**Syntactic categories in child language acquisition: innate, induced or illusory?**

Citation: Ambridge, B. (in press). Syntactic categories in child language acquisition: innate, induced or illusory? In Cohen, H. & Lefebvre, C. (Eds). *Handbook of Categorization in Cognitive Science*. Amsterdam: Elsevier.

Ben Ambridge

University of Liverpool

ESRC International Centre for Language and Communicative Development (LuCiD)

Psychological Sciences, Institute of Psychology, Health and Society, University of Liverpool, Bedford Street South, L69 7ZA.

**Running head: Syntactic categories in child language acquisition**

Ben Ambridge is a Reader in the International Centre for Language and Communicative Development (LuCiD) at The University of Liverpool. The support of the Economic and Social Research Council [ES/L008955/1] is gratefully acknowledged.

**Summary/Abstract**

Traditionally, it has generally been assumed that adult speakers are in possession of syntactic categories such as NOUN (e.g., *boy, girl*) and VERB (e.g., *see, dance*). This chapter evaluates three possibilities for how children acquire these categories. The first is that syntactic categories are **innate** (i.e., present from – or even before – birth) and that acquisition consists of using incoming words’ semantic or distributional properties to assign them to the relevant category. The second possibility is that syntactic categories are **induced** from the input, via some kind of distributional clustering procedure (e.g., words that frequently appear after both *the* and *a* are often NOUNs). On this view, acquisition consists of gradually building classes that are, ultimately, largely the same as those posited under nativist approaches. The third possibility is that syntactic categories are **illusory**, and that acquisition consists of storing individual words, phrases and sentences, along with their meanings. Speakers produce and comprehend novel utterances by generalizing across these stored *exemplars*, but at no point form a free-standing abstraction that corresponds to a category such as NOUN or VERB. It will be argued that, although implausible on the surface, the third approach is ultimately the one most likely to yield a successful explanation.

**1.0 Introduction**

*’Twas brillig, and the slithy toves*

*Did gyre and gimble in the wabe.*

When it comes to discussions of syntactic category acquisition, an opening quotation from Lewis Carrol’s *Jabbwerocky* has become a cliché of the genre, but with good reason: Even if you have somehow never heard this poem before, it will be immediately clear that – for example *– I saw a tove; The tove danced; The boy gyred* and *The girl often gimbles* are possible sentences, whereas *\*A Tove saw I, \*Danced the tove, \*Gyred the boy* and *\*Gimbles often the girl* are not.

 The traditional explanation for such phenomena is that adult speakers have (a) syntactic categories such as NOUN (e.g., *boy, girl*) and VERB (e.g., *see, dance*), into which newly-learned words can be assimilated, and (b) knowledge of how these categories can be combined into phrases and sentences; for example, English sentences place the NOUN PHRASE before the VERB PHRASE, rather than after (*The tove danced*, not *\*Danced the tove*). Focussing mainly on lexical categories such as NOUN and VERB (as opposed to functional categories such as COMPLEMENTIZER and INFLECTION), the present chapter investigates three different types of proposal for how syntactic categories are acquired.

 The first holds that syntactic categories are **innate** (i.e., present from – or even before – birth) and that acquisition consists of assigning words in the input to the appropriate categories. The second holds that syntactic categories are **induced** from the input, and that acquisition consists of gradually building classes that are, ultimately, largely the same as those posited under nativist approaches (at least for lexical categories). The third holds that syntactic categories are **illusory**, and that acquisition consists of storing individual words, phrases and sentences, along with their meanings. Speakers produce and comprehend novel utterances by generalizing across these stored *exemplars*, but at no point form a free-standing abstraction that corresponds to a category such as NOUN or VERB.

 Many previous debates on the acquisition of syntactic categories (e.g., Valian-Pine-Yang on DETERMINER; Tomasello-Fisher on VERB) have focussed on the question of *when* these categories are acquired; i.e., “adult-like” (Valian, 1986; Pine & Martindale, 1996; Pine & Lieven, 1997; Valian, Solt & Stewart, 2009; Yang, 2010; Pine, Freudenthal, Krajewski & Gobet, 2013; Meylan, Frank, Roy & Levy, submitted; Tomasello, 2000; Fisher, 2002; Tomasello & Abbot-Smith, 2002; see Ambridge & Lieven, 2015, for a review). Rather than following this well-trodden path, the present chapter approaches the debate from a different angle, by focusing on the *what* and *how* of syntactic category acquisition. For each type of theory – innate, induced, illusory – we ask (a) *What* does adults’ knowledge of syntactic categories look like, and are the representations assumed consistent with the data? and (b) *How* do children acquire this knowledge, and are the acquisition mechanisms assumed consistent with the data?

**2.0 Innate syntactic categories**

A particularly clear statement of the nativist position is provided by Valian, Solt and Stewart (2009: 744).

The child begins with an abstract specification of syntactic categories and must learn the details of how those categories behave in her target language… Since, on our approach, the child begins with the abstract category, it is a form of nativism

In terms of adult representations, the classes assumed under such approaches are the linguist’s traditional basic level “syntactic categories such as Noun [e.g., *boy, girl*], Verb [e.g., *see, dance*], Preposition [e.g., *in, on*], and Determiner [e.g., *the, a*]” (Valian, 2014: 80 [examples added]). The question of the extent to which these representations are consistent with the adult data is one that is very difficult to answer. One possible argument against this claim is that languages vary greatly with regard to the syntactic categories that they employ (e.g., Haspelmath, 2007; Evans & Levinson, 2009). A counterargument is that these categories are merely a “toolkit” (Jackendoff, 2002: 263) from which individual languages may select. A second argument against the claim that adult knowledge is best characterized in terms of basic-level syntactic categories (e.g., VERB, NOUN) is that – *Jaberwocky* notwithstanding – membership of such coarse categories is actually quite a poor predictor of possible uses. For example, both *amuse* and *giggle* are VERBs, yet one can *amuse someone*, but not *\*giggle someone*. Similarly, both *snow* and snowflake are NOUNs, yet one can talk about *a snowflake* but not *\*a snow*. A counterargument is that the existence of basic-level syntactic categories does not preclude the existence of more detailed subcategories (e.g., intransitive vs transitive verbs [*giggle* vs *amuse*]; count vs mass nouns [*snowflake* vs *snow*]), and is required to explain some (almost?) exceptionless generalizations. For example, every English (main) VERB takes the morpheme *–s* to mark third-person singular present tense [e.g., *he see****s****; he dance****s***].

 Assuming, for the sake of argument, that every language contains at least one basic level syntactic category (perhaps NOUN) that captures at least some important distributional facts about the adult grammar, we now consider possible explanations for how children could identify instances of this category in the input.

 Pinker’s (1984) ***semantic bootstrapping*** hypothesis assumes not just innate syntactic categories of NOUN and VERB but also innate linking rules which map “name of person or thing” onto the former and “action or change of state” onto the latter (p.41). The advantage of this approach is that, unlike syntactic categories, semantic categories like “person or thing” are directly observable in both the real world and – provided the child knows the label – the linguistic input. At least under the original version of the theory, these linking rules are used merely to break into the system, and are rapidly discarded in favour of distributional analysis (of the type discussed in more detail in the “Induced” section). This allows children to assimilate nouns that do not name a person or thing (e.g., *idea, situation*) and verbs that do not name an action or state-change (e.g., *justify, cost*) into the relevant categories. As Pinker (1984: 61) acknowledges, a potential problem is that children will occasionally hear sentences such as *You will get a spanking* in which the action is denoted by a noun rather than a verb. In a later version of the theory, Pinker (1987) attempts to deal with this problem by making the linking rules probabilistic. A problem with this solution (Braine, 1992) is that it makes Pinker’s theory difficult to differentiate from non-nativist theories which assume that children group together items with similar semantic properties and syntactic distributions (as discussed in subsequent sections).

Christophe, Millotte, Bernal and Lidz’s (2008) ***prosodic bootstrapping*** hypothesis assumes that children first use prosodic information to split clauses into syntactic phrases (e.g. *[The boy] [is running]*; see Soderstrom, Seidl, Kemler-Nelson & Jusczyk, 2003, for empirical evidence). The child then uses function-word ‘flags’ – here, *the* and *is* – to assign the lexical content words to the relevant categories; here, *boy* to NOUN and *running* to VERB. Ambridge, Pine and Lieven (2014: e59) outline a number of problems for this hypothesis, but perhaps the major one is that these flags do not exist in all languages for which a NOUN and VERB category would normally be assumed (e.g., the equivalent Russian sentence *Mal'chik bezhit* consists solely of a noun and a verb). Even if this problem could be solved (perhaps, for example, by using noun case-marking and verb person-number marking morphemes as flags), it is unclear how children would know which word or morpheme is a flag to which category. For example, the knowledge that DETERMINERs (*the, a*) flag NOUN PHRASEs could be innately specified; but this raises the problem of how children know that *the* and *a* are determiners.

Mintz’s (2003: 112) ***frequent frames*** hypothesis is essentially a distributional-learning proposal of the type discussed in the “Induced” section, but differs in that it additionally posits a possible mechanism to “guide the linking between distributional categories and grammatical categories”, specifically assigning the label NOUN to the distributionally-defined cluster that contains concrete objects (using a Pinker-style innate linking rule). The label VERB would then be assigned to the largest unlabeled category – or if this does “not turn out to be viable cross-linguistically…the category whose members take the nouns as arguments”. Because the linking rule plays a role only after distributionally-defined clusters have been formed, Mintz’s (2003) learner is more resilient than Pinker’s (1984) to utterances with non-canonical linking (e.g., *You will get a spanking*; see above). As such, Mintz’s proposal is probably the most promising of the nativist accounts discussed here. A weakness of this account, however, is its failure to offer an explanation of how children link their distributionally-defined clusters to categories other than NOUN and VERB (e.g., DETERMINER, AUXILIARY, WH-WORD, and PRONOUN) that are also generally assumed to be innately specified by Universal Grammar.

In summary, nativist accounts have yet to solve the bootstrapping problem and provide a workable account of how the child identifies in the input all of the major syntactic categories. Indeed, the generativist research programme seems to have largely abandoned this question. In some ways this is surprising, given that if the bootstrapping problem could be solved, the rewards – in terms of explanatory power of nativist theory – would be huge.

**3.0 Induced syntactic categories**

While arguing against the idea that syntactic categories are innate, most researchers in the constructivist tradition make reference to categories such as NOUN and VERB when discussing the adult end-state. That said, it is often unclear whether such terms represent a theoretical commitment to something like traditional linguistic categories in the adult grammar (though not the young child’s) or simply a convenient shorthand for particular types of generalizations, perhaps including on-the-fly exemplar-based generalizations that are not represented independently (as discussed in the “Illusory” section). Indeed, many constructivist researchers have probably never asked themselves this question, or have done so and decided to remain agnostic. One exception is Tomasello (2003: 301) who discusses “paradigmatic categories such as noun and verb”, arguing that they are

formed through a process of functionally-based distributional analysis in which concrete linguistic items (such as words or phrases) that serve the same communicative function in utterances and constructions over time are grouped together into a category.

Although, in later work (Abbot-Smith & Tomasello, 2006), Tomasello argues for exemplar-based learning (as discussed in the “Illusory” section), the account proposed is “a kind of hybrid model comprising both abstractions and the retention of the exemplars of which those abstractions are composed” (p.276). Although the abstractions discussed by Abbot-Smith and Tomasello (2006) are mainly at the level of constructions (e.g., the SVO English transitive construction [*The man kicked the ball*]), presumably these include categories that correspond at least roughly to traditionally syntactic categories (e.g., the VERB slot of the SVO construction).

 Thus, at least on my reading of the literature, constructivist accounts generally assume that, although certainly absent at the start of development, something like traditional syntactic categories are present in the adult grammar (hence these accounts inherit many of the advantages and disadvantages of nativist approaches that share the latter part of this assumption). This raises the question of how children form these categories. The constructivist answer is via some form of distributional analysis. As we have seen above, many nativist accounts also assume a role for distributional analysis (e.g., Chomsky, 1955: section 34.5; Yang, 2008: 206; Pinker, 1984: 59; Mintz, 2003: 112; Valian et al, 2009: 744). The difference is that constructivist accounts assume that the distributionally-defined clusters are used directly in comprehension and production, rather than being linked up to innate categories.

Essentially, distributional analysis involves probabilistically grouping together words that appear in a similar set of distributional contexts; usually a window of predetermined size either side of the target (e.g., *a \_ is; the \_ on*). As this example illustrates, the type of distributional analysis that is typically investigated in experimental and computational clustering studies is based purely on form, rather than function or meaning Although this decision is understandable (since coding for function would be both laborious and very difficult to do objectively), it means that such studies are likely to underestimate the power of the *functionally-based* distributional analysis that, according to Tomasello (2003), children perform.

 That said, the classifications yielded on the basis of purely formal distributional analysis are reasonably impressive (see Cassani, Grimm, Daelemans and Gillis, in press, for a recent review, as well as a study of the type of distributional contexts that are most useful for classification). Computational studies (e.g., Cartwright & Brent, 1997; Chrupała & Alishahi, 2010; Leibbrandt, 2009; Mintz, 2003; Monaghan & Christiansen, 2004; Redington, Chater, & Finch, 1998; St. Clair, Monaghan, & Christiansen, 2010; Chemla, Mintz, Bernal & Christophe, 2009; Wang, Höhle, Ketrex, Küntay, & Mintz, 2011; though see Erkelens, 2008; Stumper, Bannard, Lieven & Tomasello, 2011) typically yield accuracy scores in the region of 70-80%, though for many models, completeness scores are much lower (often under 50%). (Accuracy is the percentage of words classified as – say – nouns that are actually nouns. Completeness is the percentage of words that are actually nouns that the model correctly classifies as such). Similar findings have also been observed in experimental artificial grammar learning studies (e.g., Mintz, Wang, & Li, 2014; Reeder, Newport, & Aslin, 2013; Monaghan, Chater & Christiansen, 2005), and studies conducted with infants (e.g., van Heugten & Johnson, 2010; Zhang, Shi, & Li, 2014). Finally, a number of corpus studies have demonstrated that useful distributional information is available in child-directed speech (e.g., Mintz, Newport & Bever, 2002; Monaghan et al, 2005; Feijoo, Muñoz, & Serrat, 2015).

 In an important sense, however, the success of distributional learning as a mechanism for inducing syntactic categories is unsurprising. After all, while some may have rough semantic correlates (e.g., NOUNs are often things [e.g., *ball*], VERBs often actions [*kick*]), ultimately syntactic categories are *defined* distributionally (e.g., *situation* and *cost* are, respectively, a NOUN and a VERB, but not a thing and an action). Thus, the fact that large corpora contain distributional cues to syntactic categories that can be uncovered by suitable algorithms is not so much an empirical finding as true by definition. The important question is *how* child learners exploit this distributional information (certainly not, as many models do, by batch-processing a large corpus).

 One attempt at a developmentally plausible account of distributional category formation is the model of Freudenthal, Pine, Jones and Gobet (2016). These authors applied the distributional learning approach of Redington et al (1998) to a previous computational model that learns language from the input in a slow and gradual fashion, by storing progressively-longer utterances (i.e., initially storing only the final word of an utterance, then the final two words and so on). This severely restricts the amount of distributional information that the model needs to consider, in line with the assumption that young children have a relatively low processing capacity (e.g., Gathercole, Pickering, Ambridge & Waring, 2004). For the same reason, the model does not store detailed frequency information, but simply a list of the words that have preceded or followed each target word in the input. Even with these modifications in place, this model shows similar accuracy levels to the much more computationally intensive models outlined above, and also simulates the so-called noun-bias in children’s learning. A problem for this model, however, is that completeness scores are low, even by the standards of the clustering models discussed above. Regardless of the successes and failures of this particular model, it highlights the importance of considering developmental plausibility as an important criterion for future models of syntactic-category induction.

**4.0 Illusory syntactic categories**

Although few researchers have proposed *exemplar-based* accounts of child language acquisition in particular (though see Abbot-Smith & Tomasello, 2006; Bod, 2009), exemplar-based models of language are well established in both theoretical linguistics (e.g., Langacker, 2000; Gahl & Yu, 2006; Bybee, 2013) and the computational literature (e.g., Skousen, 1992; Nosofsky, Kruschke & McKinley, 1992; Daelemans, Zavrel, Van der Sloot & Van den Bosch, 2010), as well as the categorization literature more generally (e.g., Medin, Altom & Murphy, 1984; Nosofsky, 1984; though see Smith, 2002). The distinction between the present chapter’s “Induced” and “Illusory” syntactic categories, as succinctly set out by Abbot-Smith and Tomasello (2006: 276), is

whether learners develop abstractions that supersede (and essentially efface) the experienced exemplars…or, alternatively, whether generalization occurs solely through online analogical comparison to a set of previously learned exemplars.

At first glance, exemplar-based models appear highly implausible, as they require speakers to store literally millions of utterances; perhaps every utterance ever encountered. Indeed, the *prima facie* impossibility of such rampant storage – as well as the apparent redundancy involved – is presumably what led linguists to propose syntactic categories in the first place: We surely can’t learn how each and every individual verb behaves, but we can learn how VERB – the category – behaves.

 However, when we set aside our intuitions and examine the empirical evidence, there is actually considerable support for an exemplar-based view of linguistic knowledge. Studies of sentence processing (e.g., Clifton, Frazier & Connine, 1984; Trueswell, Tanenhaus and Kello, 1993; MacDonald, 1994; Gahl, Jurafsky & Roland, 2004) show that adults are acutely sensitive to the fine-grained distributional properties of individual verbs, for example that *remember* is more often followed by **noun phrase** than ***sentence complements***

The detective remembered **the woman** > The detective remembered (***that) the woman was tall***

while *suspect* shows the opposite pattern

The detective suspected (***that) the woman was tall*** > The detective suspected **the woman**

Similar studies provide evidence that adults are sensitive to the frequency of particular *n*-grams; i.e., the frequency with which they have encountered particular two-word combinations (bigrams, e.g., *cup+of*) three-word combinations (trigrams, e.g., *cup+of+tea*) and so on (e.g., Sosa & MacFarlane, 2002;McDonald & Shillcock, 2003; Tremblay & Baayen, 2010; Arnon & Snider, 2010; Siyanova-Chanturia, Conklin, and van Heuven, 2011; Janssen & Barber, 2012). One way to account for this knowledge would be to posit that learners maintain – and continuously update in real-time – an inventory of the frequency with which every possible bigram, trigram (and so on) has occurred in the input. By comparison with this possibility, the exemplar-based alternative – that learners are simply storing the original sentence strings – begins to look far less implausible.

 The fact that detailed probabilistic information about the attested combinatorial properties of individual words is (somehow) stored in learners’ memories has an important implication for the debate regarding syntactic category acquisition. In addition to storing detailed probabilistic information about how each and every individual verb behaves, learners *could* also store information about how an induced VERB category behaves. And it might well turn out that they do. But they do not *need* to: All of the information that would be stored at the level of an induced VERB category is already stored in memory. On the assumption that individual lexical strings are stored, all the distributional information that a clustering algorithm would use to induce a VERB category is already there. Thus standalone syntactic categories need not be represented at all. The work traditionally done by a category can instead be done by an on-the-fly analogical generalization across the stored exemplars. As Langacker (2000) puts it, a VERB category (or any other category or generalization) is “immanent in its instantiations”.

 Thus, there is reasonably good evidence for at least some degree of exemplar representation amongst adult speakers. When it comes to building an exemplar-based account of language *acquisition*, the devil is in the detail. It is one thing to provide evidence that learners could in principle acquire language by storing strings and generalizing across them; quite another to provide a testable account of how this generalization procedure works, which, ultimately, would presumably have to take the form of a computational model. To date, no such detailed account has been proposed. Although there exist models that generalize across input strings in a limited way (e.g., Freudenthal, Pine & Gobet, 2005; 2016), both their performance and their psychological plausibility are limited by the fact that they operate solely on the basis of surface form, and do not represent words’ functional or semantic properties. Given that such similarities are almost certainly a powerful cue to shared syntactic category membership for real learners (e.g., Tomasello, 2003), a future goal must be to build models that incorporate this type of information.

 In the meantime it is important to note that, at the level of surface form, even young children demonstrate considerable evidence of exemplar storage. This includes knowledge of verb’s fine-grained distributional properties (e.g., that *give* occurs more frequently in double-object than prepositional datives, e.g., *Give me that > Give that to me*) and – more generally – of *n-*gram frequencies (e.g., that *drink+of+milk* is a more frequent trigram than *drink+of+tea* in child-directed English [even British English!]). Evidence of this knowledge comes from a number of different paradigms including grammaticality judgments (e.g., Ambridge, Bidgood, Twomey, Pine, Rowland & Freundenthal, 2015), syntactic priming (e.g., Peter, Chang, Pine, Blything & Rowland, 2015) and elicited production/imitation tasks (Bannard & Matthews, 2008; Matthews & Bannard, 2010; Arnon & Clark, 2011). Thus, while we have barely begun to understand how they generalize across them, the fact that even young children store multi-word linguistic exemplars is already well established.

**5.0 Conclusion**

Although positing **innate** categories might appear, on the surface, to solve the problem of syntactic category acquisition, this is simply to replace one difficult question with another: how learners identify instances of these innate categories in their input. Accounts under which syntactic categories are **induced** are on a stronger empirical footing, given that (unlike, for example, the *semantic*- and *prosodic-bootstrapping* hypotheses) they have been implemented as testable computational models that yield reasonably accurate classifications. However, given the evidence that both adults and children copiously store exactly the kind of distributional information on which these clustering algorithms rely, there seems to be no good reason to assume that learners additionally store the outputs of these algorithms in the form of free-standing syntactic categories. Thus, although there remains much work to be done to arrive at an account of exactly *how* this process works, an account under which learners produce and comprehend language by generalizing across stored exemplars seems to offer the most promising solution. As such, my tentative conclusion is this: Syntactic categories are **illusory**.

**References**

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

Aha, D. W., Kibler, D., & Albert, M. K. (1991). Instance-based learning algorithms. *Machine Learning, 6*(1), 37-66. doi:10.1023/A:1022689900470.

Alishahi, A., & Chrupała, G. (2009). Lexical category acquisition as an incremental process. Paper presented at the CogSci-2009 Workshop on Psychocomputational Models of Human Language Acquisition, Amsterdam.

Alishahi, A., & Chrupała, G. (2012). Concurrent Acquisition of Word Meaning and Lexical Categories. In J. i. Tsujii, J. Henderson, & M. Pasca (Eds.), 2012 Joint Conference on Empirical Methods in Natural Language Processing (EMNLP) and Computational Natural Language Learning (CoNLL) (pp. 643-654). Stroudsburg, PA: Association for Computational Linguistics.

Ambridge, B., & Lieven, E.V.M. (2015). A Constructivist account of child language acquisition. In B. MacWhinney and W. O'Grady (Eds). *Handbook of Language Emergence.* Hoboken, NJ: Wiley Blackwell. (pp. 478-510).

Ambridge, B., Bidgood, A., Twomey, K. E., Pine, J. M., Rowland, C. F., & Freudenthal, D. (2015). Preemption versus Entrenchment: Towards a Construction-General Solution to the Problem of the Retreat from Verb Argument Structure Overgeneralization. *PloS one*, *10*(4), e0123723.

Ambridge, B., Kidd, E., Rowland, C. F., & Theakston, A. L. (2015). The ubiquity of frequency effects in first language acquisition. *Journal of Child Language, 42*(2), 239-273. doi:10.17/S030500091400049X

Ambridge, B., Pine, J. M., & Lieven, E. V. M. (2014). Child language acquisition: Why universal grammar doesn't help. *Language, 90*(3), e53-e90. doi:10.1353/lan.2014.0051

Arnon I, Snider N (2010) More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language, 62*: 67–82. doi: 10.1016/j.jml.2009.09.005.

Arnon, I. & Clark, E. V. (2011). When ‘on your feet’ is better than ‘feet’: Children’s word production is facilitated in familiar sentence-frames. *Language Learning and Development, 7*, 107-129.

Bannard, C., & Matthews, D. (2008). Stored Word Sequences in Language Learning The Effect of Familiarity on Children's Repetition of Four-Word Combinations. *Psychological Science*, *19*(3), 241-248.

Bod, R. (2009). From exemplar to grammar: A probabilistic analogy‐based model of language learning. *Cognitive Science*, *33*(5), 752-793.

Bod, R. (2009). From exemplar to grammar: A probabilistic analogy‐based model of language learning. *Cognitive Science*, *33*(5), 752-793.

Braine, M. D. (1992). What sort of innate structure is needed to “bootstrap” into syntax?. Cognition, 45(1), 77-100.

Braine, M. D.S. (1992). What sort of innate structure is needed to "bootstrap" into syntax? *Cognition, 45*(1), 77-100.

Bybee, J. (2013). Usage-based theory and exemplar representations of constructions. *The Oxford handbook of construction grammar*. Oxford: Oxford University Press (pp. 49-69).

Bybee, J. (2013). Usage-based theory and exemplar representations of constructions. *The Oxford handbook of construction grammar*, 49-69. Oxford: Oxford University Press.

Cartwright, T. A., & Brent, M. R. (1997). Syntactic categorization in early language acquisition: Formalizing the role of distributional analysis. Cognition, 63(2), 121-170. doi:10.1016/S0010-0277(96)00793-7

Cassani, G., Grimm R., Daelemans W., & Gillis S. (2015). Which distributional cues help the most? Unsupervised contexts selection for lexical category acquisition.

Chemla, E., Mintz, T. H., Bernal, S., & Christophe, A. (2009). Categorizing words using ‘frequent frames’: what cross-linguistic analyses reveal about distributional acquisition strategies. Developmental Science, 12(3), 396-406. doi:10.1111/j.1467-7687.2009.00825.x

Chomsky, N. (1955). The Logical structure of linguistic theory. Retrived from <http://alpha-leonis.lids.mit.edu/wordpress/wp-content/uploads/2014/07/chomsky_LSLT55.pdf>

Christophe, A., Millotte, S., Bernal, S., & Lidz, J. (2008). Bootstrapping lexical and syntactic acquisition. *Language and Speech, 51*, 61-75.

Chrupała, G., & Alishahi, A. (2010). Online Entropy-based Model of Lexical category Acquisition. In M. Lapata & A. Sarkar (Eds.), Proceedings of the 14th Conference on Computational Natural Language Learning (CoNLL) (pp. 182-191). Stroudsburg, PA: Association for Computational Linguistics.

Clifton, C., Frazier, L., & Connine, C. (1984). Lexical expectations in sentence comprehension. *Journal of Verbal Learning and Verbal Behavior, 23*, 696–708.

Daelemans, W., Zavrel, J., van der Sloot, K., & Van den Bosch, A. (2010). Timbl: Tilburg memory-based learner. ILK Technical Report – ILK 10-01.

Erkelens, M. A. (2008). Restrictions of frequent frames as cues to categories: the case of Dutch. In H. Chan, E. Kapia, & H. Jacob (Eds.), Proceedings of the 32nd Boston University Conference on Language Development (BUCLD 32). Boston, MA: Cascadilla press.

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*(05), 429-448.

Fisher, C. (2002). The role of abstract syntactic knowledge in language acquisition: a reply to Tomasello (2000). *Cognition*, *82*(3), 259-278.

Frank, S., Goldwater, S. J., & Keller, F. (2009). Evaluating Models of Syntactic Category Acquisition without Using a Gold Standard. In N. Taatgen & H. van Rijn (Eds.), Proceedings of the 31st Annual Conference of the Cognitive Science Society (pp. 2576-2581). Austin, TX: Cognitive Science Society.

Freudenthal, D., Pine, J. M., & Gobet, F. (2005). On the resolution of ambiguities in the extraction of syntactic categories through chunking.*Cognitive Systems Research*, *6*(1), 17-25.

Freudenthal, D., Pine, J.M., Jones, G., & Gobet, F. (2016). Developmentally plausible learning of word categories from distributional statistics. Paper presented at 38th Annual Meeting of the Cognitive Science Society, Philadelphia, PA.

Gahl, S., & Yu, A. C. (2006). Introduction to the special issue on exemplar-based models in linguistics. *The linguistic review*, *23*(3), 213-216.

Gahl, S., Jurafsky, D., & Roland, D. (2004). Verb subcategorization frequencies: American English corpus data, methodological studies, and cross-corpus comparisons. *Behavior Research Methods, Instruments, and Computers, 36*, 432–443.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental psychology*, *40*(2), 177.

Haspelmath, M. (2007). Pre-established categories don't exist: Consequences for language description and typology. *Linguistic Typology*, *11*(1), 119-132.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford University Press.

Jacobs, C. L., Dell, G. S., Benjamin, A. S., & Bannard, C. (2016). Part and whole linguistic experience affect recognition memory for multiword sequences. *Journal of Memory and Language, 87,* 38-58. doi: 10.1016/j.jml.2015.11.001

Janssen, N. & Barber, H.A. (2012) Phrase Frequency Effects in Language Production. PLoS ONE 7(3): e33202. doi:10.1371/journal.pone.0033202

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

MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes, 9,* 157–201.

Matthews, D., & Bannard, C. (2010). Children’s Production of Unfamiliar Word Sequences Is Predicted by Positional Variability and Latent Classes in a Large Sample of Child‐Directed Speech. *Cognitive Science*, *34*(3), 465-488.

McDonald, S. A., & Shillcock, R. C. (2003a). Eye movements reveal the on-line computation of lexical probabilities. *Psychological Science, 14*, 648–652.

Medin, D. L., Altom, M. W., & Murphy, T. D. (1984). Given versus induced category representations: Use of prototype and exemplar information in classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(3), 333.

Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. Cognition, 90(1), 91-117. doi:10.1016/S0010-0277(03)00140-9

Mintz, T. H., Newport, E. L., & Bever, T. G. (2002). The distributional structure of grammatical categories in speech to young children. Cognitive Science, 26(4), 393-424. doi:10.1207/s15516709cog2604\_1.

Mintz, T. H., Wang, F. H., & Li, J. (2014). Word categorization from distributional information: Frames confer more than the sum of their (Bigram) parts. Cognitive Psychology, 75, 1-27. doi:10.1016/j.cogpsych.2014.07.003

Monaghan, P., Chater, N., & Christiansen, M. H. (2005). The differential role of phonological and distributional cues in grammatical categorisation. Cognition, 96(2), 143-182. doi:10.1016/j.cognition.2004.09.001

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(1), 104.

Nosofsky, R. M., Kruschke, J. K., & McKinley, S. C. (1992). Combining exemplar-based category representations and connectionist learning rules.*Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(2), 211.

Peter, M., Chang, F., Pine, J. M., Blything, R., & Rowland, C. F. (2015). When and how do children develop knowledge of verb argument structure? Evidence from verb bias effects in a structural priming task. *Journal of Memory and Language*, *81*, 1-15.

Peter, M., Chang, F., Pine, J. M., Blything, R., & Rowland, C. F. (2015). When and how do children develop knowledge of verb argument structure? Evidence from verb bias effects in a structural priming task. *Journal of Memory and Language*, *81*, 1-15.

Pine, J. M., & Lieven, E. V. M. (1997). Slot and frame patterns and the development of the determiner category. *Applied Psycholinguistics, 18*(2), 123-138.

Pine, J. M., & Martindale, H. (1996). Syntactic categories in the speech of young children: The case of the determiner. *Journal of Child Language, 23*(2), 369-395.

Pinker, S. (1984). *Language learnability and language development.*Harvard University Press: Cambridge, MA.

Pinker, S. (1987). The bootstrapping problem in language acquisition. In B. MacWhinney (Ed.), Mechanisms of language acquisition. Hillsdale, NJ: Lawrence Erlbaum Associates.

Pinker, S. (1987). The bootstrapping problem in language acquisition. In B. Macwhinney (Ed.), *Mechanisms of language acquisition* (pp. 339-441). Hillsdale, NJ: Lawrence Erlbaum.

Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. Cambridge: MIT.

Redington, M., Chater, N., & Finch, S. (1998). Distributional Information: A Powerful Cue for Acquiring Syntactic Categories. Cognitive Science, 22(4), 425-469. doi:DOI 10.1207/s15516709cog2204\_2

Reeder, P. A., Newport, E. L., & Aslin, R. N. (2013). From shared contexts to syntactic categories: The role of distributional information in learning linguistic form-classes. Cognitive Psychology, 66(1), 30-54. doi:10.1016/j.cogpsych.2012.09.001

Siyanova-Chanturia, A., Conklin, K., & van Heuven, W.J.B. (2011) Seeing a phrase “time and again” matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition 37*: 776.

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

Smith, J. D. (2002). Exemplar theory's predicted typicality gradient can be tested and disconfirmed. *Psychological Science*, *13*(5), 437-442.

Soderstrom, M., Seidl, A., Kemler-Nelson, D.G. & Jusczyk, P. W. (2003). The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *Journal of Memory and Language, 49*(2), 249-267.

Sosa, A.V. & MacFarlane, J. (2002) Evidence for frequency-based constituents in the mental lexicon: Collocations involving the word of. *Brain and Language 83*: 227–236.

St. Clair, M. C., Monaghan, P., & Christiansen, M. H. (2010). Learning grammatical categories from distributional cues: Flexible frames for language acquisition. Cognition, 116(3), 341-360. doi:10.1016/j.cognition.2010.05.012

Stumper, B., Bannard, C., Lieven, E. M., & Tomasello, M. (2011). ‘‘Frequent Frames’’ in German Child-Directed Speech: A Limited Cue to Grammatical Categories. Cognitive Science, 35(6), 1190-1205. doi:10.1111/j.1551-6709.2011.01187.x

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition, 74*(3), 209-253.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Tomasello, M., & Abbot-Smith, K. A tale of two theories: Response to Fisher. *Cognition, 83*(2), 207-214.

Tremblay, A. and Baayen, R. H. (2010) Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In D. Wood, *Perspectives on formulaic language: Acquisition and communication*. London: The Continuum International Publishing Group

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.

Valian, V. (1986). Syntactic categories in the speech of young children. *Developmental Psychology*, *22*(4), 562.

Valian, V. (2014). Arguing about innateness. *Journal of Child Language, 41 (S1)*, 78-92.

Valian, V., Solt, S., & Stewart, J. (2009). Abstract categories or limited-scope formulae? The case of children's determiners. *Journal of Child Language, 36*(4), 743-778.

van Heugten, M., & Johnson, E. K. (2010). Linking infants’ distributional learning abilities to natural language acquisition. Journal of Memory and Language, 63(2), 197-209. doi:10.1016/j.jml.2010.04.001.

Wang, F. H., Höhle, B., Ketrex, F. N., Küntay, A. C., & Mintz, T. H. (2011). Cross-linguistic distributional analyses with frequent frames: the cases of german and turkish. In N. Danis, K. Mesh, & H. Sung (Eds.), Proceedings of the 35th Annual Boston University Conference on Language Development (BUCLD 35) (Vol. 2, pp. 628-640). Boston, MA: Cascadilla Press.

Yang, C. (2010). *Who’s afraid of George Kingsley Zipf?*Unpublished manuscript.#

Yang, Charles. 2008. The great number crunch. *Journal of Linguistics, 44*, 205–28

Zhang, Z., Shi, R., & Li, A. (2014). Grammatical Categorization in Mandarin-Chinese-Learning Infants. Language Acquisition, 22(1), 104-115. doi:10.1080/10489223.2014.943900.