Human Development 53 (2010), pp. 276-284
 DOI: 10.1159/000321290

Productivity, proceduralisation and SLI: Comment on Hsu and Bishop

Ewa Dąbrowska
Northumbria University

Short title: Productivity, proceduralisation and SLI

Address for correspondence:
Department of Humanities
School of Arts and Social Sciences
Northumbria University
Newcastle upon Tyne
NE1 8ST
United Kingdom
tel. 0191 2273497
e-mail: ewa.dabrowska@northumbria.ac.uk

Keywords: Specific Language Impairment, usage-based models, productivity, language acquisition, procedural learning, analogy, schema

Abstract

Hsu and Bishop argue that the underlying deficit in SLI is impaired statistical/procedural learning (in particular, difficulties in learning non-adjacent dependencies). This prevents affected children from developing abstract syntactic representations which in turn leads to diminished productivity. However, grammatical productivity need not be based on abstract syntactic representations. Moreover, abstraction is a matter of degree, and most everyday language use relies on relatively concrete, low-level schemas. I suggest that the grammatical problems in SLI may be due not to failure to develop abstract representations, but to a disruption of later consolidation processes which support the development of fast, automatic and error-free performance.

Productivity, proceduralisation and SLI: Comment on Hsu and Bishop

In the 1990s, many researchers tried to explain SLI by appealing to an innately specified "language module", which was assumed to be impaired in affected individuals. Over the years, language researchers have become increasingly dissatisfied with this explanation, and number of alternative accounts have been put forward which attempt to explain the language difficulties in SLA as a consequence of an underlying processing or learning deficit. Hsu and Bishop's contribution is part of this research tradition. Specifically, they argue that

> "… grammatical deficits arise when the learning system is biased towards the memorisation of exemplars, and is poor at extracting statistical dependencies from the input…. SLI involves deficits in extracting nonadjacent dependencies from input, leading to reliance on rote learning." (abstract)

This is an interesting proposal, and Hsu and Bishop cite an impressive amount of suggestive evidence. However, while the general idea is promising, the view of language and language learning that they assume is based on some questionable assumptions. I begin this commentary by discussing these implicit assumptions, and then show how their basic insight can be reformulated in a usage-based theory of language acquisition.

Grammatical productivity needn't be based on abstract representations

Hsu and Bishop draw a sharp distinction between concrete (exemplar-based) representations on the one hand and "system-wide abstract syntactic schemes" (5) on the other, and argue that the latter are needed "to support production and accurate comprehension of sentences that a child never heard before" (5). However, we do not need to postulate abstract syntactic structures in order to explain linguistic productivity: novel forms can also be constructed by using a stored exemplar as a model, and applying analogy (KISS+PAST is to *kissed* as HISS+PAST is to ???).

Exemplar models work particularly well for morphology (e.g. Eddington, 2000), but have also been applied successfully to a number of syntactic phenomena (see Bod, 2006; Skousen, 1989, 1992). Since analogies based on a single model are frequently incorrect, successful analogical models compute multiple analogies in parallel, and then choose the best solution using an algorithm that trades frequency off against similarity to the target. The problem with this approach is that it results in a computational explosion when the system needs to compute a number of analogies simultaneously, each involving a large number of models. For this reason, most language researchers assume that in the course of language acquisition analogies computed "on the fly" are gradually replaced by stored schemas. (Note, however, this property of analogical models offers a natural account of the problems experienced by children with SLI: computing multiple analogies at the same time will be particularly difficult when processing resources are limited, and the whole process may be abandoned.)

In usage-based approaches, schemas are regarded as abstract symbolic units which capture the relational similarities between the concrete exemplars from which they emerge. Since analogy also relies on relational similarity, applying analogy and extracting schemas are closely linked. Applying analogy involves three stages: retrieving relevant exemplars from memory (which requires assessing the similarity of

the target to potential models), establishing correspondences between matching parts, and computing the novel form. If a speaker repeatedly retrieves the same, or substantially overlapping, set of exemplars for the purposes of computing an analogy and compares them to establish correspondences between subparts, the links between the corresponding subparts will be strengthened, resulting in the emergence of a schema. Thus, schema extraction can be regarded as a result of applying analogy to produce and understand novel forms (Abbot-Smith & Tomasello, 2006; Dąbrowska, 2008a; Langacker, 2000, 2010).

Abstraction is not an all-or-nothing affair

Even if we accept that productivity involves relying on stored schemas or rules, it is important to realize that abstraction is not an all-or-nothing affair but a continuum: that is to say, speakers have not only fully concrete units (*Get a life!, I don't think so, What's a nice girl like you doing in a place like this?*) and fully abstract schemas (*SVO, AUX S VP?),* but also partially schematic constructions such as *take a look at NP, will you VP for me?, NP learn from NP-GEN mistakes*). Such partially schematic constructions are ubiquitous in language (Boas, 2010; Goldberg, 2006). Many are idiomatic in that they involve combinations of words or categories that violate some basic rules of the language, such as, for instance, the *The X-er the Y-er* construction exemplified by expressions such as *The more the merrier, The sooner the better, The bigger they come the harder they fall* (Fillmore, Kay, & O'Connor, 1988). Others, such as the *What's X doing Y?* construction, follow the general structural patterns of the language but have idiosyncratic meanings (Kay & Fillmore, 1999).

Much of the early work in construction grammar was devoted to documenting the existence of such idiosyncratic constructions. But there is a growing consensus that even form-meaning pairings that are fully predictable from general rules may nevertheless acquire unit status, i.e. become constructions in their own right (Goldberg 2006; Langacker, 2000). For instance, when adult native speakers of Polish form the dative case, they appear to rely on low-level rules applying to clusters of words sharing certain morphophonological properties (such as "add *-u* to nouns ending in *-ęcie*", "add *-u* to nouns ending in *-isko*", "add *-u* to nouns ending in *-nie*", etc.) rather than on more general rules ("add *-u* to neuter nouns"), in spite of the fact that the more general rule subsumes the more specific ones and thus describes the data more economically (Dąbrowska, 2004a, 2008a).

As a second example, consider English questions with long-distance dependencies (LDDs), i.e. questions which contain a dependency between a filler in the main clause and a gap in a subordinate clause, as in *Who$_1$ do you think [__$_1$ won the race]?,* where *who* "goes with" the subordinate clause verb, *won,* rather than *think.* Such dependencies are often called "unbounded" because, in principle, there can be any number of clauses intervening between the filler and the gap (cf. *Who$_1$ do you think [Steve believes [they said [Maria imagined [__$_1$ won the race]]]]?*). However, sentences involving dependencies spanning more than one clause boundary are extremely rare – perhaps non-existent – in spontaneously produced texts. In fact, attested examples of LDD questions are extremely stereotypical: the matrix auxiliary is nearly always *do,* the matrix subject *you,* and the verb *think* or *say*; and the matrix clause contains no complementizer and no additional modifiers. This has lead some usage-based linguists (Dąbrowska, 2004b, 2008b; Verhagen, 2006) to propose that speakers' knowledge of these constructions is best explained in terms of two lexically-specific templates: *WH do you think S-GAP?* and *WH did you say S-GAP?* According to such accounts, speakers produce "prototypical" LDD questions, i.e., those that match one of the templates, simply by inserting appropriate lexical material

into the WH and S-GAP slots. “Unprototypical” questions are produced by modifying the template, which requires extra effort. There is considerable evidence that prototypical LDD questions are indeed more basic, and easier to produce, than unprototypical ones. They are acquired earlier by children (Dąbrowska, Rowland, & Theakston, 2009), produced more fluently (Dąbrowska, 2010), remembered better (Dąbrowska et al., 2009), and judged to be more acceptable (Dąbrowska, 2008b).

The accounts proposed by Dąbrowska and Verhagen suggest that the complicated syntactic machinery postulated by formalist theories to account for such sentences (cyclic WH-movement, empty nodes that serve as landing sites, constraints on movement) may not be necessary. Of course the fact that they are not necessary does not entail that they do not exist: it is perfectly possible that speakers represent “the same” information redundantly at different levels of abstraction. However, it is important to note that in this case, the burden proof is with researchers who claim that they do exist.

Normal productive use is sanctioned by low-level schemas

Langacker suggests that while mental grammars may contain both abstract high-level schemas and relatively specific, low-level generalizations, it is the latter that are normally employed for assembling novel expressions; high level schemas “may be of only secondary significance, serving more of an organizing function than an active computational one” (1991: 265). Abstract high-level schemas may be needed to account for novel usages like *He sneezed the napkin off the table* or *Don’t giggle me* (although they could also be produced on analogy to more conventional expressions like *The wind blew the napkin off the table* and *Don’t tickle me*); but ordinary, garden-variety sentences like *He put the car in the garage* or *Do you like apples?* can be produced simply by inserting material into lexically specific templates such as *NP_1 put NP_2 DIR* and *Do you like NP?* respectively.

There is a considerable body of evidence suggesting that speakers prefer low-level schemas, and that performance is faster and more accurate when a low-level schema is available (see Dąbrowska, 2010, for a review). As observed earlier, prototypical questions with long distance dependencies (which can be produced simply by inserting novel phonological material into a lexically-specific template) are produced more fluently and remembered better than non-prototypical questions. A large number of psycholinguistic studies have demonstrated that verbs are processed faster when they occur in a construction with which they are frequently associated (Garnsey et al., 1997; Holmes, Stowe, & Cupples, 1989; Trueswell, Tanenhaus, & Kello, 1993). The simplest explanation for such findings is that speakers store lexically specific templates consisting of the verb and a schematic characterization of its frequent arguments, e.g. *NP claim S*, *NP accept S*, and that such templates are easier to access than more general representations (if the latter are available).

Similar effects have been observed in morphology. Dąbrowska (2004a, 2008a) found that adult native speakers of Polish were much better at inflecting nonce nouns from densely populated morphophonological neighbourhoods (for which they are likely to have low level schemas) than on nouns from sparsely populated neighbourhoods (where they have to apply the general rule); and in fact some speakers did not appear to have a general rule at all. The differences are particularly large for nouns belonging to the smallest class (neuters): in the Dąbrowska (2004a) study, speakers supplied the target inflection 98% of the time with neuter nouns with stems ending in *-anie* and *-enie* (an ending shared with a large number of real nouns), but only 47% of the time with neuter nouns belonging to more sparsely populated

morphophonological neighbourhoods – in spite of the fact that the same ending is required by all canonical neuter nouns.

Acquisition of grammar is not just distributional learning

Hsu and Bishop contrast the generative approach, according to which the acquisition of grammar involves setting the values of a small set of parameters, with the statistical learning approach, which claims that language learners extract statistical regularities from the input. It is worth noting that generative linguistics and the particular version of the statistical learning approach that Hsu and Bishop draw their inspiration from share an important characteristic: they both focus on the purely formal aspects of linguistic organisation, and seem to overlook the fact that language is used to communicate meaning. But knowing a language involves more than knowledge about the transitional probabilities between elements, whether adjacent or even non-adjacent; it involves the knowledge of constructions, i.e., form-meaning pairings.

Meaning is not only what language is for; it also plays a important role in shaping its structure, and it provides crucial clues about grammatical organization (including constituency and category membership) to the language learner. It also mediates analogical extension: the most useful analogies are not between pairs of forms, but between form-meaning pairings. Learners can apply analogy to create novel utterances by replacing a word or phrase in the model utterance with an expression that plays an analogous role in the situation described by the new utterance. Thus, the sentence *I painted the red barn* describes a situation in which a particular agent (the speaker) paints a particular object (a barn) of a particular colour (red). If the painter is Claire, we can describe the situation by saying *Claire painted the red barn;* if the object is a house, the sentence becomes *I painted the red house;* if the painter used blue paint instead of red, this will be described as *I painted the blue barn*, and so on. As Chomsky (in Searchinger, 1994) points out, purely formal analogies (i.e., analogies between pairs of forms) are often misleading. Thus, both of the sentences in (1) are grammatical; but substituting *see* for *paint* produces a grammatical result in (2a) and an ungrammatical one in (2b). Analogies based on both form and meaning, on the other hand, are much more reliable.

(1) a. I painted a red barn.
b. I saw a red barn.
(2) a. I painted a barn red.
(2) b. *I saw a barn red.

There is no doubt that statistical learning plays an important role in language acquisition. However, statistical learning involves more than just learning about how frequently a particular sound or word (or class of sounds or words) occurs next to some other formal element or class of elements. It also involves learning how frequently a particular linguistic form is used to express a particular function, and tracking the frequency of co-occurrence of *meaningful* elements. And, importantly, other abilities, such as the ability to infer meaning from context and to apply analogy, are also critical for the development of language (Tomasello, 2003).

So what precisely is the problem in SLI?

<u>Non-adjacent dependencies?</u>

Hsu and Bishop summarise a number of studies suggesting that children with SLI are less productive in their use of grammatical constructions, even when compared with younger children matched on MLU. They also point out that in several artificial grammar learning experiments, adults and adolescents with SLI were unable to learn nonadjacent dependencies, and suggest that the latter explains the former: learning non-adjacent dependencies helps learners discover abstract structure and hence become productive.

I am not convinced by this part of their argument. It seems plausible enough that SLI children have particular problems with non-adjacent dependencies, and that this difficulty may have further, and non-obvious, consequences for grammar learning (see also Leonard and Deevey in press). However, it does not necessarily follow that establishing non-adjacent dependencies is the underlying difficulty in SLI. Non-adjacent dependencies are more difficult than adjacent ones, even for normally developing children and adults. Thus, it is not surprising that they should be particularly problematic for children with SLI. More importantly, although such children are *less* productive with some grammatical markers – notably tense/agreement morphology – than normally developing children, they still demonstrate a fair degree of productivity with these morphemes. Five-year-old children with SLI use tense inflections about 40% of the time; by seven years of age, they supply them in approximately 80% of obligatory contexts (Rice, Wexler, & Hershberger, 1998); and they overregularize at rates similar to normally developing children of the same age (Marchman, Wulfeck, & Weismer, 1999). Thus the problem is not the lack of the requisite grammatical representations, but rather failure to *consistently* apply grammatical markers in obligatory contexts – in other words, it seems to have more to do with the proceduralisation of linguistic knowledge.

Proceduralisation?

Hsu and Bishop point out that grammatical SLI is also associated with poor procedural learning, and observe that "these findings point to a common basis for procedural and statistical learning, as both involve implicit learning of sequential patterns and both distinguish typically-developing children from children with poor grammar" (16). However, they caution against a strict dichotomisation of linguistic processes *à la* Pinker (e.g. 1998) and Ullman (e.g. 2004), where vocabulary and irregular morphology are subserved by declarative memory while regular morphology and syntax rely on the procedural system, pointing out that statistical learning can account for both regular and irregular morphology.

An additional reason to be sceptical about the Pinker/Ullman approach is that it does not generalise well to languages other than English. The dual mechanism theory offers a convincing account of English past tense and plural marking systems which show clear dissociations between a highly productive regular inflection and the irregular patterns, which are only a marginally productive. However, in English, regularity is confounded with several other properties. The regular and irregular inflections rely on different morphological mechanisms (regular past tense and plural forms are formed by suffixation, while most irregular inflections involve stem changes), and they differ in frequency (the two regular inflections have very high type frequencies, while most irregular forms have high token frequencies) and applicability (the regular inflections combine freely with a variety of different stems, while the irregulars apply to individual words or clusters of phonologically similar words). These differences exaggerate the contrast between regular and irregular process, making what is a quantitative difference appear qualitative (Dąbrowska, 2001). Languages with more complex inflectional systems such as Polish (cf. Dąbrowska,

2001, 2004a), Italian (Orsolini, Fanari, & Bowles, 1998), Icelandic and Norwegian (Ragnarsdottir, Simonsen, & Plunkett, 1999) often contain patterns with intermediate productivity, which share some properties of English regulars and some with irregulars; moreover, in these languages frequency and phonological similarity effects are often observed for regulars as well as irregulars.

Thus, linguistic systems (lexicon v. grammar) don't line up neatly with memory systems (declarative v. procedural). Both memory systems are probably involved in both types of linguistic knowledge (which explains the co-occurrence of lexical and grammatical deficits – cf. Bates & Goodman, 1997), though to different degrees and at different stages.

Learning a language involves four main processes: (1) acquiring a database of memorized form-meaning pairings, (2) segmenting the form into smaller chunks (phrases, words, morphemes) and matching these chunks with salient semantic substructures, (3) forming slots by generalizing over items which express similar meanings and occur in the same position in the construction, and (4) optimizing the retrieval and integration of units for fluent processing. The first three of these processes rely strongly on attention, controlled processing, and the declarative memory system (although statistical/procedural learning is also involved, of course, particularly in the acquisition of the phonological system – cf. Jusczyk, 1997; Velleman & Vihman, 2006). Learning during this stage is relatively fast and results in obvious changes in behaviour: new words and constructions appear in the child's speech, and usage becomes increasingly flexible and creative. The optimization stage relies almost entirely on procedural memory. Learning is slow, and does not involve the acquisition of new knowledge, but rather the consolidation and restructuring of existing knowledge. Consequently, there are no dramatic behavioural changes: performance gradually becomes faster and more accurate.

An explanation in terms of impaired procedural learning provides a plausible explanation of the linguistic problems experienced by older children with SLI: they are able to use most constructions productively, but their performance is slow, effortful and inconsistent; and they experience particular difficulties with the more grammaticised aspects of linguistic structure – function words and inflectional endings. Of course, language problems in SLI become evident much earlier: affected children are late talkers, have difficulties learning words, short MLUs, and unstable phonological representations. These difficulties could be due to problems with constructing phonological representations, since procedural learning is critically involved in the development of phonology (Velleman & Vihman, 2006).

Concluding remarks

As pointed out earlier, proceduralisation is a slow process, even in normal language development. English speaking children begin to use past tense inflections at about 2;4, and begin to use them productively (as evidenced by overgeneralization errors) a few months later. They supply them fairly consistently, in about 90% of obligatory contexts, by age 4;0; by 5;0, regular marking rates in spontaneous speech are close to 100%. Thus, even for a relatively simple and frequent construction, the time lag between emergence and full mastery can be quite substantial.

We don't know very much about these later processes of consolidation and restructuring that lead to fast and accurate performance. Most child language research concentrates on the early stages of development. Once children demonstrate above-chance performance on comprehension tests, or productive use of a particular construction, researchers conclude that they have acquired the relevant knowledge. Inconsistent performance is regarded as just that, *performance* – noise in the data that

adds little to our understanding of language acquisition. But knowledge about these processes could provide important clues to understanding SLI, where it is clear that it more than just noise. Conversely, a better understanding of exactly what goes awry in SLI may help us understand normal development.

References

Abbot-Smith, K., & Tomasello, M. (2006). Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *Linguistic Review, 23*(3), 275-290.

Bates, E., & Goodman, J. (1997). On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia, and real-time processing. *Language and Cognitive Processes, 12*, 507-584.

Boas, H. C. (2010). Verb meanings at the crossroads between higher-level and lower-level constructions. *Lingua, 120*, 22-34.

Bod, R. (2006). Exemplar-based syntax: How to get productivity from examples. *The Linguistic Review, 23*.

Dąbrowska, E. (2001). Learning a morphological system without a default: The Polish genitive. *Journal of Child Language, 28*, 545-574.

Dąbrowska, E. (2004a). Rules or schemas? Evidence from Polish. *Language and Cognitive Processes, 19*, 225-271.

Dąbrowska, E. (2004b). *Language, Mind and Brain. Some Psychological and Neurological Constraints on Theories of Grammar*. Edinburgh: Edinburgh University Press.

Dąbrowska, E. (2008). The effects of frequency and neighbourhood density on adult speakers' productivity with Polish case inflections: An empirical test of usage-based approaches to morphology. *Journal of Memory and Language, 58*, 931-951.

Dąbrowska, E. (2008). Questions with 'unbounded' dependencies: A usage-based perspective. *Cognitive Linguistics, 19*, 391-425.

Dąbrowska, E. (2010). The mean lean grammar machine meets the human mind: Empirical investigations of the mental status of rules. In H.-J. Schmid & S. Handl (Eds.), *Cognitive foundations of linguistic usage patterns. Empirical approaches* (pp. 151-170). Berlin: Mouton de Gruyter.

Dąbrowska, E., Rowland, C., & Theakston, A. (2009). The acquisition of questions with long-distance dependencies. *Cognitive Linguistics, 20*, 571-596.

Eddington, D. (2000). Analogy and the dual-route model of morphology. *Lingua, 110*, 281-293.

Fillmore, C. J., Kay, P., & O'Connor, M. C. (1988). Regularity and idiomaticity in grammatical constructions: the case of let alone. *Language, 64*, 501-538.

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contribution of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language, 37*, 58-93.

Goldberg, A. E. (2006). *Constructions at Work. The Nature of Generalization in Language*. Oxford: Oxford University Press.

Holmes, V. M., Stowe, L. A., & Cupples, L. (1989). Lexical expectations in parsing complement-verb sentences. *Journal of Memory and Language, 28*, 668-689.

Jusczyk, P. (1997). *The Discovery of Spoken Language*. Cambridge, MA: MIT Press.

Kay, P., & Fillmore, C. J. (1999). Grammatical constructions and linguistic generalizations: the *What's* X *doing* Y construction. *Language, 75*, 1-33.

Langacker, R. W. (1991). *Concept, Image and Symbol: The Cognitive Basis of Grammar*. Berlin: Mouton de Gruyter.

Langacker, R. W. (2000). A dynamic usage-based model. In M. Barlow & S. Kemmer (Eds.), *Usage-Based Models of Language* (pp. 1-63). Stanford, CA: CSLI Publications.

Langacker, R. W. (2010). How not to disagree: The emergence of structure from usage. In K. Boye & E. Engberg-Pedersen (Eds.), *Language Usage and Language Structure* (pp. 107–144). Berlin & New York: De Gruyter Mouton.

Leonard, L. B., & Deevy, P. (in press). Input distribution influences degree of auxiliary use by children with Specific Language Impairment. *Cognitive Linguistics*.

Marchman, V., Wulfeck, B., & Weismer, S. E. (1999). Morphological productivity in children with normal language and SLI: A study of English past tense. *Journal of Speech, Language and Hearing Research, 42*, 206–219.

Orsolini, M., Fanari, R., & Bowles, H. (1998). Acquiring regular and irregular inflection in a language with verb classes. *Language and Cognitive Processes, 13*, 425-464.

Pinker, S. (1998). Words and rules. *Lingua, 106*, 219-242.

Ragnarsdottir, H., Simonsen, H. G., & Plunkett, K. (1999). The acquisition of past tense morphology in Icelandic and Norwegian children: an experimental study. *Journal of Child Language, 26*, 577-618.

Rice, M., Wexler, K., & Hershberger, S. (1998). Tense over time: The longitudinal course of tense acquisition with specific language impairments. *Journal of Speech, Language, and Hearing Research, 41*, 1412-1431.

Searchinger, G. (1994). The Human Language. Part 2: Acquiring the Human Language: "Playing the Language Game". Hohokus, N.J.: Equinox Films.

Skousen, R. (1989). *Analogical Modeling of Language*. Dordrecht: Kluwer.

Skousen, R. (1992). *Analogy and Structure*. Dordrecht: Kluwer.

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence-processing -- separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning Memory and Cognition, 19*, 528-553.

Ullman, M. T. (2004). Contributions of memory circuits to language: the declarative/procedural model. *Cognition, 92*, 231–270.

Velleman, S. L., & Vihman, M. M. (2006). Phonological development in infancy and early childhood: Implications for theories of language learning. In M. C. Pennington (Ed.), *Phonology in Context* (pp. 25-50). Luton: Macmillan.

Verhagen, A. (2006). On subjectivity and 'long distance *Wh*-movement'. In A. Athanasiadou, C. Canakis & B. Cornillie (Eds.), *Various Paths to Subjectivity* (pp. 323-346). Berlin/New York: Mouton de Gruyter.