Neuroscience and Biobehavioral Reviews 80 (2017) 646–656

Contents lists available at ScienceDirect

Neuroscience and Biobehavioral Reviews

journal homepage: www.elsevier.com/locate/neubiorev

Meta-analysis

Reviewing the functional basis of the syntactic Merge mechanism for language: A coordinate-based activation likelihood estimation meta-analysis

Emiliano Zaccarella*, Marianne Schell, Angela D. Friederici

Max Planck Institute for Human Cognitive Brain Sciences, Department of Neuropsychology, Stephanstraße 1a, 04103 Leipzig, Germany

ARTICLE INFO

Keywords:
Merge
Language
Meta-analysis
ALE
Pars opercularis

ABSTRACT

The ability to create structures out of single words is a key aspect of human language. This combinatorial capacity relies on a low-level syntactic mechanism—Merge—assembling words into hierarchies. Neuroscience has explored Merge by comparing syntax to word-lists. Here, we first review potential issues with the word-lists materials. We then perform an activation likelihood estimation (ALE) on the reported foci, to reveal functional convergence for Merge at whole-brain level. Finally, we run probabilistic tractography on an independent population to observe how these convergent activations anatomically connect. Functionally, we found that when confounding activity was removed, consistency for Merge was only observable in the left pars opercularis (BA44) and in the inferior part of the posterior superior temporal sulcus/gyrus (pSTS/STG; BA22). Structurally, we could confirm that the two regions are connected through dorsal fiber bundles. We therefore suggest that the cortical implementation of linguistic Merge consists of a left fronto-temporal interaction between BA44 (syntactic processor) and pSTS/STG (integrative processor), which communicate to each other along dorsal white matter fascicles.

1. Introduction

In our everyday life we are constantly faced to linguistic expressions we have never encountered before. Although such expressions can be new to us and might vary in terms of complexity, we are always immediately able to identify what their structures and related meanings might be. This adaptive linguistic capacity appears to be grounded on a very basic computation—known as Merge—which creates complex structures out of single words according to the grammar of the language in use (Chomsky, 1995). Because Merge constitutes the cognitive basis upon which our linguistic competence is founded, the understanding of its cortical implementation is strongly advocated by both linguistic theory (Rizzi, 2012), evolutionary linguistics (Bolhuis et al., 2014) and neurobiology of language (Zaccarella and Friederici, 2016). Merge is taken to be the universal fundamental structure-building computation of natural language syntax (Berwick et al., 2013; Chomsky, 1995). We can formally represent this computation as: $\alpha\ \beta \rightarrow \{\alpha\ \beta\}$, which means "take two elements α and β, and string them together to form a new set containing both". The two elements in the input, α and β can be two lexical items (for example *the* and *man*)—which together form a bigger object {*the man*}. Depending on the relationship between the categories involved, sets are labeled and the hierarchy is established, e.g. *the man* is a determiner phrase (DP). As such, linguistic sequences are hierarchical assemblies of words forming more composite phrasal and sentential constituents (represented in square brackets below) which may in turn be part of bigger ones: *[the man], [[the man] [eats [an apple]]]*, etc. At the most atomic level, words—the basic building blocks of linguistic structures—can be distinguished into two different categories: content class categories and functional class categories. Content words are those items primarily carrying lexical-semantic information, which serve descriptive content and referential weight, and are context-independent. They refer to events (either states or actions) and entities participating in them (Baker, 2003). Content words typically are nouns, adjectives, verbs and adverbs, and are often called members of the open class since the class can increase its members. Function words, conversely, primarily carry syntactic information, since they have reduced semantic content, and work for structural assignment by linking other items. They are therefore context-dependent. Function words typically include prepositions, pronouns, determiners, conjunctions and auxiliary verbs, and are also called members of the closed class since they do not increase in number. The way content and function words combine together under Merge is essential to the construction of linguistic sequences, in which the hierarchical relationships are strictly established according the syntactic nature of the words entering the computation. In this sense, a determiner cannot combine with another determiner (*the the*) since it would not be possible to establish a hierarchical

* Corresponding author.
E-mail address: zaccarella@cbs.mpg.de (E. Zaccarella).

http://dx.doi.org/10.1016/j.neubiorev.2017.06.011
Received 24 February 2017
Available online 23 July 2017

relationship in which one of the two elements would dominate the other. Conversely, a determiner can successfully combine with a noun (*the man*) since an asymmetrical relationship between the two elements exists, in which the determiner dominates the noun. At the neuroanatomical level, the functional studies interested in the cortical implementation of Merge have taken direct advantage of the possibility to generate lists of words that do not form structures, to compare the load generated by simple word-string processing (*the the*) against the load generated by syntactic-building processing (*the man*). This type of experimental manipulation allows the direct investigation of Merge while leaving both stimulus length and lexical access balanced across conditions. The possibility to use this type of contrast has stimulated the production of a growing number of functional studies, which have been published in the last twenty-five years in the neurolinguistic literature, across languages and modalities, using both Positron Emission Tomography (PET) and functional Magnetic Resonance Imaging (fMRI). See Appendices A and B for a detailed list of the studies. Across the experiments, all the perisylvian language network and neighboring regions have been basically found to participate in sentence processing compared to word-list processing, including Broca's area (BA44/45) in the left inferior frontal cortex; the frontal operculum/anterior insula (FOP/aINS); the anterior portion of the left temporal lobe (BA38); the posterior portion of the left superior temporal sulcus/gyrus (pSTS/STG; BA22). As a matter of fact, a review on the neurological distribution of formal syntactic operations could not report any functional characterization for Merge, as no consistency across the data was available at the time (Grodzinsky and Friederici, 2006).

In the attempt to overcome such inconsistency across the data, a very recent meta-analysis used the frequency of localization of functional activity of different linguistic operations across the studies (label-based distribution meta-analysis), to offer a quantitative measurement of the neuroanatomical localization of the linguistic processes beyond single word level (Hagoort and Indefrey, 2014). The reliability for a particular region to be considered as active was assessed by obtaining a relative binomial distribution based on the average number of activated regions reported per experiment divided by the total number of brain regions taken in the analysis (n = 112). Interestingly, the authors included in the report a subset of studies contrasting sentence-level processing with lower-level control conditions, ranging from resting conditions, or fixation of a hair cross, or word-lists—a subgroup (n = 15) of which is included in the present study. Compared to control conditions below sentence level, the meta-analysis found that the temporal lobes and the posterior IFG bilaterally were significantly more active than the resting condition, with clear left hemisphere dominance. Moreover, large parts of the right parietal and right inferior temporal cortices were not found to be active, which led the authors to conclude that these areas are thus not reliably involved during sentence comprehension. Within the word-lists control condition set, they reported activity in the temporal cortex—the anterior/superior temporal gyrus (BA38/22), the middle temporal gyrus (BA21), the posterior/superior temporal gyrus (BA21/22)—in the inferior frontal cortex—pars triangularis (BA45), pars orbitalis (BA47)—and in the frontal cortex—BA8/9. Importantly, they noted that the most dorsal part of the IFG (pars opercularis; BA44), was not reliably activated during passive listening to simple sentences, which lead them to hypothesize that this area is either not involved in any sentence-level combinatorial process, or that this process is not necessarily active for passive listening because of good-enough processing strategies not requiring full compositional analysis (Ferreira et al., 2002).

Compared to label-based distribution meta-analysis, the activation likelihood estimation (ALE) analysis (Laird et al., 2005; Turkeltaub et al., 2002) calculates the degree of agreement across studies by modeling the activated foci as centers of a Gaussian probability distribution, rather than as single points. The three-dimensional Gaussian maps are then summed to one other following several thousand of random iterations, whose final output would correspond to p-value-based statistical parametric images. On these images, statistically significant activation for a specific voxel then can be calculated following classical threshold methodologies. The main advantages to use such qualitative, objective coordinate-based algorithm are essentially three: first, concordance is examined at voxel-level using standardized local information, compared to label-based meta-analyses which use instead anatomical labels assigned to specific locations. As such, the ALE analysis is not susceptible to errors due to the use of generalized location information taken from different atlases or the adoption of anatomical labels, which are assigned by the individual investigators to the reported foci. Second, ALE methods make use of subject size information for each foci group to calculate how blurred the Gaussian function for a specific focus has to be (the Full-Width Half-Maximum; FWHM) to more realistically implement the variable uncertainty. In this respect, while label-based meta-analyses give the same activation weight to two studies differing in terms of experimental sample size and power effect (e.g. 3 subjects vs. 18 subjects), the ALE algorithm will assign a higher weight (a tighter and taller Gaussian) to the foci of the study with a bigger sample size, compared to the ones belonging the study with a smaller sample size. Third, ALE random effects methods allows testing for qualitative differences across groups of studies, by pooling the studies in subgroups and measuring the corresponding source of variation based on such qualitative differences.

With respect to word-list studies taken into account in the present study, one qualitative difference, which can be measured by ALE methods concerns with the nature of the word-list control condition used in the specific studies. A recent review on the brain basis of language processing, Friederici (2011) noticed that the word-list conditions often employ both content words (e.g. nouns, verbs, adjectives) and function words (e.g. determiners, prepositions, conjunctions) together, which may erroneously enhance rather than decrease the construction of minimal structures in the non-syntactic conditions, for example: "*Money* the [*the client washed*]" (Kuperberg et al., 2000); "[*Her eyes during close*] *the she ceremony*" (Vandenberghe et al., 2002). Crucially, these studies report high involvement of temporal regions, compared to the infero-frontal regions, which are conversely not found as active during subtraction analysis in the same subjects. One hypothesis would then be that if control conditions consist of remaining syntactic chunks enhancing low-level syntactic processing—because of the use of both content and function words together—then the comparison of licit structures against word-lists might have removed the Merge effect, given the high automaticity of this type of process (Hahne et al., 2002). Conversely, studies using content-only or function-only word-list control conditions, in which no syntactic process is required, might most truly reveal neuroanatomical correlates of Merge during the comparison between the syntactic and the non-syntactic condition. Interestingly, this subclass of studies rather reports a selective involvement of the inferior frontal regions, therefore suggesting syntax-sensitivity in the area (Bornkessel and Schlesewsky, 2006; Friederici, 2011; Hagoort, 2005). This may in turn reflect functional dissociation between content words and function words during sentence comprehension (Bastiaansen et al., 2005; Bradley and Garrett, 1983; Brown et al., 1999; Friederici, 1985; Mohr et al., 1994; Neville et al., 1992; Osterhout et al., 1997; Pulvermuller, 1995; Shapiro and Jensen, 1986; Small et al., 1998; Wang et al., 2008). A major role for BA44 in syntactic processing finds strong support in artificial language studies, which show the involvement of the area during the processing of hierarchical phrase-structure grammars (Bahlmann et al., 2008; Friederici et al., 2006a), as well as during the processing of natural language structures, in which the stimulus material is operationalized in terms of syntactic complexity manipulation, either by using embedding or scrambling manipulations (Friederici et al., 2006b; Makuuchi et al., 2009; Meyer et al., 2012; Santi and Grodzinsky, 2010). At the structural level, Diffusion Tensor Imaging (DTI) studies further indicate that BA44 is connected to the posterior temporal region (pSTS/STG) via long-range fiber bundles—anatomically identified with the arcuate fascicle/superior