fully fledged languages, ready to differentiate and to be influenced by neighbors.

Processing cost and its consequences

doi:10.1017/S0140525X15000916, e87

William O'Grady

Department of Linguistics, University of Hawaii at Manoa, Honolulu, HI 96822.
ogrady@hawaii.edu

Abstract: I focus on two challenges that processing-based theories of language must confront: the need to explain why language has the particular properties that it does, and the need to explain why processing pressures are manifested in the particular way that they are. I discuss these matters with reference to two illustrative phenomena: proximity effects in word order and a constraint on contraction.

Christiansen & Chater's (C&C's) proposal has much to recommend it: Processing resources are severely limited, and Chunk-and-Pass is a promising strategy for accommodating those limitations. The hope and promise of this type of work is that in addition to shedding light on the nature of incremental processing, it can help explain specific properties of linguistic systems. C&C focus their attention on very general features of language, such as duality of patterning, the bounded nature of linguistic units, and the existence of multiple levels of representation. But many properties at a finer level of granularity also call for attention. Why, for example, do we find certain systems of agreement and case marking, but not others? Why are some languages ergative? Why are filler-gap dependencies subject to certain types of locality constraints? Traditionally, the answers to such questions invoke principles of grammar, not processing. However, a wave of recent research by C&C and others (e.g., Hawkins 2004; 2014; O'Grady 2005; 2013; 2015a) proposes a very different approach: Languages are the way they are because of their need to adapt to processing pressures.

At least two challenges immediately arise. On the one hand, it is necessary to demonstrate that processing pressures can help resolve the baffling puzzles that spring up everywhere in the phonology, morphology, and syntax of natural languages. On the other hand, it is necessary to develop a theory to explain why the effects of the processing bottleneck are felt when and where they are. Two examples help illustrate this point.

As C&C note (sect. 6.1.2), items that enter into a relationship with each other should occur in close proximity, for obvious processing reasons. But how close? In Thai, not even a determiner can intervene between a verb and the head of its direct object (one says "I read book that"). But the picture is complicated by data from other languages.

(1) a. A determiner intervenes: (English, French, Mandarin)
read [**that** book]

b. A possessor NP intervenes: (English, Mandarin)
read [**a good friend's** book]

c. A relative clause intervenes (Mandarin):
read [**that I just bought**] books

(compare English: read books [**that I just bought**])

Hawkins (2004, p. 123ff) offers a key insight: All other things being equal, if a language permits a more costly implementation of a particular relationship, it will also permit a less costly implementation. For example, Mandarin allows a relative clause to appear between the verb and the head of its direct object, as in (1c) – a costly option in terms of working memory; as predicted, however, Mandarin also allows a less complex possessor phrase and a simple determiner to occur in that position. English sets the bar lower, allowing only possessor phrases and determiners to intervene – as in (1a,b) – but not a relative clause. The cut-off point for French is still lower: A determiner can intervene, as in (1a), but not a possessor or a relative clause. Most restrictive of all is Thai, in which even determiners cannot intervene. The processing bottleneck, it seems, is not absolute; it is manifested in different ways in different languages.

Another example of systematic variation in processing effects involves the notorious constraint on *want to* contraction illustrated below.

(2) a. Contraction allowed:
Ask whether they want to stay there. (cf. They want to stay there.)
wanna

b. Contraction prohibited:
Ask who they want to stay there. (cf., They want Mary to stay there.)
*wanna

Jaeggli (1980) proposed that contraction is blocked in (2b) by the presence of an invisible Case-marked trace between *want* and *to* – a classic example of grammatical analysis. In contrast, O'Grady (2005) outlined a processing-based alternative that turns on the interplay between two pressures: (a) for reasons related to working memory, filler-gap dependencies are best resolved at the first opportunity; (b) for articulatory reasons, contraction is most natural when *want* and *to* combine with each other without delay. Matters are straightforward in (2a), where the articulatory system moves seamlessly from *want* to *to*, producing a contracted pronunciation.

(3) Ask whether they **want-to** stay there.
↓
wanna

The situation is very different in (3) than in (2b), in which the transition from *want* to *to* is interrupted by the need to promptly resolve the filler-gap dependency by associating the *wh* word with *want*, which is transitive here (cf. *We want* ***her*** *to stay*). The resulting delay, often accompanied by prosodic reflexes such as lengthening of *want* (Warren et al. 2003), compromises the naturalness of contraction.

(4) Ask who they **want # to** stay there.

Here too, though, there is evidently room for variation. Ito (2005) reported that 5 of the 41 English speakers who she studied allowed *wanna* in patterns like (2b). Crucially, however, they also permitted contraction in the less-demanding (2a). The reverse is, of course, not true: Many speakers permit contraction in the easy pattern but not the difficult one.

In sum, case studies such as these help confirm that processing pressures (C&C's Now-or-Never bottleneck) shape the way language works, creating an explanatory narrative that is fundamentally different from traditional grammar-based accounts. At the same time, we gain insight into the nature of processing itself, for which an intriguing story is beginning to emerge. Because processing cost can never be reduced to zero, there is no perfect language and no single way to manage processing costs. What we find instead is systematic variation in what languages (and speakers) tolerate, with a preference for less-costly options over more-demanding alternatives. The end result is an array of effects in phenomena ranging from typological variation to developmental order (O'Grady 2013; 2015b).

Processing cost offers an important idea on which to build. The next step requires further close-range attention to the details of how languages work, how they differ from each other, and how they are acquired. Here, in the traditional data fields of linguistics, lie the clues needed to settle the disputes that define the contemporary study of language.

Downloaded from http://www.cambridge.org/core. University of Hawaii at Manoa Library, on 06 Dec 2016 at 21:31:51, subject to the Cambridge Core terms of use, available at http://www.cambridge.org/core/terms. http://dx.doi.org/10.1017/S0140525X15000916