CellPress

The biological basis of language: insight from developmental grammatical impairments

Heather K.J. van der Lely and Steven Pinker

Department of Psychology, Harvard University, William James Hall 970, 33 Kirkland Street, Cambridge, MA 02138, USA

Specific language impairment (SLI), a genetic developmental disorder, offers insights into the neurobiological and computational organization of language. A subtype, Grammatical-SLI (G-SLI), involves greater impairments in 'extended' grammatical representations, which are nonlocal, hierarchical, abstract, and composed, than in 'basic' ones, which are local, linear, semantic, and holistic. This distinction is seen in syntax, morphology, and phonology, and may be tied to abnormalities in the left hemisphere and basal ganglia, consistent with new models of the neurobiology of language which distinguish dorsal and ventral processing streams. Delineating neurolinguistic phenotypes promises a better understanding of the effects of genes on the brain circuitry underlying normal and impaired language abilities.

Developmental disorders as a window into the biology of language

Given the lack of animal models for language, and the inability to use invasive procedures with humans except out of medical necessity, our knowledge of the neurobiology of language has long depended upon natural experiments. During the 19th and 20th centuries, studies of patients with acquired brain lesions provided key insights [1–3]. Understanding of language in the 21st century promises to be enriched by data from developmental disorders. SLI, a family of language impairments in otherwise normal children, is highly heritable and has been linked so far to four genes. These discoveries provide a new route to understanding the complex pathways from genes and environment to the neural systems underpinning language.

This understanding depends, however, on breaking down the coarse categories of 'language' and 'language impairment' and examining the way that specific components of language are affected in specific disorders, and how they correlate with brain function and structure. That is, rather than searching for a direct link from genotype to behavior, we suggest linking genetic variants with alterations in the neural substrates of subcomponents of language processing.

Specific language impairment (SLI)

SLI is a heterogeneous family of impairments which affect the acquisition of language in 7% of children, an average of two in every classroom [4]. It frequently co-occurs with other disorders such as dyslexia, autistic spectrum disorders, and attention deficit and hyperactivity disorder [5,6], with which it also shares some phenotypic and genotypic characteristics[7,8]. Many genetic variants contribute to SLI across individuals, consistent with the heterogeneity of the disorder [7]. Despite this heterogeneity, the majority of children are impaired in grammatical functions, particularly syntax and morphology, and often phonology as well (Box 1) [4,9,10]. One of us has identified a subtype of the broad SLI category called G-SLI, which is concentrated in grammar, though it may embrace secondary deficits, for example, in the lexicon ([4,9–11]; but see also[12–14]). Though it was discovered in English-speaking children, G-SLI has since been identified in other languages [15–17].

Grammatical phenotypes of SLI

This review focuses on what G-SLI can reveal about the structure and neural instantiation of language. Crucially, G-SLI is not a global impairment of language or even of grammar, but is strongly manifested in certain aspects of linguistic performance while leaving others largely intact. This raises the possibility that the contrast reflects a key division within the neural or genetic substrates of language. In particular, children with G-SLI have difficulty interpreting and producing syntactic structures such as 'wh'-questions, the passive voice, and tense-marking; words that must be grammatically inflected in real time; and complex phonological structures embracing multiple syllables and clusters of phonemes. However, they have age-typical performance in syntactic tasks in which lexical semantic information is sufficient; in morphological tasks in which stored, nondecomposed forms are sufficient; and in phonological tasks in which strings of phonological units are sufficient. The problems with composed forms, moreover, persist into adulthood. We suggest that this pattern of deficits may reflect two modes of grammatical representation and processing which we call Basic and Extended (Box 1), and that individuals with G-SLI are specifically impaired in processing Extended representations.

Corresponding author: Pinker, S. (pinker@wjh.harvard.edu).

Keywords: specific language impairment; neurolinguistics; genetics of language; brain and language; language disorders.

1364-6613/

 http://dx.doi.org/10.1016/j.tics.2014.07.001

Trends in Cognitive Sciences November 2014, Vol. 18, No. 11

Box 1. Components of language

Language is traditionally divided into several subsystems [79]:

- **Syntax**: The combination of words into phrases and sentences, and assignment of grammatical relations (subject, object, head, etc.) which determine their compositional meaning.
- **Morphology**: The combination of words or parts of words (morphemes) into new words, further subdivided into 'inflection' (modifying a word according to its role in the sentence) and 'derivation' (creating a new word from old ones).
- **Phonology**: The combination of sounds into morphemes, and the modification of sounds according to their contexts.
- **Pragmatics**: Principles governing the use of language in a discourse and communicative context.
- **Lexicon**: The component of memory which stores words, idioms, and other fixed forms.

We concentrate on the first three, and propose that they are cross-classified by a distinction in representation and processing:

Extended grammatical representations are:

- abstract, consisting of categories defined by their grammatical privileges rather than their semantic content;
- hierarchical, defined by a tree of constituents embedded in larger constituents;
- nonlocal, potentially spanning long distances in the string;
- composed, namely assembled into meaningful combinations by rules.

Basic grammatical representations are:

- semantic and lexical, consisting of words or features of meaning;
- linear, defined by left-to-right ordering;
- local, involving adjacent or nearby elements;
- holistic, consisting of entire assemblies stored in memory.

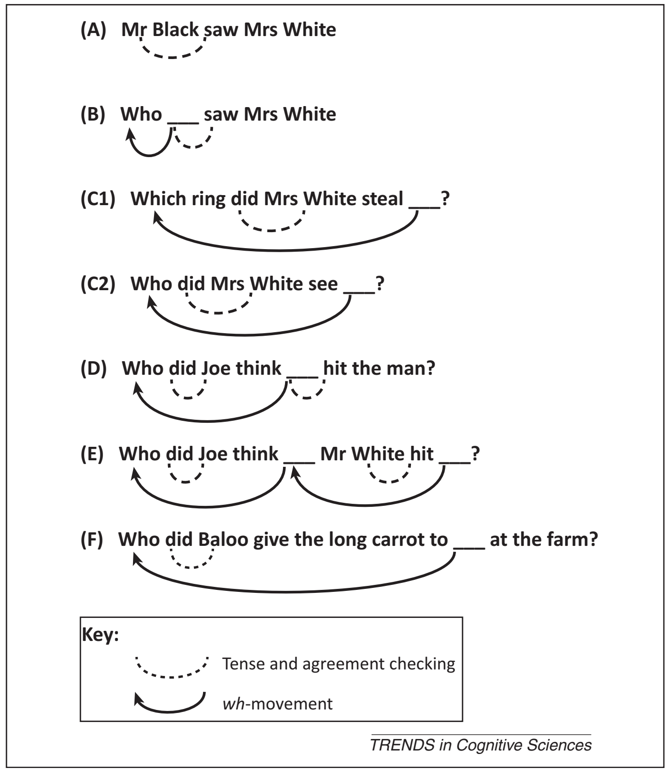

Figure 1. Examples of Extended syntax. The English rules for marking tense and agreement **(A–F)** and forming questions are complex and belong to what we call Extended syntax. Questions with 'wh'-words (who, what, which, etc.) such as those in **(B–F)** require movement (unbroken arrows) from an underlying position (underscore, cf. **(A)**) to a position in the surface string. The result of the movement is audible when the 'wh'-word corresponds to an object **(C)**, the subject of an embedded clause **(D,E)**, or the object of a preposition **(F)**, but it is covert when the 'wh'-word corresponds to the subject of a main clause **(B)**. Tense and subject–verb agreement are obligatory in English main clauses; in 'wh'-questions they must be marked on the auxiliary 'do' if the clause lacks any another auxiliary, and the auxiliary must be inverted with the subject (Did he leave? rather than Left he?). These operations are handled by a complex sequence of movement and feature-checking operations, which we abbreviate here with broken lines. If checking or movement has not been reliably computed, as (we hypothesize) is common in G-SLI, then for verbs that should be marked for tense, the infinitival form may be used instead, and the auxiliary may be omitted. For 'wh'-questions, the problem may be manifested as a filled gap or, if the movement is partial, the 'wh'-word may be copied in the medial moved position **(D, E)**. See Table 1 for examples of errors produced by such children.

Extended versus Basic syntax

Extended syntax involves hierarchical structures and dependencies between words, often spanning the entire clause, which are computed in real time. For example, in 'wh'-questions such as 'Who did Joe see __?', the 'wh'-word and the empty position after 'see' are in a dependent relation, which may be analyzed as the movement of the word from its original position in an underlying structure (Figure 1). Additionally, Extended syntax is abstract: the assembly and interpretation of phrases depend on their grammatical categories (noun, verb, tense) and relations (subject, head, complement), each defined by a pattern of intercorrelated privileges (where they may occur, how they may be inflected, what can substitute for what). Abstraction is central to grammar: the acquisition of abstract symbolic rules enables a person to generalize a pattern learned from a finite number of exemplars to an infinite number of new ones which need not resemble them in sound or meaning [18,19].

Basic syntax, in contrast, involves relations between words that can be determined from the meanings of the words themselves or from dependencies between adjacent words. Basic syntax may consist of holistic representations, in which sequences are stored and retrieved without necessarily analyzing their grammatical structure, and instead are linked directly to their semantic and pragmatic properties and their ordering relative to adjacent units.

Extended and Basic representations differ in their processing requirements. In Basic syntax, words and their features (number, gender, meaning) can be inserted directly from the lexicon, whereas in Extended syntax, relations between words within and across hierarchical units must be computed by operations such as movement and feature checking or unification. Figure 1 shows some of the extended syntactic relations which must be computed, according to a major theory of grammar [20], in assigning tense to a clause and in producing or interpreting 'wh'-questions; other theories require operations of comparable complexity.

Impairments in syntax

Table 1 shows the results of a variety of experiments, differing widely in their methods and processing demands, in which children with G-SLI display problems with Extended syntax (specifically, 'wh'-questions similar to those illustrated in Figure 1), but perform well in control conditions requiring only Basic syntax or lexical semantics [9,10,21].

Experiments testing other syntactic constructions show similar patterns. For example, teenagers with SLI rely on number information as a shortcut to understanding relative clauses in the same way that unimpaired six-year-olds do [22]: upon hearing 'The cat that is chasing the dogs is black', they understand that the cat is black rather than