Commentary

Ben Ambridge

Island constraints and overgeneralization in language acquisition

Abstract: Ambridge and Goldberg (2008) found that long distance dependency (LDD) questions (e.g., *Who did she mumble that she saw?*) do not seem to be formed by analogy with similar, more frequent sentences of the same type (e.g., *What do you think X?; What did he say X?*), but, rather, that such questions are acceptable to the extent that the main verb backgrounds the complement clause (e.g., *say > mumble*). Kalyan (2012) argued that this finding is compatible with a similarity-based account, provided that similarity between the verb and *say/think* is defined as similarity in the extent to which the verb backgrounds the complement clause. In the present article, I argue that Kalyan (2012) is correct, and that this phenomenon can be seen as an instance of a broader phenomenon whereby the fit between the properties of a particular item (e.g., a verb) and those of a particular construction slot (e.g., the VERB slot in the LDD question construction) is the primary determinant of the degree of (un)grammaticality of a possible generalization.

Keywords: subjacency, island constraints, long distance dependency questions, FIT account

My goal in this note is not to refute but to endorse Kalyan's (2012) reinterpretation of Ambridge and Goldberg's (2008) data on island constraints. In doing so, I argue that Kalyan's analysis provides support for a more general theory of (over)generalization in language acquisition outlined in Ambridge and Lieven (2011: 256–268) and a recent paper in this journal (Ambridge et al. 2011: 316–321).

To recap, Ambridge and Goldberg (2008) demonstrated that utterances such as **Who did she mumble that he saw?* are rated as less acceptable than equivalent utterances with semantically "lighter" verbs (e.g., *Who did she say that he*

Ben Ambridge, Institute of Psychology, Health & Society, University of Liverpool; ESRC International Centre for Language and Communicative Development (LuCiD),
E-mail: Ben.Ambridge@Liverpool.ac.uk

saw?) and further that, across 12 verbs, the greater the extent to which the main verb is judged to "background" the complement clause (by virtue of being semantically "heavy"), the greater the unacceptability of the relevant long distance dependency (LDD) question. Crucially, Ambridge and Goldberg (2008) found no correlation between the semantic similarity of each verb to *think/say*, as assessed by human raters and *Latent Semantic Analysis [LSA]*, and the acceptability of the LDD question. This null finding, the authors argued, constitutes evidence against the possibility that LDD questions are formed by analogy to the high frequency exemplars *WH do you think S-GAP* and *WH did you say S-GAP?*, as exemplar- or similarity-based accounts (e.g., Bybee and Eddington 2006; Bybee 2010; Dąbrowska 2008) would seem to predict.

Kalyan argues that if "similarity" between each verb and *think/say* is defined as similarity in *the extent to which the verb backgrounds/foregrounds the complement clause*, as opposed to the more general type of similarity measured by human similarity judgments and LSA, then Ambridge and Goldberg's (2008) result is predicted by a similarity-based account. Indeed, Kalyan (2012: 549) notes that his explanation is "very similar to that of Ambridge and Goldberg, in that the acceptability of an LDD question is a matter of the degree of fit between the verb's specifications and those of the construction".

I agree fully with this assessment. My goal in this commentary is to argue that the fit between the properties of a particular item (e.g., a verb) and those of a particular construction slot is the primary determinant of the degree of (un)grammaticality of a given generalization. The claim is that most – possibly all – overgeneralization errors reflect the use of a lexical item in a construction slot with which it is less than optimally compatible in terms of its semantic, phonological, pragmatic, or other properties (or a combination of several). I begin by outlining some cases where the fit between the item and construction slot is defined in terms of semantic and phonological properties, before returning to LDD questions, where the relevant properties are mainly pragmatic.

The FIT account (so called because it emphasises the importance of the *Fit* between *Item* and *Template* properties) arose from work on verb argument structure overgeneralization errors (e.g., Ambridge et al. 2008). Of course, the idea that filler items must be compatible with their slots is not new, but rather a mainstay of cognitive linguistics (e.g., Langacker 2000). The novel contribution of the FIT account is its claim that **all** argument structure overgeneralization errors – and, as we will see, many other different types of error – reflect a slot-filler mismatch. Taking the example of the English causative, the claim is that generalizations such as **The clown giggled the man* (which are produced by many young children; see Bowerman 1988) arise when the speaker uses an item (here *giggle*) in a construction slot ([AGENT] **[ACTION]** [PATIENT]) with which it

is less than optimally compatible in terms of its semantic properties.[1] For this example, the clash arises because *giggle* has the semantic properties of denoting an internally-caused, single-participant action, whilst the ACTION slot in the transitive causative construction denotes an event of direct, prototypically physical, external causation (which is why one cannot say **John crashed the car* if John simply distracted the driver).[2] It is important to emphasise that, under this account, fit is graded and probabilistic, not all-or-nothing. Thus grammatical acceptability is also probabilistic; an utterance ranges from completely acceptable (perfect fit between item and construction properties) to completely unacceptable (no fit between item and construction properties), with all possibilities in between.

Children acquire the properties of particular verbs (and other items) through repeated exposure. Similarity, the properties of construction slots are acquired through repeated exposure to utterances that instantiate the relevant construction. If all the items that appear in a particular slot share a particular property (whether this is semantic, phonological, pragmatic etc.), the slot will take on this property. If the items that appear in a particular position are highly variable along a particular dimension (e.g., the verbs that appear in the ACTION slot of the transitive causative construction do not share any particular phonological properties), then the slot will not take on any particular properties with regard to this dimension (Suttle and Goldberg 2011).[3] This is essentially the argument

---

1 The account also posits other factors, such as the frequency with which the verb has appeared in this and competing constructions (for a full list, see Ambridge and Lieven 2011: 260). However, only the notion of fit is relevant for our present purposes.

2 Goldberg (1995: 169) discusses some apparent counterexamples. For example, one can say *The owner ran his favourite horse (in the race)* or *Chris cut her hair at the salon* (when, in fact, a hairdresser did the cutting). Goldberg analyses these as "conventionalized scenarios [that can be] cognitively packaged in such a way that their internal structure is ignored". However, one could argue that one can *run a horse* (or *walk a dog*) but not *run* or *walk a child* precisely because children are less amenable to direct causation than horses or dogs, and it is for this reason that only the former examples have become conventionalized. Similarly, for the present author, *Chris cut her hair* is certainly less acceptable as a description of a visit to the salon than *Chris had her hair cut*, presumably for the same reason. Of course, it will nearly always be possible to come up with an *ad hoc* justification for the grammaticality of a particular sentence in terms of semantic properties of the verb or event. What are needed are studies in which participants rate particular actions or scenarios on these dimensions, with these ratings then correlated with grammatical acceptability scores. A number of studies that follow this logic are outlined below.

3 The present account does not posit type frequency of each slot in the construction (the number of different items that have appeared in the slot) as a separate factor. Rather, apparent type-frequency effects are held to be emergent from variability effects: Slots with high type frequency are more productive only to the extent that they are likely to be filled by a wide range of highly variable items. Connectionist models of morphological generalization (e.g., Hare,

proposed by Kalyan, where the "category" in question is the slot (i.e., potential slot-fillers): "(...) features exhibited by all members [of the category] assume high salience, and hence have a large effect on the similarity between each exemplar of the category and the item to be categorized" (Kalyan 2012:548).

The flip-side, of course, is that features shared by only a few members are not relevant for categorization (i.e., slot formation). For example, a handful of verbs that may appear in the ACTION slot of the transitive causative construction begin with the phoneme /b/, but because most do not, the slot does not take on this phonological property.

1 Semantic fit

So far, our studies have provided evidence that – for three different construction types (locatives, datives and verbal *un-* prefixation) – the relative ungrammaticality of overgeneralization errors is determined by the fit between semantic properties of the verb and the relevant construction slot (amongst other factors).

For the ground- or container-locative construction [AGENT] [ACTION] [CONTAINER/LOCATION] *with* [CONTENTS] (e.g., *Lisa filled the box with paper*), the [ACTION] slot is associated with the semantic property of state-change (e.g., from empty to full; Gropen et al. 1991). For the figure- or contents-locative construction [AGENT] [ACTION] [CONTENTS] *into/onto* [CONTAINER/ LOCATION], the [ACTION] slot is associated with the semantic property of motion in a particular manner. Hence one can say *Lisa filled the box with paper* but not **Lisa filled paper into the box* and *Lisa poured juice onto the floor* but not **Lisa poured the floor with juice*. In an experimental study (Ambridge et al. 2012), adult participants rated 142 locative verbs for the extent to which each denotes (a) state-change and (b) manner of motion. In another part of the study, different participants (both adults and children) rated figure- and ground-locative uses of each of the verbs for grammatical acceptability. It was found that these semantic ratings could predict not only which verbs could appear in each of the two constructions, but also the relative ungrammaticality

---

Elman, and Daughtery 1995; Forrester and Plunkett 1994; Plunkett and Nakisa 1997) have shown that slots with low type frequency can be productive, provided that the items that have previously appeared in the slot are dissimilar on the relevant dimension (here phonology). An example here would be the German *[NOUN] + s* plural construction. Conversely, slots with high type frequency can be highly restrictive, if all the items that have appeared in this slot share a salient (here, phonological) feature.

of overgeneralization errors. This result held even when verb frequency – both overall and in the two locative constructions – was taken into account.

Very similar findings were obtained for the dative constructions (Ambridge et al. 2014). Adult ratings of the extent to which different verbs denote properties such as *possession-change* and *manner-of-transfer* were able to predict the relative acceptability of these verbs in the prepositional-object and double-object dative constructions (e.g., *John sent a package to Sue/John sent Sue a package*), as rated by adults and children.

Ambridge (2013) and Blything et al. (2014) reported similar findings for the morphological *un*-prefixation construction (*un-VERB*). Although the semantic properties of verbs that may and may not appear in this slot are not easily characterized (they include *enclosing, covering, surface attachment* and *circular motion*; Whorf 1956; Li and MacWhinney 1996) participants' ratings of these properties could again predict the relative unacceptability – and production probability – of particular errors (e.g., **unbend, *unclose, *unfreeze*).

It is not only [ACTION] (or [VERB]) slots that are associated with particular semantic properties. For example, the [AGENT] slot of the transitive causative ([AGENT] [ACTION] [PATIENT]) construction must be filled by a relatively direct causer (e.g., as we noted earlier *John crashed the car* is acceptable only if John is the driver).[4] Accordingly, another study (Ambridge et al. 2009) found that both adult and child participants rated overgeneralization errors with a more-direct causer (e.g., ***The magician's spell** *disappeared Bart*) as less unacceptable than comparable errors with a less-direct causer (e.g., ***The magician** *disappeared Bart*; causation is construed to be less direct here, because the intervening cause – the spell – is not mentioned).

2 Phonological fit

For the examples discussed so far, the compatibility between the slot and its filler (or lack thereof) has been semantic. We now consider errors where an item is overgeneralized into a slot with which it is less than optimally compatible in terms of its phonology. It is well known (e.g., Pinker 1989) that the double-object

---

4 Although, for simplicity, the theory is presented as though slots were filled in isolation, this example illustrates the point that, in fact, slots must "work together" in choosing suitable fillers. For example, for the scenario discussed here, *John* would be a suitable filler for the [AGENT] slot if the [ACTION] and [PATIENT] slots were filled by *distracted* and *the driver*, but not if they are filled by *crashed* and *the car*. I thank Adele Goldberg for helpful discussion on this point.

dative construction resists Latinate-sounding bi-syllabic verbs with stress on the second syllable (e.g., **John suggested Mary the trip*). This can be reformulated in positive terms: The VERB slot of the double-object dative construction is prob-abilistically associated with the phonological properties of monosyllabicity and first-syllable stress (and perhaps also of Germanic-sounding morphemes). Verbs whose phonological properties are not sufficiently compatible with those of the slot may not be generalized into the construction. Ambridge et al. (2012) found that, when taught novel pseudo Germanic (monosyllabic) and Latinate (iambic) verbs, adults rated the former as significantly more acceptable in the double-object dative construction than the latter.

Similar findings have arisen in the English past-tense debate. Studies with real English verbs are inconclusive due to the possibility of rote-learning. However, for both adults (e.g., Albright and Hayes 2003) and children (e.g., Ambridge 2010), studies have shown that the extent to which a novel stem (e.g., *spling*) is judged to be grammatical in various possible morphological past-tense constructions (e.g., *splinged*, *splang*, *splung*) is a function of the extent to which the stem is phonolo-gically compatible with the relevant word-level morphological construction. Importantly, this is true for the various "regular" past-tense constructions (STEM + [əd], e.g., *needed*; STEM + [t], e.g., *wished, kissed*) as well as the irregular constructions. For example, to be judged as acceptable in the STEM + [t] past-tense construction, a novel stem must share with the slot (i.e., all the other items that have appeared in the slot) the property of ending in a voiceless consonant.

3 Pragmatic fit

Having considered cases of *semantic* and *phonological* (in)compatibility between item and construction slot, we now come full circle to look at LDD questions as a case of *pragmatic* incompatibility. Let us assume that, as for other constructions, speakers acquire an LDD question construction that consists of a series of slots ([WH-WORD] [AUX] [SUBJECT] [VERB] *that* [S-GAP]). As for the other constructions discussed, each slot will be associated with particular properties. Given that virtually all the items that have previously appeared in the [VERB] position (primarily *think* and *say*) have the pragmatic property of foregrounding the S-GAP, only other items that share this property (e.g., *believe*) may appear in this position. Items that are increasingly incompatible with this pragmatic prop-erty – i.e., that increasingly background the S-GAP – (e.g., *mumble* < *realise*) are deemed increasingly ungrammatical in this slot. Of course, this is precisely the argument that is made by Kalyan (2012: 550), who states that "a complement-backgrounding verb would be a poorly-sanctioned filler of the main verb slot of

the LDD question construction". The point is simply that the ungrammaticality of such utterances is not an isolated phenomenon, but is one example of a wider phenomenon whereby the use of any item in a construction slot with which it is less than fully compatible in terms of its semantic, phonological, pragmatic/ discourse (or other) properties is less than fully grammatical.[5]

Indeed, there is at least one other type of "overgeneralization error" (though it is not normally described as such) that, under the present analysis, results from suboptimal fit between a slot and its filler in terms of pragmatic/discourse properties. These are so-called Principle-C violations such as *She$_i$ said that **Sarah$_i$** likes poetry*. The [SUBJECT] slot of the co-reference construction *[PRONOUN$_i$] said that **[SUBJECT$_i$]***[VP] is associated with the pragmatic property of referring to a discourse-old referent. Thus the use of a full nominal (*Sarah*) – which has the pragmatic property of referring to a discourse-new referent – in this slot leads to a mismatch between the discourse-pragmatic properties of the slot and filler, and hence to an ungrammatical utterance. This pragmatic analysis is not new (e.g., Van Hoek 1997); the point is that it is possible to frame this analysis in terms of a clash between properties of slots and their fillers.

4 Other types of fit

Finally, although the fit between item and construction properties has been characterized as (mainly) semantic, phonological or pragmatic in nature, there are doubtless other possibilities too. Wherever fillers share a particular property, the slot will probabilistically exhibit that property, and items that do not will make less than optimal fillers. This holds even if the precise nature of the relevant property as semantic, phonological etc. is difficult to characterize. This is the case, for example, for the [ADJECTIVE] slot of the [ADJECTIVE] [NOUN PHRASE] construction. Boyd and Goldberg (2011) showed that this slot

---

5 Note that the argument does not depend crucially on the assumption that the speaker acquires an abstract VERB slot with (predominantly) the semantic properties of *say* and *think*, as opposed to forming novel LDD questions by direct analogy with stored lexically-specific templates containing these items (e.g., *What do you think S-GAP?*). That is, we need take no position with regard to whether abstract constructions are stored or generated online as they are needed, or between prototype and exemplar-based models (see Ambridge and Lieven 2011: 133 for discussion of this point). Neither does the argument depend crucially on the characterization of the relevant properties of the VERB slot as "pragmatic" or "discourse functional". If "foregrounding the relevant clause" is viewed as a property that is semantic rather than pragmatic, the analysis still holds, provided that the semantic (or whatever) property can still be learned by abstracting across items that appear in this position in LDD questions in the input.

must be filled by an adjective that has the property of being non-decomposable into the prefix *a-* plus a stem that is semantically related to the meaning of the adjective. For example, one can say *The living man* but not **The alive man*, because *alive* **is** decomposable into the prefix *a-* and the stem *live*. Note however, that the prohibition is not purely phonological, as one can say – for example – *The awkward man*, because *awkward* cannot be decomposed into *a + wkward*. Thus while it is not clear precisely how to characterize the properties of the [ADJECTIVE] slot of the [ADJECTIVE] [NOUN PHRASE] construction (partly phonological, partly semantic, partly etymological?), there is clear evidence that the slot nevertheless exhibits them. For example, Boyd and Goldberg (2011) found that participants would use novel *a*-adjectives (e.g., *ablim*) in the [ADJECTIVE] [NOUN] construction, only when they exhibited these properties (i.e., **not** being segmentable into *a-* plus a meaningful stem).

5 Conclusion

Although different construction slots may have very different properties, the claim of the FIT account is that once the relevant properties have been identified, most – perhaps all – overgeneralization errors reflect the use of an item in a construction slot with which it is less than optimally compatible in terms of these properties. Errors in which verbs that background the complement clause are used in a verb slot that has the pragmatic property of foregrounding the complement clause (e.g., **Who did she mumble that he saw?*) are just one example of this phenomenon.

References

Albright, Adam & Bruce Hayes. 2003. Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition* 90. 119–161.

Ambridge, Ben. 2010. Children's judgments of regular and irregular novel past tense forms: New data on the dual- versus single-route debate. *Developmental Psychology* 46(6). 1497–1504.

Ambridge, Ben. 2013. How do children restrict their linguistic generalizations?: An (un-)grammaticality judgment study. *Cognitive Science* 37(3). 508–543.

Ambridge, Ben & Adele E. Goldberg. 2008. The island status of clausal complements: Evidence in favor of an information structure explanation. *Cognitive Linguistics* 19(3). 357–389.

Ambridge, Ben & Elena V. M. Lieven. 2011. Child language acquisition: Contrasting theoretical approaches. Cambridge: Cambridge University Press.

Ambridge, Ben, Julian M. Pine & Caroline F. Rowland. 2011. Children use verb semantics to retreat from overgeneralization errors: A novel verb grammaticality judgment study. *Cognitive Linguistics* 22(2). 303–323.

Ambridge, Ben, Julian M. Pine & Caroline F. Rowland. 2012. Semantics versus statistics in the retreat from locative overgeneralization errors. *Cognition* 123(2). 260–279.

Ambridge, Ben, Julian M. Pine, Caroline F. Rowland & Franklin Chang. 2012. The roles of verb semantics, entrenchment and morphophonology in the retreat from dative argument structure overgeneralization errors. *Language* 88(1). 45–81

Ambridge, Ben, Julian M. Pine, Caroline F. Rowland, Daniel Freudenthal & Franklin Chang. 2014. Avoiding dative overgeneralization errors: Semantics, statistics or both? *Language, Cognition and Neuroscience* 29(2). 218–243.

Ambridge, Ben, Julian M. Pine, Caroline F. Rowland, Rebecca L. Jones & Victoria Clark. 2009. A semantics-based approach to the "No Negative Evidence" problem. *Cognitive Science* 33(7). 1301–1316.

Ambridge, Ben, Julian M. Pine, Caroline F. Rowland & Chris R. Young. 2008. The effect of verb semantic class and verb frequency (entrenchment) on children's and adults' graded judgements of argument-structure overgeneralization errors. *Cognition* 106(1). 87–129.

Blything, Ryan P., Ben Ambridge & Elena V. M. Lieven. 2014. Children use statistics and semantics in the retreat from overgeneralization. *PLoS One*. 9(10). doi:10.1371.

Bowerman, Melissa. 1988. The "No Negative Evidence" problem: How do children avoid constructing an overly general grammar? In J. A. Hawkins (ed.), *Explaining language universals*, 73–101. Oxford: Blackwell.

Boyd, Jeremy K. & Adele E. Goldberg. 2011. Learning what not to say: Categorization and statistical preemption in 'a-adjective' production. *Language* 87(1). 1–29.

Bybee, Joan L. & David Eddington. 2006. A usage-based approach to Spanish verbs of 'becoming'. *Language* 82(2). 323–355.

Bybee, Joan L. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.

Dąbrowska, Ewa. 2008. Questions with long-distance dependencies: A usage-based perspective. *Cognitive Linguistics* 19(3). 391–425.

Forrester, Neil & Kim Plunkett. 1994. Learning the Arabic plural: The case of minority default mappings in connectionist networks. *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

Goldberg, Adele E. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Gropen, Jess, Steven Pinker, Michelle Hollander & Richard Goldberg. 1991. Affectedness and direct objects – The role of lexical semantics in the acquisition of verb argument structure. *Cognition* 41(1–3). 153–195.

Hare, Mary, Jeff L. Elman & Kim G. Daughtery. 1995. Default generalisation in connectionist networks. *Language and Cognitive Processes* 10(6). 601–630.

Kalyan, Siva. 2012. Similarity in linguistic categorization: The importance of necessary properties. *Cognitive Linguistics* 23(3). 539–554.

Langacker, Ronald W. 2000. A dynamic usage-based model. In M. Barlow & S. Kemmer (eds.), *Usage-based models of language*, 1–63. Stanford: CSLI.

Li, Ping & Brian MacWhinney. 1996. Cryptotype, overgeneralization, and competition: A connectionist model of the learning of English reversive prefixes. *Connection Science* 8(1). 3–30.

Pinker, Steven. 1989. *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: MIT Press.

Plunkett, K. & R. C. Nakisa. 1997. A connectionist model of the Arabic plural system. *Language and Cognitive Processes* 12(5–6). 807–836.

Suttle, Laura & Adele E. Goldberg. 2011. Partial productivity of argument structure constructions. *Linguistics* 49(6). 1237–1269.

Van Hoek, Karen. 1997. *Anaphora and conceptual structure*. Chicago: University of Chicago Press.

Whorf, Benjamin L. 1956. A linguistic consideration of thinking in primitive communities. In J. B. Carroll (ed.), *Language, thought, and reality*, 65–86. Cambridge, MA: MIT Press.