
Rothstein, S., 2010. Counting and the mass - count distinction. Journal of Semantics, in press.
Soja, N., Carey, S., Spelke, E., 1991. Ontological categories guide young children’s induction of word meaning: Object terms and substance terms. Cognition 38, 197-211.
Treves, A., 2005. Frontal latching networks: a possible neural basis for infinite recursion. Cognitive Neuropsychology. **6:** 276–91
van der Velde, F., de Kamps, M., 2006. Neural blackboard architectures of combinatorial structures in cognition. Behavioral and Brain Sciences 29: 37-70.