December 2015, *Linguistic Approaches to Bilingualism*.
Target article + 17 commentaries + this response

Learning Obscure and Obvious Properties of Language

Colin Phillips & Lara Ehrenhofer, University of Maryland

[response to commentaries on Phillips & Ehrenhofer, 2015]

Some aspects of an account of successful native language learning are non-negotiable. First, learners are unlikely to succeed if they cannot analyze the input correctly. Second, learners need to recognize discrepancies between their current grammar and the target language to progress to the appropriate endstate. Both requirements depend on an understanding of how learners process the language around them. In order to understand why other learner groups are less successful, it is useful to scrutinize the most successful learners, i.e., children: when and how do they succeed, and with what cognitive resources? We can then ask which of those mechanisms fall short in the less successful learners.

These two connections are the main motivations for our target article (Phillips & Ehrenhofer, 2015). We are grateful to the authors of the 17 commentaries for engaging seriously with our arguments. In attempting to draw connections across fields we inevitably exposed gaps in our own knowledge, and we are glad to have these filled in by the commentators. In our response we aim to clarify some key steps in the argument that are engaged by a number of the commentaries. In particular, we ask:

1. Why is an account of language processing in learners necessary? And why is the notion of a “processing account” so unhelpful in theories of grammar and learning?
2. Why are predictive mechanisms so valuable for successful learning?
3. How is language processing relevant to the (non-)mastery of more ‘obvious’ linguistic phenomena such as grammatical gender?

Language processing in learners: what it is and is not. In our view, ‘language processing’ refers to real-time mental computation that builds, maintains, and queries linguistic representations. ‘Language comprehension’, 'language production', and 'acceptability judgments' simply describe tasks that use this mental machinery in situations where the message is unknown, or the form is unknown, or both are known. Like standard linguistic analysis, our enterprise aims to elucidate which representations are possible, which are not, and why. But it also aims to add detail where standard linguistic analyses set aside issues of time, memory architecture etc.; and it adds mechanisms that address the problem of what to do in situations of uncertainty in the input, whether in comprehension or production. As such, an account of the processing of a given linguistic phenomenon is not a substitute for a standard grammatical analysis, but rather an elaboration at a finer grain of analysis. Note that this does not entail that there is an exact correspondence between what is *acceptable* and what is *representable*. Those notions may diverge in limited ways, which could be helpful or hurtful for learners. (This difference is relevant to a query raised by **de Villiers** at the start of her commentary.)

This approach contrasts with a common perception of language processing research as an exercise in classifying what is easy or difficult in language. It is true that much language processing research relies on measures of ease or difficulty to draw inferences about underlying representations and processes. Nevertheless, the perception is unfortunate. A corollary of this is so-called "processing accounts" of grammatical phenomena, i.e., claims that a generalization that is standardly attributed to grammatical constraints is instead an artifact of language processing difficulty. A classic example is the claim that syntactic 'island' phenomena reflect comprehension difficulty rather than grammatical constraints (Hofmeister & Sag, 2010; Kluender & Kutas, 1993). We distinguish such theories, which we call "reductionist accounts," from theories that specify the implementation of grammatical constraints at finer grains of analysis. Elsewhere we have argued that reductionist accounts are sometimes right, sometimes wrong, and that they need to be evaluated on a case-by-case basis (Phillips, 2013).

This ambiguity is evident in certain commentaries. Some commentators wish we would reduce developmental or grammatical phenomena to effects of difficulty or frequency (**O'Grady**, **Sharwood-Smith**, **Kidd**). For example, O'Grady states his goal as showing that the properties of linguistic rules and representations should be reducible to consequences of "processing pressures".

While such cases exist, these should not be the primary concern. And they may not reduce the learner's burden as much as is sometimes thought: it is little consolation to the learner if a linguistic generalization is an artifact of comprehension or production difficulty, because the learner must somehow distinguish those generalizations from true generalizations about the grammar of the target language. The most basic concern for an account of language processing in learners surely must be whether and how learners assign the correct representations to input sentences. If they misanalyze the input, then they are at risk of learning the wrong grammar (and perceived frequencies will mismatch actual frequencies in the input). That is why it is such a concern that recent laboratory studies on children's language processing have revealed that children are prone to misanalyses of the input, and are strikingly resistant to reanalysis (Omaki & Lidz, 2015; Snedeker, 2013; Trueswell, Sekerina, Logrip & Hill, 1999). The limitations revealed by laboratory studies may turn out to have less damaging effects "in the wild", but this is an area in dire need of further research.

Children are not conservative learners. If children were "conservative" learners, adopting into their grammar only things for which they have unambiguous evidence in the input, then the learning task would be easier (though still far from trivial: Fodor & Sakas, 2005). But children seem not to be conservative learners, and that is a key motivation for our interest in predictive mechanisms in learners. When learners hypothesize an over-general grammar that admits a superset of the expressions allowed by the target grammar, they face the well-known problem that all input sentences from the target grammar appear to be consistent with the incorrect hypothesized grammar. Prediction offers a possible remedy for non-conservative learning.

O'Grady introduces a well-known example of how a conservative learner might take advantage of knowledge of language processing difficulty facts. According to a famous typological

generalization, relativization possibilities in a given language are governed by a hierarchy of possible extraction sites (Keenan & Comrie, 1977). Relativization of subjects is easier, and hence more highly ranked, than relativization of direct objects, which in turn is easier than relativization of indirect objects. Typologically, languages are expected to respect this hierarchy, so if the learner observes relativization from position X, she can infer that the more highly ranked extraction sites are also available for relativization, without directly observing such cases. The practical utility of this strategy may be limited, since the more highly ranked alternatives are likely also more frequent in the input, creating little need to compensate for missing input data. But a more serious concern is that this approach presupposes a relatively conservative learner. This presupposition appears to be incorrect.

Child learners do overgeneralize: there is ample evidence from such diverse domains as verb argument structure (Bowerman, 1982; Pinker, 1989), quantifier scope (Goro, 2007), *wanna*-contraction (Zukowski & Larsen, 2011), and wh-scope marking (Thornton, 1990), among others.

This challenges accounts of learning that rely on the assumption that learners avoid things that they have no evidence for, or that presume that learners' choices are reducible to frequency. Children appear willing to adopt analyses that are 'harder' than what they have encountered in the input, and that have zero frequency in the input. And they show biases in interpreting the input that are unlikely to mirror frequency biases, e.g., Japanese children's bias for long-distance construals of ambiguous wh-questions (Omaki et al., 2014).[1]

We can think of only two possible solutions to overgeneration: (i) denial, (ii) distributional learning. Our account of children's mixed profile of success with constraints on anaphora could be classified as a "denial" account. We argued that children's non-adultlike interpretations of pronouns reflect their limited reanalysis abilities ("adults' first interpretation is children's only interpretation"), rather than a non-adultlike grammar. But we think that distributional learning is necessary as a viable approach to all of children's overgeneralizations. Recovery from overgeneralization requires the learner to recognize that the input is under-generated by her current grammar. Such distributional learning is sometimes called "indirect negative evidence" (Pinker, 1979). Learning models of this type are intuitively attractive, but hard to get to work. That is where prediction becomes useful.

[1] In her commentary, **de Villiers** addresses a discrepancy between the findings that we report for ambiguous English wh-questions (Omaki et al., 2014) and her own findings in a number of previous studies (e.g., de Villiers & Roeper, 1995). We found that English-speaking children showed a bias for a local construal of the wh-phrase, whereas de Villiers and Roeper reported a long-distance bias. We attributed the contrast to de Villiers and Roeper's use of a specific main clause verb, bare *say*, but de Villiers points out that she obtained similar findings with some other verbs. We stand corrected. As de Villiers notes, the plausibility of individual stories likely contributed to the observed biases. However, in the studies by Omaki et al., extreme care was taken to balance the plausibility of the alternative parses, and we tested identical stories with different main clause verbs. Most relevant to the current discussion, Japanese and English-speaking children showed opposite biases in their answers to questions about identical stories. This reversal pattern almost certainly reflects word order differences in the two languages, rather than plausibility or input frequency biases.

The attractive intuition behind using distributional evidence to recover from overgeneralization is this: if something fails to occur, then that must be because it is ungrammatical. But this is not so straightforward. (i) Input to learners is finite, and many things go unsaid. So what counts as sufficient evidence of non-occurrence? (ii) Many things do occur in the input, though, and the learner somehow needs to appropriately tally all of these in order to notice systematic gaps, and also re-classify prior input when her grammatical hypothesis changes. (iii) Learners must also be able to figure out when to attribute a non-occurrence to ungrammaticality. These challenges become especially acute in some of the linguistic ephemera where we encounter overgeneralization in children, such as long-distance movement and quantifier scope.

Predictions and recovery from overgeneralization. Predictive mechanisms are an attractive way to implement a distributional learning mechanism. If learners can use their current grammar to generate expectations about how a sentence will unfold, and compare that against the actual input, then they might use that as a way to incrementally adjust the grammar that is responsible for generating those expectations. This idea has been successfully applied to some simple cases such as subject-verb agreement (Elman, 1993) and verb argument structure (Chang, Dell, & Bock, 2006). It remains to be seen whether it can be scaled up to handle the kinds of more obscure linguistic phenomena where 5-year olds still overgeneralize.

Learning via prediction would be especially useful if learners can encode not only *what* they predict but also *why* they are making that prediction, and so constrain the task of assigning blame for failed predictions. We do not know whether children can do this, but we do have some evidence that adults can track the source as well as the content of their linguistic predictions, based on ERP evidence from Hindi, a language where verb tense can be predicted based on semantic/discourse cues, as in English, but also based on morphosyntax, due to the split ergative case system (Dillon, Nevins, Austin & Phillips, 2012).

Learning by prediction might also help to explain why recovery from overgeneralization takes so long, lasting until age 6 or later. When distributional learning is slow, it is tempting to say that this simply reflects the scarcity of relevant evidence in the input. But that is risky: if the relevant input is so rare that learners typically don't encounter enough of it until age 6, then some learners might never encounter enough. Alternatively, it may be that the relevant input is always present, but cannot be effectively recognized until suitable processing mechanisms develop. **Gabriele, Fiorentino**, and **Johnson** make relevant observations about the relation between language comprehension mechanisms and individual differences in cognitive measures.

In his commentary, **Lew-Williams** pushes back against our suggestion that delayed development reflects the development of language processing abilities, rather than limitations of the input to learners. He argues that pattern recognition in earlier stages of development deserves more attention, and highlights the potentially dramatic effects of individual differences in language experience on language development, especially those arising from socio-economic status (Hart & Risley, 1995; Hoff & Naigles, 2002; Huttenlocher et al., 2010; Rowe, 2008). Lew-Williams is correct that there are many readily observable generalizations that children master early, and that the quality and quantity of input matters to language skills that are essential for

school success. What is much less clear at present, however, is the relation between those aspects of language and the ones our target article focuses on: phenomena that children master late, that are only evident in laboratory settings; obscure generalizations that are underrepresented even in input to privileged children; and phenomena that reveal "native-like" L2 learners to be not quite native-like. These literatures have proceeded largely independently until recently, but we eagerly await the results of new work combining these approaches.

Advanced predictive abilities are our current best guess for how children recover from overgeneralization, and we hope that highlighting the potential value of these mechanisms stimulates further research on them. In this context, **Kaan**'s commentary is particularly relevant. She points out that learning can occur without successful prediction, as shown by examples where L2 learners exhibit grammatical knowledge but are unable to use it to generate rapid predictions. She also argues that some grammatical knowledge cannot be acquired via prediction, since it depends on words that are unpredictable: e.g. in *The butler decided to kill himself*, the reflexive cannot be predicted unless the listener knows in advance what message the speaker wishes to convey. Consequently, Kaan argues, other learning mechanisms are needed.

We agree that prediction is not a cure-all. The specific domain where we see it as most valuable is cases of recovery from overgeneralization. If there are domains where learners are more conservative, then the input can provide errors to guide learning, even in the absence of prediction (for extensive discussion of such examples see Snyder, 2011). And we agree that in some domains, including Kaan's example of anaphora, the utility of learning-by-prediction depends on learners having substantial prior knowledge of the speaker's intended message. An alternative that we have not yet given sufficient attention to is that learners might engage in *post-hoc* comparison of an incoming sentence with how they would have expressed the same message themselves. If such mechanisms exist, they could potentially address some of Kaan's reasonable concerns.

In our target article we suggested that prediction abilities in adult learners should, in principle, be as good as in children, since they should merely be a function of proficiency and fluency in the language, unaffected by maturational constraints. But **Hopp** points out in his commentary several reasons why adult learners might be less confident or effective predictors.

Further contributions of language processing effects. We argued that accurate parsing of the input is unavoidable if learning is to succeed, and that predictive mechanisms are our best bet for ensuring recovery from overgeneralization. But there are other ways that properties of language processing contribute to accounts of learning. In our extended discussion of child-adult parallels in constraints on anaphora we argued that children's well-documented difficulties should be attributed to their limited reanalysis mechanisms rather than to an immature grammar. These findings contribute to an account of learning success only in the respect that they remove some challenges for recovery from overgeneralization, and potentially avoid a mystery involving why some constraints on anaphora are mastered earlier than others. Relatedly, in their commentary, **Pozzan and Trueswell** report interesting cases from Kannada

and Tagalog that show that children are able to respond more effectively to cues that appear early in a sentence. In these examples, language processing mechanisms help explain differences in the time course of two ultimately successful learning achievements. These are certainly important contributions of processing mechanisms, but we distinguish them from mechanisms that are essential for the ability to ever recognize a generalization.

Where do children outperform adult learners? An assumption in our target article is that there is a connection between (i) children's most impressive learning successes, (ii) children's later learning successes, and (iii) domains where adult learners have greater difficulty. This was the basis for our suggestion that children outshine adults specifically in areas that they master at a later age, and for our corresponding skepticism that children's language processing limitations are *directly* responsible for their learning prowess (Newport, 1990). We should distinguish this claim about timing from our additional suggestion that the late-learned phenomena primarily involve complex optional phenomena that relate sentence forms with semantic and pragmatic properties (Coppieters, 1987; Sorace, 2011).

The commentaries take issue with these suggestions in different ways. **Westergaard** argues, following Slabakova (2013), that adult learners' primary difficulty lies in domains such as inflectional morphology that children master rather early (cf. Phillips, 2010). If that is the primary source of child-adult differences, then clearly our hypothesis is in trouble. But we suspect that inflectional morphology is an important but not exclusive barrier for adult learners. **Schaeffer** argues that specific semantic phenomena involving the relationship between specificity and word order in Dutch are mastered (relatively) early. This early learning is not predicted by our claims, but nor is it incompatible with them. It highlights the need to more precisely characterize the time course with which different semantic/discourse phenomena develop in children, and in particular which of them require recovery from overgeneralization.

Learning the Obvious: what's the matter with gender? One linguistic domain that is highlighted by a number of commentators is grammatical gender (**DeKeyser**, **Gabriele et al.**, **Hopp**, **Perez-Leroux, Sekerina**). Some commentators correctly point out that adult learners' struggles with grammatical gender are not predicted by our account. Children master grammatical gender at an early age, but adult learners show great difficulty with it, especially if their native language lacks grammatical gender. Even advanced adult learners who have excellent knowledge of grammatical gender in the L2 show limited ability to use gender as a predictive cue (Grüter, Lew-Williams & Fernald, 2012, but see Hopp, 2013). Children outshine adults at an early age, and prediction and learning clearly don't coincide.

Since grammatical gender ought to be very easy, we focused our attention on learning domains that present clear challenges for learners, due to evidence that is scarce, obscure, or misleading. But the grammatical gender of a noun is straightforwardly observable, and there are no obvious learnability problems, since the memorization involved can be no harder than vocabulary learning. There are other such domains where idiosyncratic properties of words prove difficult even for advanced L2 learners, such as lexical tone in Mandarin (Wong &

Perrachione, 2007), irregular plurals in German and Arabic (Parodi, Schwartz, & Clahsen, 2004), and aspectual alternations in Slavic languages.

We should note that there is little reason to think that children's advantage in learning lexical properties like gender should be related to their limited cognitive development or their language processing difficulties. The "Less is More" hypothesis (Newport, 1990; Elman, 1993) suggests a link between limited memory in children and the need to generalize rather than rote memorize. But rote memorization should go a long way when learning arbitrary lexical properties. (For sure, we often find sub-regularities in grammatical gender systems, irregular plural systems, etc. But we are unaware of evidence that adult learners' failure lies specifically in detecting sub-regularities.) Even in the computational simulations cited by **DeKeyser**, monolingual L1 mastery of native language gender systems shows no benefit from starting with a reduced memory capacity (Monner, Vatz, Morini, Hwang, & DeKeyser, 2013).

Our best guess is that L2 difficulties with grammatical gender and other lexical properties are due to entrenchment of L1 features, rather than to maturational change. (This is consistent with the simulations that DeKeyser cites.) It is becoming clear that even advanced L2 learners who have reliably learned the relevant property of an L2 lexical item (e.g. gender or tone) do not use that property in language processing in the same way as native speakers. The properties are not encoded in memory in the same way for native and non-native speakers, and presumably the nature of these memory encodings is a consequence of L1 learning, rather than maturation. Experimental evidence by Hopp shows that child L2 learners fare well with gender, as long as they have sufficient input (Hopp, 2011).

Distinguishing effects of age vs. L1 is an important issue that a number of commentators highlight (**Foucart**, **Marinis**, **Perez-Leroux**, and **Schimke**). We take it as well established that late learners' difficulties are not exclusively due to L1 interference or entrenchment. Findings about age-of-onset effects for development of L1 sign languages provide rather clear evidence in that regard (Newport, 1990). But we know less about the specific contributions of age vs. L1.

We conclude that when assessing the role of language processing for language learning, whether in L1, L2, or other populations, it is important to distinguish two roles for language processing. First, successful language processing is necessary in order for learners to access critical evidence, without which they simply cannot detect the appropriate generalizations about the target language. This is especially relevant in cases of phenomena that are hard to observe. These are our focus in the target article, and they implicate the ability to accurately parse and reanalyze, and the ability to use prediction to recover from overgeneralization. Second, language processing mechanisms, especially the nature of memory encoding and access, are important for achieving native-like fluency, even in cases where the target language facts should be rather easy to observe. We paid little attention to these in our target article, though memory encoding and access mechanisms are an independent fascination of ours (Phillips, Wagers, & Lau, 2011; Dillon et al., 2014). The contributions of language processing to mastering obvious vs. non-obvious properties of a language should be distinguished in future research, but we concede that they may rely on closely related mechanisms.

References

Bowerman, M. (1982). Reorganizational processes in lexical and syntactic development. In E. Wanner, & L. Gleitman (Eds.), *Language Acquisition: The State of the Art* (pp. 319-346). New York: Academic Press.

Chang, F., Dell, G., & Bock, J. K. (2006). Becoming syntactic. *Psychological Review, 113*, 234-272.

Coppieters, R. (1987). Competence differences between native and near-native speakers. *Language, 63*, 544-573.

de Villiers, J.G. & Roeper,T. (1995) Relative clauses are barriers to Wh-movement for young children. *Journal of Child Language, 22*, 389-404.

Dillon, B., Chow, W.-Y., Wagers, M., Guo, T., Liu, F., & Phillips, C. (2014). The structure-sensitivity of memory access: Evidence from Mandarin Chinese. *Frontiers in Psychology, 5*: 1025. doi: 10.3389/fpsyg.2014.01025

Dillon, B., Nevins, A., Austin, A. C., & Phillips, C. (2012). Syntactic and semantic predictors of tense in Hindi: An ERP investigation. *Language and Cognitive Processes, 27*, 313-344.

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition, 48*, 71–99.

Fodor, J. D. & Sakas, W. (2005). The Subset Principle in syntax: Costs of compliance. *Journal of Linguistics, 41*, 513-569.

Goro, T. (2007). *Language-specific constraints on scope interpretation in first language acquisition*. PhD dissertation, University of Maryland.

Grüter, T., Lew-Wiliams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem. *Second Language Research, 28*, 191-215.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Paul H Brookes Publishing.

Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon. *Child Development, 73*, 418-433.

Hofmeister, P. & Sag, I. (2010). Cognitive constraints and island effects. *Language, 86*, 366-415.

Hopp, H. (2011). Internal and external factors in the child L2 development of the German determiner phrase. *Linguistic Approaches to Bilingualism, 1*, 238-264.

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research, 29*, 33-56.

Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. V. (2010). Sources of variability in children's language growth. *Cognitive Psychology, 61*, 343-365.

Keenan, E., & Comrie, B. (1977). Noun phrase accessibility and Universal Grammar. *Linguistic Inquiry, 8*, 63-100.

Kluender, R. & Kutas, M. (1993). Subjacency as a processing phenomenon. *Language and Cognitive Processes, 8*, 573-633.

Monner, D., Vatz, K., Morini, G., Hwang, S., & DeKeyser, R. (2013). A neural network model of the effects of entrenchment and memory development on grammatical gender learning. *Bilingualism: Language and Cognition, 16*, 246-265.

Newport, E. (1990). Maturational constraints on language learning. *Cognitive Science, 14*, 11-28.
Omaki, A., & Lidz, J. (2015). Linking parser development to acquisition of syntactic knowledge. *Language Acquisition, 22*, 158-192.
Parodi, T., Schwartz, B. D., & Clahsen, H. (2004). On the L2 acquisition of the morphosyntax of German nominals. *Linguistics, 42*, 669-705.
Phillips, C. (2010). Syntax at age two: cross-linguistic differences. *Language Acquisition, 17*, 70-120. [Originally published 1995 in MIT Working Papers in Linguistics, vol. 26.]
Phillips, C. (2013). Some arguments and non-arguments for reductionist accounts of syntactic phenomena. *Language and Cognitive Processes, 28*, 156-187.
Phillips, C. & Ehrenhofer, L. (2015). The role of language processing in language acquisition. *Linguistic Approaches to Bilingualism*.
Phillips, C., Wagers, M. W., & Lau, E. F. (2011). Grammatical illusions and selective fallibility in real-time language comprehension. In J. Runner (ed.), *Experiments at the Interfaces* (Syntax and Semantics, vol. 37), pp. 147-180. Bingley, UK: Emerald Publications.
Pinker, S. (1979). Formal models of language learning. *Cognition, 7*, 217-283.
Pinker, S. (1989). *Learnability and Cognition*. Cambridge, MA: MIT Press.
Rowe, M. L. (2008). Child-directed speech: Relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of Child Language, 35*, 185-205.
Slabakova, R. (2013). What is easy and what is hard to acquire in a second language: A generative perspective. In M. García Mayo, M. Junkal Gutierrez Mangado, & M. Martínez Adrián (Eds.), *Contemporary approaches to second language acquisition*, (pp. 5-28). Amsterdam: John Benjamins.
Snedeker, J. (2013). Children's sentence processing. In R. van Gompel (ed.), *Sentence Processing*, pp. 189-220. New York, New York: Psychology Press.
Snyder, W. (2011). Children's grammatical conservatism: Implications for syntactic theory. In N. Danis, K. Mesh, & H. Sung (eds.), *BUCLD 35: Proceedings of the 35th Annual Boston University Conference in Language Development*, vol. 1, pp. 1-20. Somerville, MA: Cascadilla Press.
Sorace, A. (2011). Pinning down the concept of "interface" in bilingualism. *Linguistic Approaches to Bilingualism, 1*, 1-33.
Thornton, R. (1990). *Adventures in long-distance moving: The acquisition of complex wh-questions*. PhD dissertation, University of Connecticut.
Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition, 73*, 89-134.
Wong, P. C. M., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics, 28*, 565–585.
Zukowski, A. & Larsen, J. (2011). Wanna-contraction in children: Retesting and revising the developmental facts. *Language Acquisition, 18*, 211-241.

colin@umd.edu
ehrenhof@umd.edu