Systematicity and Arbitrariness in Language: Saussurean Rhapsody

Charles Yang*
University of Pennsylvania
charles.yang@ling.upenn.edu

July 2020

> In fact, the whole system of language is based on the irrational principle of the arbitrariness of the sign, which would lead to the worst sort of complication if applied without restriction. But the mind contrives to introduce a principle of order and regularity into certain parts of the mass of signs, and this is the role of relative motivation. ... Since the mechanism is but a partial correction of a system that is by nature chaotic, however, we adopt the viewpoint imposed by the very nature of language and study it as it limits arbitrariness.
>
> Ferdinand de Saussure
>
> *Cours de linguistique générale* (CLG; 1916, 133)

1 Beyond Arbitrariness

The original title of this chapter was *Productivity and the Lexicon* but that would probably strike one as an oxymoron. The lexicon, at least in the tradition widely recognized as Saussurean, is a depository of arbitrariness, the very opposite of productivity. As Di Sciullo and Williams (1987, 3) vividly put it,

*I thank Bob Berwick, Lila Gleitman, Annika Heuser, Barbara Lust, John Trueswell, Hector Vazquez, Virginia Valian, and two anonymous reviewers for helpful discussions and comments.

the lexicon is "like a prison — it contains only the lawless, and the only thing that its inmates have in common its lawlessness."

Strictly speaking, of course, the lexicon is not always arbitrary. As Saussure duly noted, along with some of the most prominent contributors to linguistics from Jespersen (1922) to Bloomfield (1933), and from Harris (1951) to Chomsky (1957), sound symbolism can be observed to varying degrees in all languages. For example, English words with the initial consonant cluster /gl/ strongly convey the meaning of light, vision, and brightness – *glitter*, *gleam*, *glory*, etc. – which was readily confirmed when I met a co-editor of the present volume over twenty years ago.[1] But the more serious problem with the principle of arbitrariness, especially when over-emphasized, is that it lends to an incomplete reading of Saussure. As he made abundantly clear throughout CLG, the principle of arbitrariness, an observation that borders on banality, is important not so much as a fundamental property of language, but as a fundamental property of language *for language to overcome*. On my (whig-history) reading, Saussure advocated a radical, and deeply psychological, perspective on words, language, and cognition.

> Psychologically our thought – apart from its expression in words – is only a shapeless and indistinct mass. Philosophers and linguists have always agreed in recognizing that without the help of signs we would be unable to make a clear-cut, consistent distinction between two ideas. Without language, thought is a vague, unchartered nebula. (CLG, 111-112)

A principle of order and regularity, as quoted in the epigram, serves to limit such arbitrariness. As a point of example, Saussure contrasted the French numerals *vingt* '20' and *dix-neuf* '19' (CLG, 131). The former is purely idiosyncratic and the sound-meaning pair is arbitrary. By contrast, the latter clearly has a compositional structure as both *dix* and *neuf* recur into other numerals such as *dix-huit* '18' and *trente-neuf* '39'. The "relative motivation" between *dix-neuf* and *dix-huit*/*trente-neuf* is referred to by Saussure as "associative". It is nevertheless clear, especially from many other examples in Saussure's

[1]There are 19 such words (more precisely, stems) with non-negligible frequencies: *glad*, *glass*, *<u>glove</u>*, *glory*, *<u>glue</u>*, *glow*, *<u>globe</u>*, *glance*, *glimpse*, *glamour*, *glitter*, *<u>glitch</u>*, *glare*, *gloat*, *<u>gland</u>*, *gloss*, *gloom*, *glee*, and *glum*, and I count 6, underlined, as not supporting the noted semantic field. For what it's worth, 13 items out of 19 just about support a generalization according to the Tolerance Principle; see Section 3.2. Of course, such a sound-meaning mapping, if genuine, can at best guide the English speaker into the right semantic ballpark when encountering a novel word *gleit*: the residual, and arbitrary details still have to be filled out.

discussion that primarily draws from English, French, and Latin morphology (e.g., *enseignement-enseinger-enseignons*, *painful-delightful-frightful*), that such associative relations refer to higher-order rules and abstractions over words. "elements of sound and thought combine", language makes the infinite use of finite means, the Humboldtian notion brought to modern prominence through generative grammar and Chomsky.

This chapter is the discussion of productivity on the Saussurean theme: How the formal systematicity enables language to go beyond and overcome the arbitrariness of the lexicon. A special focus will be placed on learning: to go beyond arbitrariness, children must discover, and subsequently generalize, the language-specific regularities that reside in their finite linguistic experience. Like Saussure, our discussion will mostly focus on morphology or word formation; see also Demuth (this volume). However, I will also suggest that an appropriate theory of productivity should be applicable to form-meaning mappings at any linguistic level, including sound symbolism alluded earlier (fn 1) as well as syntax-semantics correspondences, with potentially interesting implications for how language acquisition shapes children's conceptual development.

In Section 2, I review some widely used behavioral tests of productivity, along with the associated methodological complications that may give rise to conflicting findings. In Section 3, I provide a cross-linguistic survey of productivity in children's morphology. The evidence is unequivocal: productivity is a categorical phenomenon, as traditionally held, despite an assortment of pleas for gradience in the recent literature. In Section 4, I discuss how productivity may be detected by the child learner in a psychologically plausible setting. The developmental evidence suggests that an appropriate learning model must be similarly categorical in nature. In Section 5, I offer some initial thoughts on how a learning-theoretic approach to productivity impacts the theory of language and cognition. What can be effectively learned from data needn't be built into Universal Grammar, and a learning model that gives rise to qualitative changes in the child's grammar may also help understand how language may play a causal role in children's conceptual development.

In what follows, I will assume that the child has acquired a reasonable vocabulary so that a formal system of productivity can be established; see Gleitman and Trueswell (this volume). These two processes of acquisition are logically independent but are clearly intertwined. For example, Brown's seminal study

(1957) shows that the form in which a word is used – *to sib*, *a sib*, *sibbing* – can guide children to focus on special aspects of word meanings. Similarly, according to the theory of syntactic bootstrapping (Gleitman 1990; Lidz, this volume), children exploit syntactic regularities such as word order and case marking, which must be derived from their existing vocabulary, to deduce the semantic properties of novel words. Even an incomplete grasp of productivity can contribute to the acquisition of vocabulary. For instance, if children have learned that determiners and nouns productively combine, they may infer that a novel word following a determiner is likely a noun (Shi and Melançon 2010; see Dye et al. 2019 for a recent review). The acquisition of the initial vocabulary appears slow, which in part has to do the very high degree of referential ambiguity in the linguistic input (Chomsky, 1959, Quine, 1960, Landau and Gleitman, 1985, Gillette et al., 1999, Trueswell et al., 2016) and the considerable computational challenges it poses (Siskind, 1996, Yu and Smith, 2007, Frank et al., 2009, Fazly et al., 2010, Stevens et al., 2017). The rapid rise in children's vocabulary during later stages is likely a consequence of having established a formal system of productivity (e.g., Bates et al., 1988, Fenson et al., 1994, Anisfeld et al., 1998, Hoff and Naigles, 2002, Gleitman et al., 2005).

2 Tests for Productivity

The simplest assessment of productivity is the celebrated Wug test as it directly measures the linguistic capacity to generalize. In a landmark study, Berko (1958) introduced young children (age range 4 to 7) to novel words including nouns, verbs, and other categories and recorded their responses.

(1) This is a WUG.

Here is another one.

These are two _____.

WUGS.

It is important to note that children are far from perfect in their Wug performance (Berko, 1958, 160). For instance, in the regular past inflection of novel verbs, first graders successfully produced *rick-ricked* and *spow-spowed* in as few as 25% of the test cases. But we must remember that most if not all English-learning children acquire the *-d* rule by three (Marcus et al., 1992), as indicated by their spontaneous use of overregularized forms (e.g., *fall-falled*). Thus there seems to be a gap of at least three years between acquiring the regular past tense and using it in a specific experimental design (e.g., Huttenlocher 1964, Anisfeld and Tucker 1967): failure to pass the Wug test does not imply imperfect knowledge of morphology.

The picture gets murkier for words that do not follow productive rules, with the English irregular verbs as the paradigm example. Many irregular verbs are residue of historically productive processes; as such, some may exhibit a certain degree of similarity. For instance, some irregular verbs ending in /ɪŋ/ form past tense by changing the vowel to /æ/ (e.g., *sing-sang*, *ring-rang*) or /ʌ/ (e.g., *swing-swung*, *cling-clung*). These similarities are at best partial: for instance, *wing*, *ping*, *ding*, etc. actually take *-d*, and there are irregular verbs that have nothing in common yet undergo the same process to form past tense: *bring*, *buy*, *catch*, *fight*, *teach*, *think*, and *seek* all replace the rime with /ɔt/.[2] Nevertheless, these similarities are the primary motivation for analogical account of morphology and associative accounts of morphological acquisition (Rumelhart and McClelland, 1986, Pinker and Prince, 1988).

Children, however, are oblivious to the pull of analogy. The pattern is already clear in the original Wug study although it has been overshadowed by the better-known result on productive rules. Berko found that when children failed to produce *spowed* for *spow*, they produced no response at all rather than, say, *spew*, which would follow the analogy of *know-knew*, *grow-grew*, *blow-blew*, etc. In addition, Berko systematically investigated the role of analogy with very irregular-like stimuli. Children were presented with novel verbs such as *bing* and *gling* that are strongly similar to existing irregular verbs (*sing-sang*, *sting-stung*, etc.), and thus with great potential for analogical extension.

(2) This is a man who knows how to GLING.

He's GLINGING. (Picture of a man exercising.)

He did the same thing yesterday.

What do he do yesterday?

[2] The English irregular past tense formation has been referred to as *product-oriented schema* (e.g., Bybee and Slobin, 1982) as the verbs are characterized by undergoing the same structural change: which verb can undergo the change must be lexically learned (Yang, 2002). This is to be contrasted with *source-oriented schema* where the structural change automatically applies if some structural condition is met, e.g., *-d* applies to a word if it is a verb. We will retain the more traditional non-productive vs. productive distinction in the current discussion.

Yesterday he _____.

Children were overwhelmingly conservative in their responses. Only one of the eighty-six children in Berko's study supplied the analogical form *bang* and another produced *glang*. Ivimey (1975) replicated Berko's results with the same stimuli: virtually no children analogized the real or the pseudo irregulars. (Graves and Koziol 1971 tested children's response on real irregular nouns such as *foot* and *mouse* – which were indeed over-reguarlized – but did not report results on *goot* and *touse*.) These findings are strongly consistent with children's production data from many cross-linguistic acquisition studies (Section 3.2).

The paucity of analogical forms and the abundance of rule-based generalizations stand in contrast with claims about the probabilistic and gradient nature of productivity and grammar (Keller 2000, Hay and Baayen 2005, Bresnan and Nikitina 2009). In the words of McClelland and Bybee (2007, 439), "there is no dichotomous distinction between productive and unproductive phenomena; rather, there are only degrees of productivity".

The appearance of gradience, however, seems methodological in nature. This is already evident in Berko's original study. While almost no child produced forms such as *bang* and *glang*, adults "clearly felt the pull of the irregular pattern, and 50% of them said *bang* or *bung* for the past tense of *bing*, while 75% made *gling* into *glang* or *glung* in the past" (Berko, 1958, 165). Similarly, Ivimey (1975) found that adults' tendency to use the irregular forms increased as a function of age and especially the level of education. Indeed, most case studies that support to the claim of gradience were carried out with adult subjects (e.g., Bybee and Moder, 1983, Marcus et al., 1995, Clahsen, 1999, Hahn and Nakisa, 2000, Hayes et al., 2009, Becker et al., 2011). Some, however, have design flaws. For instance, Bybee and Moder (1983) and Hahn and Nakisa (2000) presented the subjects with questionnaires consisting of real irregular and pseudo-irregular words; the facilitatory effect is left uncontrolled for. Furthermore, while children clearly regarded the nonce words as if they were being taught new English words (Berko, 1958, 157), it is not clear how adult subjects approach the Wug test, as they may try to uncover the purpose of the task as in many experimental settings (Anderson, 1980). As Schütze (2005) notes, the instructions provided (to adults) generally do not rule out what he calls the dictionary scenario, in which the subject approaches the test item as an obscure word in their native language and thus attempts to offer a reasonable guess of what the inflected form would be. This may have the effect of leading the subject to deviate from their "normal" linguistic behavior. It turns out that subjects frequently reported the Wug task as some kind of joke that required bizarre analogies (Derwing and Baker, 1977).

Compounding the matter further is the use of rating tasks in the assessment of productivity (Kim et al., 1991, Prasada and Pinker, 1993, Ramscar, 2002, Albright and Hayes, 2003, Ambridge, 2010). But these tasks, which are gradient in nature, would almost guarantee a gradient conception of productivity. In general, participants are inclined to spread responses over whatever range they are given (Parducci and Perrett 1971). For instance, the classic study of Armstrong et al. (1983) finds that gradient judgment is obtained even for uncontroversially categorical concepts such as "even number". Still, virtually all studies have found that regularized forms are rated higher than irregularized forms in head-to-head comparisons. For example, in one of the few rating studies with children in the original Wug test age group (Ambridge, 2010, Appendix), only one (*drit*) out of 40 verbs tested had the regular form *dritted* rated slightly lower than the irregular form *drit*.

This somewhat critical review is only meant to suggest that the tests for productivity, like all experimental methods, must be properly interpreted and may not provide a direct window into language and productivity.[3] There may be several factors contributing to the mixed results reviewed so far. First, as suggested earlier, it is possible that adults and children approach experimental tasks differently. For instance, in artificial language studies, children generally apply categorical rules while adults tend to match the statistical distribution of variable forms (e.g., Hudson Kam and Newport, 2005, Schuler et al., 2016). Similar differences have been observed between children and adults in behavioral tasks in other domains of cognition (Stevenson and Weir, 1959, Weir, 1964, Derks and Paclisanu, 1967). Whatever the nature of these differences are, it is important to note that adults in naturalistic settings *are* capable of learning categorical rules as in their second language however arduously (White and Genesee, 1996). Second, the previous conflicting findings may be due to the failure to distinguish productive from unproductive processes. In the former,

[3]Similarly, I have primarily focused on children's production data as evidence for productivity. Other tests are possible but all come with caveats. The most obvious alternative is comprehension. For instance, (Figueroa and Gerken, 2019) find with a preferential listening task that sixteen-month-olds can distinguish nonce words from high-frequency and likely familiar nouns when both are suffixed with *-d*, e.g., *fimmed* and *trucked*. Thus, children at this age know that verbs can appear with *-d* but nouns cannot. But this falls short of productivity, which requires that verbs *must* appear with *-d*, sometimes with the irregulars as collateral damage.

such as the English past tense *-d*, rules always apply. In the latter, such as the English irregular past tense, the absence of a productive rule forces (adult) subjects to resort to analogical similarity under experimental conditions — only up to a point, as the discussion of morphological gaps in Section 3.3 will make clear. Such an interpretation of the behavioral results is similar to the findings in the study of categorization – linguistic or otherwise – that abstract categories may co-exist with prototypical exemplars (e.g., Armstrong et al. 1983, Pierrehumbert 2001, Murphy 2004). Again, clear evidence comes from actual languages. For example, despite adult subjects' willingness to supply and accept irregularized forms, the English language has not added a new irregular verb in a very long time: see Anderwald (2009) for comprehensive discussion of the historical record. All changes in the English verb system in the past few centuries have been familiar pattern of irregular verbs drifting to the regular class (Anderwald, 2013, Ringe and Yang, 2020). Indeed, the English language of the recent past has presented two clear opportunities for the extension of irregular forms. The verb *bing* (from the Microsoft search engine) and *bling* (to display ostentatious jewelry) both take *-d* and are fully regular, despite fitting the most favorable condition for irregularization (Bybee and Moder, 1983, Albright and Hayes, 2003).

3 Productivity in Child Language

In this section we provide a review of the quantitative findings of children's morphology across languages. A categorical notion of productivity is strongly supported. We once again begin with the well-known case of English past tense.

3.1 Emergence

An idealized depiction of English past tense acquisition is given in Fig 18.1 based on the longitudinal data of "Adam", a child from Brown's 1973 study and adapted from Pinker (1995, 116); see also Maslen et al. (2004).

Adam's irregular verb acquisition follows the well-known U-shaped learning curve (Marcus et al., 1992), a developmental pattern first discussed by Ervin and Miller (1963), Brown (1973), and MacWhinney

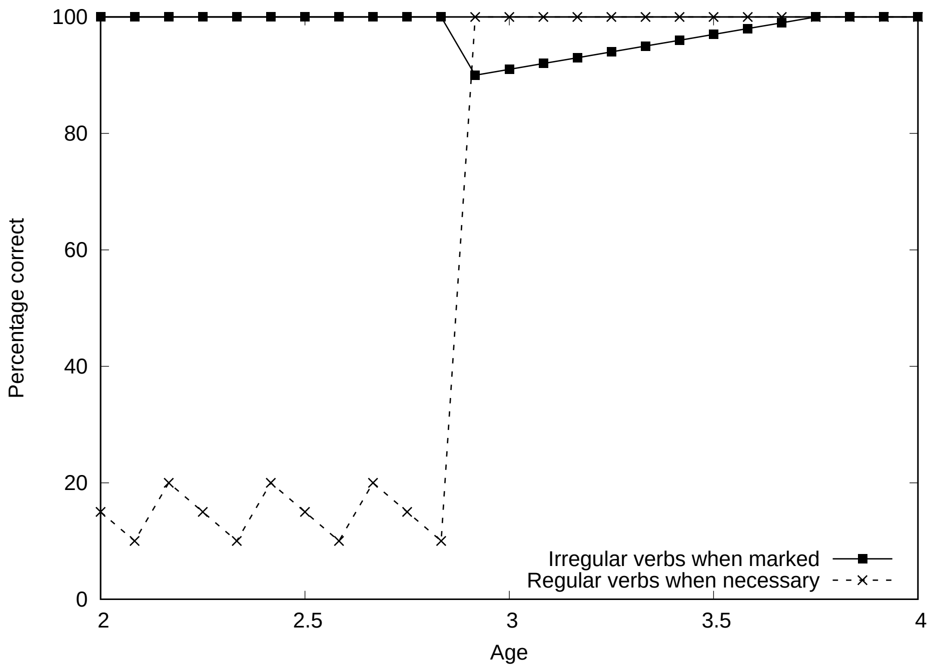


Figure 18.1 The developmental trajectory of irregular and regular past tense marking. Irregular verbs, when marked, are initially perfect before children start to over-regularize them, which indicates the emergence of *-d* productivity. The regular verbs are initially very inconsistently marked with *-d* in obligatory past tense contexts and become much more consistent at the same time when *-d* becomes productive.

(1975). Prior to 2;11, every irregular verb was used correctly before Adam succumbed to over-regularization and produced the first irregular marking error: *What dat feeled like*? Because *feeled* is inconsistent with adult input, it must have been the child's spontaneous creation. The onset of over-regularization, then, is taken to be the emergence of productivity: the suffix *-d* in this specific case. Over-regularization errors persist from this point onward and will gradually diminish, due to the cumulative exposure to the irregular forms in the input that will override the productive rule over time. The precise nature of how the irregular verbs are learned and stored is a matter of considerable debate (Rumelhart and McClelland, 1986, Pinker and Prince, 1988, Clahsen, 1999, Yang, 2002) that will not concern us here.

The development of the regular verb past tense marking is also of significant interest although this is often not considered in the same context. An important early observation is due to Ervin (1964): for some *regular* verbs, children do not mark them at all (and instead use the stem form) in obligatory past tense context until after the appearance of over-regularization of *irregular* verbs. Similarly, Marcus et al. (1992, Chapter VI)'s longitudinal study shows that before the onset of productivity, children marked both regular and irregular verbs inconsistently in obligatory contexts, producing numerous examples such as *I see bird just now* and *I walk home yesterday*. Past tense marking rose significantly across the board, for regular and irregular verbs alike, after the -*d* suffix became productive (with over-regularization errors). As Kuczaj (1977, 593) remarks, "Apparently, once the child has gained stable control of the regular past tense rule, he will not allow a generic verb form to express 'pastness,' which eliminates errors such as *go*, *eat*, and *find*, but results in errors like *goed*, *eated*, and *finded*, as well as *wented*, *ated*, and *founded*." A reasonable interpretation of the developmental change is that in the early stage, the child's regular verbs are lexically specific (i.e. Tomasello 2003) and do not go beyond the scope of the input. Once they discover that -*d* is formally productive, they realize that tense marking is an obligatory feature of their language and use to do so across the board for all verbs. In other words, the child's *grammar* becomes infinite and systematic after they detect formal productivity that resides in their finite, and necessarily arbitrary, *vocabulary*. Such an inductive step is obviously critical for the acquisition of language; it may also have important implications for children's conceptual development and change that interacts with language as we discuss in Section 5.2; see Landau (this volume) for general discussion.

The emergence of productivity is a critical milestone in child language, and has rightly become a major focus of research from many theoretical perspectives. These transient moments are no doubt real. There *is* a moment when the child overgeneralizes a rule for the very first time. And for the acquisition of language specific rules, all learning theories require some accumulation of examples: the child cannot conclude -*d* is productive for all verbs upon hearing only one instance. But these moments are nevertheless difficult to capture in "real time" as the record of individual children's longitudinal development is never complete. There is also a good deal of individual variation in the development of productivity noted from the very earliest studies of morphological acquisition (Cazden, 1968, De Villiers and De Villiers, 1973, Brown, 1973). For instance, while Adam produced the first instance of over-regularization at 2;11, Eve's first overuse of -*d* came at 1;10 (*it falled in the briefcase*). Brian, a child in the longitudinal study of (Maslen et al., 2004), over-regularized at 2;5. Abe, another child with extensive longitudinal record (Kuczaj, 1977), overregularized during the first recording session at 2;3 (*he falled*) and his regular past tense marking in obligatory context is already very high (Marcus et al., 1992). A theory of productivity acquisition must leave room for such extensive individual differences even though, in general, all children eventually converge on the same set of rules; we discuss these matters further in Section 4.1.

3.2 Irregulars and Regulars Across Languages

Once a productive rule emerges, children acquire a systematic aspect of their language while occasionally over-extending its application. The rates and duration of such errors vary from case to case, and from child to child. Earlier corpus studies suggest that 4.2% of the English irregular verbs children produce are over-regularized (Marcus et al., 1992, Pinker, 1995), but later studies have found somewhat higher rates. For instance, Yang (2002) reports a rate of 10% based on approximately 7,000 tokens from four large longitudinal corpora in the public domain (see also Maratsos 2000), and Maslen et al. (2004) find that 7.8% out of about 1,300 tokens produced from a single child are overregularized. The robustness of over-regularization is uncontroversial but the literature has often given an impression that the irregular forms are also frequently extended. As discussed in Section 2, this may be the result of task effects in various behavioral tests for productivity. The evidence from children's naturalistic production, by contrast, unambiguously points to a categorical distinction between regular and irregular processes.

The acquisition of English again provides most detailed evidence. The frequent reference to analogical errors such as *bite-bote*, *wipe-wope*, *think-thunk*, etc. (Bowerman 1982, Bybee 1985, Pinker and Prince 1988, Pinker 1999, Ambridge et al. 2015) seems anecdotal: not a single instance can be found in the over 4 million words produced by English-learning children in the CHILDES database (MacWhinney, 2000). The most comprehensive empirical study of analogical errors (Xu and Pinker, 1995) dub these "weird past tense errors" on the basis of their rarity. Xu and Pinker examined over 20,000 past-tense tokens produced by nine children, and only forty weird errors (0.02%) were identified. Similarly, an elicitation study of typically-developing and SLI children also finds that when prompted for past tense of irregular verbs, children frequently over-regularize or leave the verb unmarked, but the stem changes characteristic of irregular verbs are very rare for both groups (Marchman et al., 1999). The drastically different rates of over-regularization and overirregularization suggest that there is a (near) categorical distinction with respect to productivity between productive rules and irregular rules. When children make mistakes, they almost always employ a default or productive form (e.g., *thinked*) or omit the appropriate form altogether: they almost never substitute with an inappropriate form.

The categorical distinction between productive and unproductive morphological processes is strongly confirmed across languages. For example, in a study that targets the German agreement affixes -*st* (2SG) and -*t* (3SG), Clahsen and Penke (1992) find that while children supply an agreement affix in obligatory context only 83% of the time, almost all the errors are those of omission. When the child does produce an agreement affix, it is almost always the appropriate one (over 98% of the time); inappropriate use (e.g., substituting a -*t* for -*st*) is virtually absent. Similar patterns can be observed in the acquisition of Italian. In a cross-sectional study (Caprin and Guasti, 2009, 31), children in all age groups use a diverse and consistent range of tensed forms. Furthermore, the use of person and number agreement is essentially error free throughout, reaching an overall correct percentage of 97.5%, consistent with previous reports (Guasti, 1993, Pizzuto and Caselli, 1994). Children's impressive command of agreement is most clearly seen in the acquisition of languages with considerable morphological complexities. In a study of morphosyntactic acquisition in Xhosa (Gxilishe et al., 2007), children are found to gradually expand the use of subject agreement across both verbs and noun classes. The rate of marking in obligatory contexts as well as the diversity of the morphological contexts themselves steadily increased. In a process best described as probabilistic, the children often alternate between marking a verb root in one instance and leaving it bare in another, very much like the use/omission alternation pattern reviewed earlier. Crucially, virtually all agreement errors are those of omission: 139 out of 143 or 97.2% to be precise. Substitution errors are again very rare, confirming previous research on languages with similarly complex morphology (Demuth, 2003, Deen, 2005), including polysynthetic languages such as Inuktitut (Allen, 1996).

We turn now to two case studies that focus more specifically on the contrast between regular and irregular morphologies in children's naturalistic speech. This type of evidence has been accumulating in the literature on the dual-route approach to morphology (Pinker, 1999, Clahsen, 1999), for which a categorical distinction between regular and irregular processes is of central importance. The results are again unambiguous.

The German participle system consists of a productive default -*t* suffix (*fragen-gefragt* 'ask-asked'), as well as an unpredictable set of irregulars taking -*n* (*stehlen-gestohlen* 'steal-stolen') (Wiese, 1996). In a series of studies, Clahsen and colleagues (Clahsen and Rothweiler, 1993, Weyerts and Clahsen, 1994, Clahsen, 1999) find that children across all age groups overapply the -*t* suffix to the irregulars, where the reverse usage is virtually absent. Their longitudinal data contains 116 incorrect participle endings, out of which 108 are -*t* errors (**gekommt* instead of *gekommen* 'come', i.e., over-regularization). The rest are irregularization errors such as **geschneien* for *geschneit* (snowed). According to the authors, the overall rate of -*t* regularization is 10% of all usage, which suggests that the -*n* irregularization rate is merely 0.75% (based on the 8 -*n* errors compared to 108 -*t* errors). The acquisition of German past participles, therefore, is quite analogous to that of English past tense reviewed earlier (Xu and Pinker, 1995), because both point to the productive asymmetry between regulars and irregulars.

The inflection of Spanish verbs provides a complex but highly informative case for exploring productivity in child language. In Spanish, stems generally consist of theme vowels and roots, which are then combined with affixes for inflection. For instance, a finite form of the verb *hablar* (to talk) is *habl-a-ba-ais*, which represents the root (*habl* 'speak'), the theme vowel (*a*), the past tense (*ba*) and the second-person plural (*ais*). The irregularity in Spanish inflection comes in two broad classes concerning the stem and the suffix respectively. There are some thirty verbs that are highly irregular with the insertion of a velar stop in certain inflections. The majority of irregulars undergo an alternation known as diphthongization, a process which is not limited to verbal morphology (Harris, 1969, Eddington, 1996). However, which verbs undergo diphthongization still must learned on an individual basis: It is possible to find minimal pairs such as *contar* ('to count') and *montar* ('to mount') where the former contains the diphthong (*cuento*) but the latter does not (*monto*). And there are a few common verbs that show both diphthongization and velar insertion in some forms. Although inflectional irregularities in Spanish mostly concern the stem, the suffixes are affected as well. For the stem *querer* 'to want', for instance, the 1SG past tense is *quise*, which involves the stem change noted earlier but also takes an irregular suffix rather than the regular suffix, which would have resulted in *quisí*. The suffix in the 3SG past tense *puso* 'she/he/it put' is -*o* and the regular suffix would have formed **pusió*.

Clahsen et al. (2002) analyzed the verbal inflections of the fifteen Spanish-speaking children and found strong evidence for a categorical distinction between the regular and irregular inflections.

(3) a. The irregulars: children produced a total of 3,614 irregular verb tokens, out of which 168 (4.6%) are incorrect either in stem formation or suffixation.

i. Of the 120 stem-formation errors (see below), 116 are over-regularizations and only one is analogical irregularization.

ii. Of the 133 suffixation errors, 132 are over-regularizations with no occurrence of irregularization.

b. The regulars: children produced 2,073 regular verb tokens, only 2 of which are the inappropriate use of irregular suffixes.

Collectively, then, the rate of analogical irregularization is only 0.001% for all verbs, and also 0.001% for the irregulars: again, orders of magnitude lower than the rate of over-regularization errors. Clahsen et al.'s study does not include errors regarding diphthongs; all the stem-formation errors are failures to use a diphthong when required; it thus does not rule out the possibility of "mis"-diphthongization — for example, the child produces [ie] alternation when the correct diphthong is [ue]. To address this issue, Mayol (2007) provides a finer-grained investigation of inflectional errors focusing more specifically on the types of stem-formation errors and their underlying causes. The speech transcripts of six Spanish-learning children, almost 2,000 tokens in all, fail to yield a single misuse of diphthongization.

3.3 When Productivity Fails

The Wug test puts the unbounded creativity of language in the spotlight but I would be remiss if we left out the corners of grammar where productivity breaks downs. In a classic paper, Halle (1973) draws attention to morphological "gaps," the absence of inflected words for no apparent reason. For instance, there are about seventy verbs in Russian that lack an acceptable first-person singular nonpast form (data from Halle 1973 and Sims 2006).

(4) *lažu 'I climb'

*pobežu (or *pobeždu) 'I conquer'

*deržu 'I talk rudely'

*muču 'I stir up'

*erunžu 'I behave foolishly'

There is nothing in the phonology or semantics of these words that could plausibly account for their illicit status, yet native speakers regard them as ill-formed. Such unexpected failure of productivity is in fact widely attested in the world's languages (see Baerman et al. 2010 and Fanselow and Féry 2002 for surveys). Even the relatively impoverished morphology of English contains gaps. For example, most speakers are not sure about the past tense form of the verb *forgo* (*forgoed* or *forwent*) or the past participle of the verb *stride* (*strided* or *stridden*) (Pullum and Wilson, 1977). Similarly, in most dialects of English, the negative clitic *-n't* cannot be contracted to the auxiliary/modal verbs *am* (**I amn't*) and *may* (**you mayn't*) in the manner of *haven't*, *needn't*, and *don't* (Zwicky and Pullum, 1983). In other words, children learning these English dialects must *fail* to learn a productive rule for these words in order to acquire the dialects correctly. Finally, as Halle (1973) notes, gaps show that there is no inherent productivity distinction between inflectional and derivational morphology: that the former tends to be productive and the latter tends not to must be captured by the language learning rather than some architectural property of the grammar.

The failure of morphological productivity has only received scant attention in the acquisition literature. In an important study, Dąbrowska (2001) shows that in Polish, masculine nouns in the genitive singular either take an *-a* or *-u* suffix. A corpus analysis of child-directed Polish shows that *-a* is the numerical majority, covering over 65% of the nouns, but fails to become the productive default; see Section 4.3 for details. Loanwords, for example, take on *-a* and *-u* in unpredictable fashion. And native speakers have to resort to brute-force memorization of noun-specific suffix, a process that extends well into teenager years (Dąbrowska, 2005), as the choice of the suffix seems arbitrary (Mausch, 2003, Westfal, 1956). The Polish genitive system, and the phenomenon of gaps more generally, pose considerable challenges to many leading theories of language and language learning. It suggests that child learners should not presuppose the existence of a default rule — and look for it — as presumed by the dual-route model of morphology (Clahsen, 1999, Pinker, 1999). Likewise, the absence of a default also poses challenges to competition-based theoretical frameworks such as Distributed Morphology (Halle and Marantz 1993) and Optimality Theory (Prince and Smolensky 2004) in which a winning form is always expected to emerge.

In summary, we have seen cross-linguistic evidence that when learning morphology, children draw a categorical distinction between productive and unproductive processes. These results conform with the traditional notion of productivity (Bloch, 1947, Nida, 1949, Botha, 1969, Bauer, 1983), and are at odds with the recent claim of gradience along a continuum of productivity (Jackendoff, 1996, McClelland and Bybee, 2007). That is not to say that productivity holds uniformly across individual speakers in all cases: a rule may be categorically productive for some speakers but categorical unproductive for others. We have already seen examples of this: at the age of 2;6, for instance, *-d* was categorically unproductive for Adam but categorically productive for Eve. But if the individuals' productivity measures are pooled (as is frequently the case in experimental research), or the productivity test is inherently gradient (Section 2), it may give rise to the impression that productivity lies on a continuum. The categorical distinction between productive and unproductive processes, extensively supported across languages, should not be surprising: similarity-based analogies, which children steadfastly avoid, would not give rise to a stable and usable language shared by a community of speakers.

Returning to our Saussurean theme, the emergence of productivity allows the child to form generalizations beyond their finite, and to a great extent, arbitrary experience. Furthermore, productivity must overcome another layer of arbitariness: the irregular verbs, which derive from historical change but form essentially an arbitrary list, must be overcome for children to derive the *-d* rule. How, then, do children pick out just the productive rules in their language? Fortunately, the acquisition evidence reviewed this section places severe constraints on the space of possible models.

4 The Learnability of Productivity

4.1 Developmental constraints on learnability

The preceding section can be viewed as a descriptive study of productivity from a developmental perspective: the child knows A at age X but B at age X+Y. The learnability of productivity aims to provide a more complete explanation (Chomsky, 1965): What kind of learning mechanisms, acting on what kind of linguistic data, can facilitate the transition from A to B during the time span of Y?

A description of productivity thus provides a set of design specifications for a learning-theoretic account, for which the psychological nature places additional constraints. For example, a descriptive account of productivity can be pursued with whatever methodological tools at the scientist's disposal such as those reviewed in Section 2 and 3: corpus statistics, behavioral studies, observations of language change over time, etc. But children learning languages are much more resource-limited. They have no access to information such as "This is an irregular verb" or "That's a productive rule", or indeed behavioral or statistical correlates of productivity (e.g., that the reaction time of processing productive generally does not show whole-word frequency effects; Ullman et al. 1997, Clahsen 1999). Rather, they are presented with a finite set of words that undergo an assortment of morphological processes: they must be able to determine which ones can, and cannot, generalize beyond the input. Finally, a learning theory must has plasticity to leave room for individual variation in language development, as reviewed by Potter and Lew-Williams (this volume), while ensuring the final outcome to be generally uniform across individuals. To illustrate this point, consider the first verbs produced by the six children in the Providence Corpus (MacWhinney, 2000). The biweekly recording sessions started at 1;0 and therefore provide a reasonable approximation of the children's vocabulary growth. Using the morphological annotation in the corpus, I extracted the first 150 unique verb stems produced by these children in longitudinal order, divided into six sets with a size increment of 25 verbs, and the Jaccard similarities are computed for children's verb vocabulary.[4] Fig 18.2 reports the mean value and standard deviation across the 15 pairwise comparisons for the six stages.

[4]For each pair of children, the Jaccard similarity is the ratio between the size of their vocabulary intersection and the size of their vocabulary union.

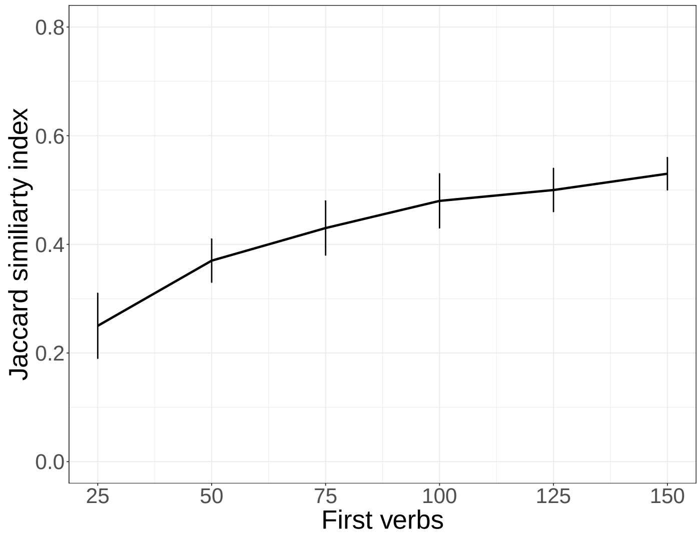


Figure 18.2 Children's early verb vocabularies differ considerably across individuals. Data from the Providence corpus.

The individual differences across children are significant, especially during the early stages. We may never be able to predict exactly which word a child learns, but a theory of productivity must be able to "normalize" such differences to ensure an essentially uniform outcome of language learning, which has been abundantly documented in the quantitative study of language variation and change within speech communities (Labov, 1972, Labov et al., 2006).

There has been no shortage of learning models of productivity. The so-called past tense debate, which has engendered much of the research on morphological productivity and its acquisition, was ignited by Rumelhart and McClelland's connectionist network to model English past tense (1986). At a high level, the network model was able to reproduce the U-shaped learning curve in child language development, i.e., an initial stage of conservatism followed by the emergence of productivity which results in over-regularization. However, the Rumelhart and McClelland model makes unpredictable errors that are unattested in child learning when trying to handle regular forms, such as *membled* for the past tense of *mail* (Pinker and Prince, 1988). While the specifics of these problems have been addressed with further improvements to this class of models (Plunkett and Juola, 1999), the core tension between productive and unproductive morphological processes remains.

The categoricity of productivity in child language acquisition documented in the previous sections has almost completely escaped the attention of learning research. While every effort to model morphological learning has (rightly) focused on the over-regularization phenomenon, the absence of over-irregularization in child language has never been taken into account, save a brief comment in Marcus (1995, 277) that connectionist models produce similar rates for both. Indeed, the failure to distinguish productive and unproductive patterns has been a persistent feature of computational models of morphological learning. For example, a recent computational study by O'Donnell (2015) tests several (nonconnectionist) models of past-tense learning. Most models are reasonably successful at passing the Wug test for regular verbs, but they all produce a high number of analogical forms on the basis of the existing irregulars: the best overall model produces 10% of overirregularization patterns on novel items, which is orders of magnitude higher than the overirregularization rate by human children (Xu and Pinker, 1995). The reason for these failures appears to be the inclusion of token frequency of words in the modeling effort. Because irregular verbs in English are quite frequent, they can lead the model to assign/reserve a large probability mass to the irregular morphological processes, in effect matching the frequencies in the input. As a result, irregular forms are analogized when models are presented with nonce words, especially those similar to existing irregulars, in contrast with the virtual absence of analogical errors by children.

Nevertheless, important insights have emerged from connectionist models, which actually converge with ideas from linguistic theorizing. It was long recognized that productivity is the result of high, and indeed, dominant coverage of words. For instance, a classic text states that "(a) form which is statistically predominant is also likely to be productive for new combinations" (Nida, 1949, 45), and the regular suffix *-d* in English is explicitly identified with statistical majority (Bloch, 1947). In the first comprehensive study of morphology in generative grammar, Aronoff (1976, 36) quantifies the productivity of a word-formation rule (WFR) as follows: "we count up the number of words which we feel could occur as the output of a given WFR (which we can do by counting the number of possible bases for the rule), count up the number of actually occurring words formed by that rule, take the ratio of the two and compare this with the same ratio for another WFR." Early psychological research on children's morphology contains similar proposals (Ivimey, 1975, MacWhinney, 1978), and that is also how Rumelhart and McClelland (1986) modeled English past tense. They first presented the network with a small sample of 10 verbs with a dominant majority of 8 irregular verbs, before a sudden influx of a much larger sample of 420 verbs with a dominant majority of 334 regular verbs. The legitimacy of these modeling assumptions aside (see Pinker and Prince 1988 for discussion), the accumulation of a critical mass of regular verbs to overwhelm the irregulars is the driving force for the emergence of productivity (Marchman and Bates, 1994).

A statistical majority (of types) will surely identify *-d* as the productive rule for English past tense: there are over 100 irregular verbs, somewhat fewer in regular circulation, and there are thousands more that take *-d*. But this strategy fails as a general principle for productivity learning. On the one hand, there are morphological gaps such as the Polish masculine genitive reviewed earlier: the suffix *-a* is by far the majority (65%) yet it fails to achieve productivity. On the other, there are cases such as the German noun plural system, where the suffix *-s* is clearly productive – along with at least some of the other suffixes – while not covering only a tiny proportion of nouns in the language (see Section 4.3). The Tolerance Principle is a model that formalizes the traditional insights into a unified solution for productivity.

4.2 The Tolerance Principle

Learning a language requires discovering rules that generalize beyond a finite sample of data. The Tolerance Principle (Yang 2005, 2016b; henceforth TP) is a theory of how such generalizations are formed. Specifically,

(5) Let a rule R be defined over a set of N items. R is productive if and only if e, the number of items not supporting R, does not exceed ϑ_N:

$$e \leq \vartheta_N = \frac{N}{\ln N}$$

If e exceeds ϑ_N, then the learner will learn these by lexically specific means and not generalize beyond them: that is, R is unproductive. Here I use the term "rule" as a theoretical-neutral term to denote a function that maps an input item to an output item. The function may be partial, as it must be for the case of morphological gaps reviewed in Section 3.3.

The TP builds on the intuition in the discussion of learning models earlier, that rules must "earn" productivity by the virtue of being applicable to a sufficiently large number of candidates it is eligible for. If there are 10 examples and all but one (9/10) support a rule, generalization ought to ensue. But no one in their right mind would extend a rule on the basis of 2/10: the learner should just memorize the two supporting examples. Productivity is a calibration of regularities and exceptions — crucially with respective to word *types* rather than tokens. The reader is referred to Yang (2016b) for the empirical motivation and formal analysis behind of the TP.

Table 1 provides some sample values of N and the associate threshold values ϑ_N. Note that ϑ_N decreases quite sharply as a proportion of N, which suggests that rules defined over a smaller vocabulary can tolerate relatively more exceptions, and are thus easier to learn. This has interesting consequences for language development and provides a theoretical underpinning for the idea of "less is more" (Newport, 1990, Elman, 1993, Yang, 2018) which will not be pursued here.

Table 1 The maximum number of exceptions for a productive rule over N items.

N	ϑ_N	%
10	4	40.0
20	7	35.0
50	13	26.0
100	22	22.0
200	38	19.0
500	80	16.0
1,000	145	14.5

In contrast to most if not all learning models in linguistics and psychology (e.g., Shepard 1987, Anderson 1991, Nosofsky et al. 1994, Tenenbaum and Griffiths 2001), the TP is parameter free. Productivity is determined by two input values (i.e., N and e) that are word counts in an individual learner's vocabulary, and a categorical prediction is made without the need for parameter tuning or curve fitting. While the TP is a claim about all language learners, it clearly allows room for variation in the transient stages of language acquisition as well as in the stable grammars of individual speakers. The relationship between N and e, which may change during the course of language acquisition, determines the status of the rule. If e is very low as a proportion of N, then children may rapidly conclude that a rule is productive. Otherwise, a protracted stage of conservatism may ensue, which may be followed by the sudden onset of productivity. It is also possible that no rule ever reaches the productivity threshold; gaps and other phenomena of ineffability arise.

To apply the TP, it is important to obtain reliable estimates of N and e, ideally at the individual level. This is possible in the controlled setting of artificial language learning, as the items children are exposed to are under the complete control of the researcher. For instance, Schuler et al. (2016) taught children the labels for nine objects. In one condition, five of the nouns share a plural suffix and the other four have individually specific suffixes. In another condition, the split was three and six. Children were given a novel label in the singular in a Wug-like test afterwards: generalization was observed in the 5/4 condition but not in the 3/6 generalization, as the TP predicts that 9 items can only tolerate 4 exceptions. In a further variation (Schuler, 2017), children were given a vocabulary test to assess their N and e, which subsequently confirmed personalized predictions of the Tolerance Principle. The TP has also been shown to be effective for even younger subjects on a passive language task (Koulaguina and Shi, 2019) where again the vocabulary size and composition can be carefully controlled.

In general, however, vocabulary estimation of language learners is difficult. The challenge is even more formidable when a child-directed speech corpus is not available. Nevertheless, there are several mitigating factors such that the data poverty problem is not debilitating. On the one hand, child language, at least the core aspects of the grammar such as morphology and word order, is acquired very early, at an age where the vocabulary size is at most just over 1,000 items (Hart and Risley, 1995, Szagun et al., 2006). On the other, there is converging evidence that lexical frequency can help to provide a reasonable approximation of vocabulary. Nagy and Anderson (1984), for instance, estimate that most English speakers know words above a certain frequency threshold (about once per million). Developmentally, it is also known that children's vocabulary acquisition correlates with word frequencies in child-directed speech (Goodman et al., 2008), especially for open-class words, the primary arena of rule productivity. These two considerations suggest that the bulk of young children's vocabulary can be found in the most frequent words in the language. For instance, the Chicago Corpus, a large longitudinal study of vocabulary acquisition by 62 children (Rowe and Goldin-Meadow 2009; listed as an appendix of Carlson et al. 2014), produced a list of words assessed to be available to most English-learning children prior to 50 months. The vast majority can be found in the top 1,000 most frequent words in child-directed speech (MacWhinney, 2000). Finally, the calibration of productivity under the TP deals with the type frequency of words, and more specifically, the proportion of *e* relative to *N*. It is immaterial exactly what these words are, or how frequently they appear — assuming, of course, they are frequent enough to be learned by young children. Therefore, while different learners necessarily know different words, for obvious and non-obvious reasons (Fig 18.2) their grammars may still be the same if the relative proportions of *N* and *e* fall on the same side of productivity. As Kodner (2019) shows for a cross-linguistic/genre study, the vocabulary necessary for children's rule learning can be effectively bootstrapped from adult corpora; the simulated samples almost always result in the same rules. The TP can thus apply, at least as a suggestive model for rule learning even in the absence of precise vocabulary data from child learners.

4.3 Applying the Tolerance Principle

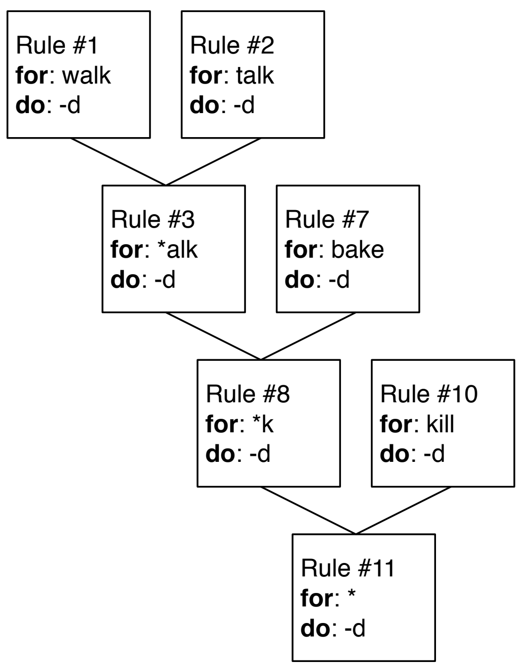
The application of the TP is a typical example of the hypothesis testing approach to language learning (Chomsky, 1965, Wexler and Culicover, 1980, Berwick, 1985, Trueswell et al., 2013). The learner detects a hypothesis, i.e., a rule *R*, and then tests its productivity numerically with the quantity of *N* and *e*. If *R* is productive, the learner will generalize it; otherwise the learner will attempt to formulate a different rule and the process repeats.

It is important to emphasize that the formation of a hypothesis and its subsequent evaluation are in principle two independent processes. For concreteness, consider linguistic rules of the form:

(6) R: IF *A* THEN *B*

where *A* provides the structural description for *R* which, if met, triggers the application of *B*, the structural change. Quite generally, learning can be framed as a search problem that identifies the structural descriptions of the items that undergo a specific structural change. All explicit learning models in the study of language (e.g., Chomsky, 1955, Ivimey, 1975, Berwick, 1985, Skousen et al., 2002, Albright and Hayes, 2003) as well as from adjacent fields such as artificial intelligence (Mitchell, 1982, Cohen, 1995, Yip and Sussman, 1997, Daelemans et al., 2009) and cognitive psychology (Medin and Schaffer, 1978, Feldman, 2000, Osherson and Smith, 1981) converge on a shared insight: inductive learning must proceed conservatively, drawing minimal generalizations from the data. For concreteness, we illustrate the inductive process using the Yip-Sussman model (Yip and Sussman 1997 and extended by Molnar 2001) on the familiar example of English past tense.

The learner constructs rules as mapping relations between input (stem) and output (past tense). Both input and output are represented as a linear sequence of phonemes specified by their distinctive features in the Yip-Sussman model but we will use English orthography for ease of presentation. The operation of the model is presented in Fig 18.3 with a sequence of input words that becomes incrementally available to the learner.



The learning of the regular rule (*-d*) adapted from Molnar 2001.

The model from ever more inclusive generalizations over the phonological properties of the verbs that take *-d*. Eventually, it concludes that *-d* has no restrictions whatever (⋆'s all around in Fig 18.3) the generality of a rule is directly related to the diversity of words it applies to.

Generalization-based models as in Fig 18.3 do not have the capacity to distinguish productive from unproductive rules. Indeed, when executed on English, it produces rules in (7) in addition to the *-d* rule:

(7) a. Rime → ɔt/ __ (e.g., *think-thought*, *catch-caught*, *buy-bought*, *bring-brought*, *seek-sought*, *fight-fought*, *teach-taught*)

b. d → t / en __ (e.g., *bend-bent*, *lend-lent*, *spend-spent*)

c. ɪ → æ / __ ŋ (e.g., *sing-sang*, *ring-rang*)

These rules, however, are not constrained for productivity. For example, the high degree of heterogeneity for words in (7a) results in a rule that places no restrictions on its application, such that any verb take on "ought" for past tense, which is clearly illicit. Similarly, the rule (7b) would replace /d/ with /t/ for verbs with the rime /end/, erroneously turning *blend* into *blent*, *mend* into *ment*, and *end* into *ent*, which no English speaker does.

The TP effectively places a cap on the productivity of rules. The rules in (7) have open-ended scopes of application but they would be almost immediately assessed as unproductive: the number of items that follow the rules is a tiny fraction of those that fit the structural description but do not the rules. It is especially interesting to study rules such as (7c), namely the verbs that end in /ɪŋ/ ('-ing'), the only kind that adults but only in experimental conditions feel the temptation to irregularize (Bybee and Moder, 1983, Albright and Hayes, 2003). There are only 14 such verbs in English that fall into this category:

(8) a. "ought": bring (1)

b. ɪ → æ / __ ŋ: sing, ring, spring (3)

c. ɪ → ʌ / __ ŋ: swing, string, sting, fling, cling, sling, wring (7)

d. Regular: wing, zing, ding (3)

None of these patterns, including the plurality pattern (8c), is numerically large enough to achieve productivity because the maximum number of exceptions is only $\vartheta_{14} = 5$. It is clear, however, that the TP leaves room for individual and dialect variation (Herman and Herman, 2014). If a learner happens to receive input data where an overwhelming majority – as determined by the numerical relationship between N

and *e* – then the speaker can achieve productivity for patterns that are not productive in the "standard" English variety.

The *-d* suffix, by contrast, can easily clear the threshold: for some 100 irregular verbs, some 650 regular verbs are sufficient to establish productivity. However, the numerical condition of the TP (Table 1) suggests that this will be a protracted development. Of the top 200 verbs inflected in the past tense (MacWhinney, 2000), 76 are irregulars. Because ϑ_{200} is only 37, it follows that children who know some 200 most frequent verbs cannot recognize the productivity of *-d* despite its statistical predominance. Productivity can only result when the number of regular verbs thoroughly overwhelm that of irregulars, as the TP provides a precise measure of the critical mass necessary for productivity suggested in the previous literature (Ivimey, 1975, Rumelhart and McClelland, 1986, Marchman and Bates, 1994). The individual differences in the emergence of productivity (e.g., Adam and Eve reviewed in Section 3.2) appears to the children's vocabulary size (Yang, 2016b, Section 4.1.2).

The TP can be applied fairly mechanically, provided that the researcher can identify plausible generalizations from the data on the basis of independently motivated developmental assumptions while quantifying the vocabulary counts that follow or defy such generalizations. It is intended to operate for productivity calculation on all linguistic levels. For instance, phonological alternation, typically described as allophonic rules, may also have exceptions. A well-known example is the Philadelphia "short-a" system (Labov, 1989). It tenses /æ/ in a well-defined set of morpho-phonological contexts but there are lexically specific exceptions. While /æ/ is lax before voiced stops, *mad*, *bad*, and *glad* tense, and while /æ/ productively tenses before tautosyllabic front nasals, three high frequency irregular past tense forms, i.e., *ran*, *swam*, and *began*, do not. The rules and their exceptions are all reliably acquired by native Philadelphia speakers Roberts and Labov (1995), and the TP can be applied to account for how the distribution of /æ/ is acquired and how it may change over time as the input experience changes (Sneller et al., 2019).

When a rule defined over a set of words fails to reach productivity, the set will be partitioned further and productive patterns may be detected within the resulting subsets. From a cross-linguistic perspective, such recursive application of the TP is likely the norm: morphological systems in general have "nested" patterns where productivity is defined over complementary distributions of words along some structural dimension (e.g., phonological, semantic, gender, and what is descriptively known as conjugations and declensions), as the simple default-plus-exception case of English past tense being an anomaly. As an illustrative example, consider the distribution of the noun plural suffixes in German: *-s*, *-n*, *-e*, *-er*, and *-ø* (the null suffix). Notably, the *-s* suffix covers only a small minority of nouns. Table 2 provides the statistics based on some 450 highly frequent noun plurals from child-directed German speech. The distribution of the suffixes is similar to data from larger corpora (e.g., Clahsen, 1999). This is a significant fact, suggesting that grammatical rules acquired from a child-sized vocabulary would generalize to larger data sets.

Table 2 Distribution of noun plural suffixes for highly frequent nouns in child-directed German

Suffix	Types	Percentage
-ø	87	18.9
-e	156	34.1
-er	30	6.5
-n	172	37.5
-s	13	2.8

An application of the TP clearly will not identify a productive suffix in Table 2 since none is anywhere near requisite threshold. Yang (2016b) reports several detailed case studies of this type: when children fail to discover a productive rule over a set of lexical items, they will subdivide the set along some suitable dimension and apply the TP recursively. For the German plural system, the relevant dimensions are gender (Mills, 1986) and the phonology of the final syllable (e.g., Wiese 1996), which children appear to acquire in conjunction (Szagun et al., 2007). Applying to the TP to the subdivided classes of nouns in Table 2 produces the correct results. The net effect is that almost all nouns are predictably accounted for by the four suffixes: each suffix will still have exceptions but the number of them fall under their respective tolerance threshold. For example, 146 out of the 166 feminine nouns take the *-n* suffix, easily clearing the threshold (134): the remaining 20 would be memorized as exceptions to the feminine rule, which children indeed occasionally over-regularize with *-n* (Gawlitzek-Maiwald, 1994, Elsen, 2002). The remaining non-feminine nouns, however, still do not productively support any of the suffixes, but further partitioning by gender (masculine vs. neuter) and the phonological condition on the final syllable yields productive rules for *-ø*, *-e*, and *-er*, although each rule has exceptions that must be rote-memorized. This removes almost all nouns from consideration when it comes to the *-s* suffix, which has no structural restrictions on the noun and thus becomes the default.

Finally, recall that the TP tolerates a relatively low level of exceptions (as indicated by the logarithmic function; see Table 1). This accounts for the protracted development of productive rules with many and especially high frequency exceptions (e.g. *-d*) and will also detect the absence of productivity when none of the alternations meets the threshold. As noted in Section 3.3, the phenomenon of morphological gaps is difficult to account for by acquisition and theoretical models that assume the existence of a productive default. It similarly poses challenges for learning models especially those probabilistic in nature (e.g., Rumelhart and McClelland, 1986, Skousen, 1989, Albright and Hayes, 2003, Tenenbaum et al., 2011, O'Donnell, 2015), that produce the most favored output or sample from alternatives.

Consider again the Polish masculine genitive singular (GEN.SG) suffixes, where neither *-a* nor *-u* is productive and the learner must resort to rote learning (Section 3.3). In contrast, the genitive plural (GEN.PL) for masculines is unproblematic: the default suffix is *-ów* with a small number of exceptional nouns taking *-i/-y*. Children acquiring Polish make very few errors in the genitive singular (GEN.SG), and frequently overextend *-ów* in the GEN.PL., just as this description leads us to expect (Dąbrowska, 2001). Applied to noun stems found in child-directed Polish, the TP provides a straightforward account of both patterns, summarized in Table 3; see Gorman and Yang (2018) for details.

Table 3 Distributions of genitive suffixes on Polish masculine nouns, the productivity predictions of the Tolerance Principle, average frequency (mean number of tokens per million words), and children's error rates

	Suffix	Types	Productive?	Avg. freq.	Child error rate
GEN.SG:	*-a*	837 (62%)	No	7.2	1.28%
	-u	516 (38%)	No	8.8	0.24%
GEN.PL:	*-ów*	551 (90%)	Yes	6.5	0.41%
	-y/-i	61 (10%)	No	11.4	15.53%

For the singulars, neither *-a* nor *-u* can be productive because the maximum number of exceptions is 187 ($1353/\ln 1353$). The absence of a productive suffix offers no opportunity for over-regularization and children's performance on both suffixes is consistently high. For the plural, however, the productivity of the *-ów* suffix is unperturbed by the presence of 61 exceptions that fall far below the threshold of 95 ($612/\ln 612$). It thus serves as the attractor for over-regularization: to wit, *-i/-y* nouns have the highest error rates by far, even though they have a higher average frequency in child-directed Polish.

5 From Form to Meaning

The conception of productivity in this review, and the learning-theoretic approach to it, take the position that the formal system in language is established by distributional means from the input (Harris, 1951, Chomsky, 1955, 1957). If so, we may begin to contemplate how it figures into the traditional conception of language and cognition. On the one hand, a learning model capable of capturing linguistic generalizations from the input data may change the perspective on explanatory adequacy (Chomsky, 1965). A primary motivation for domain specific principles of Universal Grammar is to constrain the child's hypothesis space and eliminate logically possible but empirically unattested options. But such principles may be dispensable if the learning model can prevent wild hypotheses all by itself. Children would not even attempt to generalize a property from two verbs to twenty or one hundred; see the discussion of (7). And crazy rules – e.g., only verbs ending in nasals can passive – would be instantly rejected even if they were entertained by the learner. The net result would be a simpler theory of Universal Grammar (Berwick and Chomsky, 2016). On the other, and pursuing the Sausurrean theme that the formal correspondences between linguistic form and meaning bring order to thought, it is possible that the acquisition of productivity, which crucially depends on experience (e.g., N and e), may result in differential trajectories in children's conceptual development across cultures – as a result of their language (Yang, 2016a). In this final section, I will briefly discuss one case study from each direction,

5.1 Productivity and Linguistic Theory

Consider a well-known problem in the acquisition of argument structure and syntax-semantic mapping, the English double-object construction (Baker, 1979):

(9) a. John told the students a story.

b. *John said the students a story.

c. John offered the students a book.

d. *John donated the students a book.

e. John promised the students a pizza.

f. *John delivered the students a pizza.

A substantial linguistic literature is concerned with providing precise characterizations of these verbs and their structural properties, to an extent that their distributional patterns as (9) can be predicted; see Harley and Miyagawa (2017) for a recent review. Likewise, the language acquisition literature (e.g., Pinker, 1989, Levin, 1993) assume (innate) mapping relations between the semantic properties of verbs and their syntactic manifestations; see Jackendoff and Alexiadou (this volume) for an overview of lexical semantics and its relation to syntax. The verbs in (9) indeed fall in the semantic class "caused-possession" (Green, 1974) but that is clearly only a necessary condition and not a sufficient one. In fact, Levin (1993) lists nearly 250 English caused-possession verbs but only less than half participate in the double-object construction, with a high degree of lexical idiosyncrasy as illustrated in (9), and presumably individual speaker variation as well.[5] Moreover, while all languages appear to have the semantic class of "caused possession", their syntactic distributions differ widely. In Korean, for example, the equivalent of the double-object construction is restricted to a handful of verbs (Jung and Miyagawa, 2004), and there are languages such as Chamorro that disallow the double-object construction altogether (Chung, 1998). Whatever innate syntax-semantic mapping is available, learning from language specific and lexically idiosyncratic data is inevitable.

While children must be able to grasp the concept of caused possession and form verb semantic classes, it does not seem necessary to assume that they also need to have any prior expectation of the syntactic construction associated with these verbs. In fact, a minimalist distributional learning model of hypothesis formation and testing can be proposed:

[5] In the history of English, the syntactic distributions of these verbs have changed considerably even though their meanings have remained stable (e.g., Visser, 1963, Mitchell, 1985).

(10) a. *Observation*: A child learner observes a set of verbs $V_1, V_2, \ldots, V_M$ that participate in the syntactic construction "V NP NP."

b. *Hypothesis formation*: The learner proceeds to inductively identify a semantic class C, over the verbs $V_1, V_2, \ldots, V_M$.

c. *Hypothesis testing*: The learner identifies the total number of verbs ($N, N \geq M$) in their vocabulary that belong in the semantic class C.

- If $(N - M) < \vartheta_N$ then the learner extends the use of double objects to all members of C.

- Otherwise the learner lexicalizes the M verbs as allowing double-objects but will not extend the construction to any other item.

The model in (10) can be deployed on corpora representative of language acquisition data. A five-million-word corpus of child-directed English (MacWhinney, 2000) contains 42 verbs that appear in the double-object frame "V NP NP" (*Observation*; 10a). Of these, 38 have the clear semantic property of caused possession: the 4 exceptions are the performative verbs *call*, *consider*, *name*, and *pronounce*, which nevertheless fall below the Tolerance threshold ($42/\ln 42 = 11$). These generalizations can in principle be derived automatically by models such as the Yip-Sussman learner (Fig 18.3) if the semantics of verbs is decomposed into more primitive feature representations as in the structuralist tradition of componential analysis (e.g., Nida, 1975) and its modern descendants (Fillmore, 1968, Dowty, 1979, 1991, Jackendoff, 1990). Thus, the *hypothesis formation* step (10b) succeeds as the learner may conclude that the caused possession verb class is a necessary condition for the double object construction. In the *hypothesis testing* step (10c), the learner would evaluate all of the verbs in their vocabulary with the semantics of caused possession — 49 from the corpus — to see if the subset actually attested in the construction (i.e., 38) is sufficient for generalization: just about, as 37 is required ($49 - 49/\ln 49$). In other words, a reasonable sample of English input data supports a productive correspondence between the semantics of caused possession and the syntax of the double-object construction. The well-documented errors in child English (Bowerman, 1982, Gropen et al., 1989) such as *I said her no*, *Shall I whisper something*, *I am going to deliver you some milk*, etc. are thus expected.

To get a more complete picture of the English speaker's knowledge of the construction and to account for the distributional patterns in (9), we need to go beyond the child-directed sample and approximate the linguistic input of a typical speaker. Bootstrapping off CHILDES into larger corpora (see Yang 2016b, 207ff for details), we can obtain a list of dative verbs, sorted by frequency, that provides important insight on how productivity changes as a function of the learner's vocabulary.

Table 4 Caused-possession verbs and their expected distribution in the double-object construction.

Top N	Yes	No	ϑ_N	Productive?
10	9	1	4	Yes
20	17	3	7	Yes
30	26	4	9	Yes
40	30	10	11	Yes
50	34	16	13	No
60	39	21	15	No
70	43	27	16	No
80	46	24	18	No
92	50	42	20	No

Table 4 shows that if a child only learns from the most frequent dative verbs, such as those found in CHILDES, the double object construction will be deemed productive because the vast majority of these verbs — sufficiently many as assessed by the TP — will be attested in the construction. Indeed, the semantics of the most frequent verbs have very salient manifestations in language use that involve the physical transfer of objects (e.g., *give*, *hand*, *bring*, *throw*) or the transmission of information (e.g., *tell*, *ask*, *read*), which may help the child to quickly form the semantic generalization of caused possession as described in the hypothesis formation step in (10b). However, as the learner's vocabulary expands, the construction will no longer meet the productive threshold: the hypothesis testing step in (10c) eventually fails. The learner will then retreat from the over-generalization and lexically memorize those verbs that do participate in the construction: "I said you something" will then disappear. Under the TP approach, the learner has only one hypothesis at any time: the grammar is either productive or not. Thus, the conundrum that arises from choosing between an unproductive hypothesis and a productive hypothesis is not an issue, sidestepping the Subset Problem and the ineffective use of indirect negative evidence (Yang, 2017).

If this line of approach is correct, then children can directly acquire language specific rules that grammaticalize conceptual categories and relations such as animacy, causation, intentionality, quantity, etc. These categories and relations are likely innately available to the pre-linguistic infants (Leslie and Keeble, 1987, Gergely et al., 1995, Kotovsky and Baillargeon, 1998, Woodward, 1998, Feigenson et al., 2004, Sommerville et al., 2005, Spelke and Kinzler, 2007, Carey, 2009); they are also invoked as primitives as in the decompositional representation of linguistic meaning (Fillmore, 1968, Dowty, 1991), to which generalization models such as Fig 18.3 and principles of productivity assessment such as the TP can apply. Importantly, recall that children acquire the core syntax-semantics rules very early (Brown, 1973, Golinkoff et al., 1987, Naigles, 1990); it must have been done on a relatively smaller vocabulary, under which the TP would be especially robust and effective. This reduces the need for an innate, universal, and highly domain-specific theory of syntax and semantics mappings, which in any case would be difficult to maintain given the level of lexical idiosyncrasy across languages (Pinker, 1989, Borer, 1994, Levin and Hovav, 1995).

5.2 Productivity and Conceptual Development

As a final example, consider the development of number concepts in children (Gelman and Gallistel, 1978, Dehaene, 2011). Despite the apparent connection between natural language and natural number (Chomsky, 1988), and in contrast to the very early acquisition of language, children develop number knowledge in a protracted fashion (Feigenson et al., 2004, Carey, 2009). While very small numbers can be understood innately (Wynn, 1992) or by direct experience (e.g., counting objects; Mix et al. 2002), the concept of large numbers can only be obtained via generalization: specifically, the Successor Function, that every number is followed by another number exactly one greater. The impact of language in children's understanding of number has long been noted (Hale, 1975, Miller et al., 1995, Fuson and Kwon, 1991, Siegler and Mu, 2008, Gordon, 2004, Pica et al., 2004) although the underlying causal mechanism has remained obscure.

The Successor Function is a semantic relation that holds between two number concepts. If, by hypothesis, semantic systematicity can only follow from formal systematicity, then the Successor Function can only be acquired when the child establishes the systematic formal relation between successive placeholder expressions (i.e., numerals) that represent the concepts. This is strongly analogous to the acquisition of linguistic rules that have exceptions. As discussed in Fig 18.1, children's consistency of tense marking in obligatory context increases significantly after the very first instance of over-regularization, i.e., the discovery that *-d* is *formally* productive. Prior to that point, their knowledge of tense marking, and perhaps the notion of "pastness", appears to be lexically specific (Ervin, 1964, Kuczaj, 1977, Marcus et al., 1992).

As previously observed (Fuson, 1988), English numerals have 17 expressions (1-13, 15, 20, 30, 50) that require some kind of rote memorization as they do not conform to the transparent structure of digit in combination with teens and decades. To discover the productive rules for counting, the child must learn a sufficiently large number of "regular" numerals to overcome the 17 exceptions. The problem of learning to count, then, reduces to a problem of productivity in language acquisition. According to the Tolerance Principle, English-learning children need to count to at least 73 to learn the English numeral rules as $N = 73$ is the small N such that $N/\ln N = 17$. This prediction receives support from the study of children's counting sequences (Fuson, 1988): once English-learning children can count to the 70s, they generally continue all the way to 100 where counting tasks typically conclude (Fuson et al., 1982). Moreover, it appears that productive counting is necessary for generalizing the successor relation that holds of the small numbers. For instance, only children who can count past 80 show systematic knowledge of the Successor Function (Cheung et al., 2017) although no theoretical reason was given for this observation.

The productivity based approach to number makes a strong claim across languages, as the TP makes precise predictions about the tipping point at which counting becomes productive for specific languages: significant change in children's understanding of the Successor Function should follow. A recent study (Yang et al., 2019) provides a direct test with Cantonese. The Cantonese numeral system, like the Chinese system widely adopted in East Asian languages, have a very transparent counting system. Only 12 numerals are idiosyncractic and require rote-memorization: 1-10, and two linear patterns of teens (i.e, ten-digit) and decades (i.e., digit-ten). By hypothesis, children only need to count past 46 to understand the Successor Function. Indeed, unproductive counters (i.e., those who could not count past 50) performed at chance on a binary outcome task (Sarnecka and Carey, 2008) that assesses their understanding of the Successor Function. By contrast, productive counters who counted past 50 performed at nearly 90% across a wide range of numbers. Such a quantal change in the productivity of counting and the understanding of number is very similar to the English past tense: again, the formal productivity of *-d* is critical for the development of tense. It turns out that Cantonese-learning children show an understanding of the Successor Function over a full year earlier than English-learning children: the cognitive advantage derives from nothing but a simpler linguistic system.

6 Summary

This chapter reassesses some very traditional ideas about the nature of language — Saussurean arbitrariness and Humboltian infinity – in light of progress in the study of linguistic productivity, language acquisition, and cognitive science. The formal nature of the learning-theoretic approach should not overshadow the fact that the calibration of productivity is ultimately a psychological process, which interacts with other components of cognition and perception that we share with our biological relatives (Hauser et al., 2002, Berwick and Chomsky, 2016). Challenging problems lie ahead. We need to further develop formal and quantitative learning models that conform with the developmental evidence (Section 2 and 3). To the extent this direction of research holds promise (Section 4), it raises fundamental questions about the nature of language and its place in cognition (Section 5). In particular, the Tolerance Principle provides a causal mechanism that connects the acquisition of the vocabulary (i.e., arbitrariness) to the rise of productivity (i.e., systematicity).

Saussure's vision of language as the medium between form and meaning is remarkably consistent with the view that emerged after sixty years of generative grammar and cognitive science. He likened thought to an "indefinite plan of jumbled ideas." Phonetic sounds are "equally vague" given their inherent variability in a continuous acoustic space. The direct correspondence between these "two shapeless masses" provides very little expressive power – like the English irregular verbs, essentially a random list of word forms. Such a correspondence is necessarily finite and highly dependent on experience. The principle of arbitrariness is merely the starting point, or a design specification, to be rescued by the "principle of order and regularity", which, on my view, is a theory of productivity which operates formally on quantitative basis. And that appears to be exactly Saussure's conception of language as a science: "Linguistics then works in the borderland where the elements of sound and thought combine: their combinations produce a form, not a substance" (CLG, 113).

References

Albright, A. and Hayes, B. (2003). Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition*, 90(2):119–161.

Allen, S. (1996). *Aspects of argument structure acquisition in Inuktitut*. John Benjamins Publishing, Amsterdam.

Ambridge, B. (2010). Children's judgements of regular and irregular novel past-tense forms: New data on the English past-tense debate. *Developmental Psychology*, 46(6):1497–1504.

Ambridge, B., Kidd, E., Rowland, C. F., and Theakston, A. L. (2015). The ubiquity of frequency effects in first language acquisition. *Journal of Child Language*, 42(02):239–273.

Anderson, J. R. (1980). *Cognitive psychology and its implications*. San Francisco. W. H. Freeman.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, 98(3):409–429.

Anderwald, L. (2009). *The morphology of English dialects: Verb-formation in non-standard English*. Cambridge University Press, Cambridge.

Anderwald, L. (2013). Natural language change or prescriptive influence?: Throve, dove, pled, drug and snuck in 19th-century American English. *English World-Wide*, 34(2):146–176.

Anisfeld, M., Rosenberg, E. S., Hoberman, M. J., and Gasparini, D. (1998). Lexical acceleration coincides with the onset of combinatorial speech. *First Language*, 18(53):165–184.

Anisfeld, M. and Tucker, G. R. (1967). English pluralization rules of six-year-old children. *Child development*, 38(4):1201–1217.

Armstrong, S. L., Gleitman, L. R., and Gleitman, H. (1983). What some concepts might not be. *Cognition*, 13(3):263–308.

Aronoff, M. (1976). *Word formation in generative grammar*. MIT Press, Cambridge, MA.

Baerman, M., Corbett, G. G., and Brown, D., editors (2010). *Defective paradigms: Missing forms and what they tell us*. Oxford University Press, Oxford.

Baker, C. L. (1979). Syntactic theory and the projection problem. *Linguistic Inquiry*, 10(4):533–581.

Bates, E., Bretherton, I., and Snyder, L. S. (1988). *From first words to grammar: Individual differences and dissociable mechanisms*. Cambridge University Press, Cambridge.

Bauer, L. (1983). *English word-formation*. Cambridge Textbooks in Linguistics. Cambridge University Press, Cambridge.

Becker, M., Ketrez, N., and Nevins, A. (2011). The surfeit of the stimulus: Analytic biases filter lexical statistics in Turkish laryngeal alternations. *Language*, 87(1):84–125.

Berko, J. (1958). The child's learning of English morphology. *Word*, 14(2–3):150–177.

Berwick, R. (1985). *The acquisition of syntactic knowledge*. MIT Press, Cambridge, MA.

Berwick, R. C. and Chomsky, N. (2016). *Why only us: Language and evolution*. MIT Press, Cambridge, MA.

Bloch, B. (1947). English verb inflection. *Language*, 23(4):399–418.

Bloomfield, L. (1933). *Language*. Henry Holt & Co., New York.

Borer, H. (1994). The projection of arguments. *University of Massachusetts occasional papers in linguistics*, 17:19–47.

Botha, R. P. (1969). *The function of the lexicon in transformational generative grammar*. Mouton, Berlin.

Bowerman, M. (1982). Reorganizational process in lexical and syntactic development. In Wanner, E. and Gleitman, L. R., editors, *Language acquisition: The state of the art*, pages 319–346. Cambridge University Press, New York.

Bresnan, J. and Nikitina, T. (2009). On the gradience of the dative alternation. In Wee, L. and Uyechi, L., editors, *Reality explorations and discovery: Pattern interaction in language and life*, pages 161–184. CSLI Publications, Stanford, CA.

Brown, R. (1957). Linguistic determinism and the part of speech. *The Journal of Abnormal and Social Psychology*, 55(1):1–5.

Brown, R. (1973). *A first language: The early stages*. Harvard University Press, Cambridge, MA.

Bybee, J. (1985). *Morphology: A study of the relation between meaning and form*. John Benjamins, Philadelphia.

Bybee, J. and Moder, C. L. (1983). Morphological classes as natural categories. *Language*, 59(2):251–270.

Bybee, J. L. and Slobin, D. (1982). Rules and schemas in the development and use of the English past tense. *Language*, 58(2):265–289.

Caprin, C. and Guasti, M. T. (2009). The acquisition of morphosyntax in Italian: A cross-sectional study. *Applied Psycholinguistics*, 30(1):23–52.

Carey, S. E. (2009). *The origin of concepts*. Oxford University Press, New York.

Carlson, M. T., Sonderegger, M., and Bane, M. (2014). How children explore the phonological network in child-directed speech: A survival analysis of children's first word productions. *Journal of memory and language*, 75:159–180.

Cazden, C. B. (1968). The acquisition of noun and verb inflections. *Child Development*, 39:433–448.

Cheung, P., Rubenson, M., and Barner, D. (2017). To infinity and beyond: Coutning procedures precede learning the logic of the natural numbers. *Cognitive Psychology*, 92:22–36.

Chomsky, N. (1955). The logical structure of linguistic theory. Ms., Harvard University and MIT. Revised version published by Plenum, New York, 1975.

Chomsky, N. (1957). *Syntactic structures*. Mouton, The Hague.

Chomsky, N. (1959). A review of B.F. Skinner's *Verbal Behavior*. *Language*, 35(1):26–58.

Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT Press, Cambridge, MA.

Chomsky, N. (1988). *Language and problems of knowledge*. MIT press, Cambridge, MA.

Chung, S. (1998). *The design of agreement: Evidence from Chamorro*. University of Chicago Press, Chicago.

Clahsen, H. (1999). Lexical entries and rules of languge: A multidisciplinary study of German inflection. *Behavioral and Brain Sciences*, 22:991–1069.

Clahsen, H., Aveledo, F., and Roca, I. (2002). The development of regular and irregular verb inflection in Spanish child language. *Journal of Child Language*, 29(3):591–622.

Clahsen, H. and Penke, M. (1992). The acquisition of agreement morphology and its syntactic consequences: New evidence on German child language from the Simone Corpus. In Meisel, J., editor, *The acquisition of verb placement*, pages 181–234. Kluwer, Dordrecht.

Clahsen, H. and Rothweiler, M. (1993). Inflectional rules in children's grammars: Evidence from German participles. *Yearbook of Morphology*, pages 1–34.

Cohen, W. W. (1995). Fast effective rule induction. In *Proceedings of the Twelfth International Conference on Machine Learning, Lake Tahoe, California*.

Dąbrowska, E. (2001). Learning a morphological system without a default: The Polish genitive. *Journal of Child Language*, 28(3):545–574.

Dąbrowska, E. (2005). Productivity and beyond: mastering the Polish genitive inflection. *Journal of Child Language*, 32(1):191–205.

Daelemans, W., Zavrel, J., van der Sloot, K., and van den Bosch, A. (2009). TiMBL: Tilburg Memory-Based Learner. Technical Report ILK 09-01, Induction of Linguistic Knowledge Research Group, Tilburg University.

De Villiers, J. G. and De Villiers, P. A. (1973). A cross-sectional study of the acquisition of grammatical morphemes in child speech. *Journal of psycholinguistic research*, 2(3):267–278.

Deen, K. U. (2005). *The acquisition of Swahili*. John Benjamins Publishing, Amsterdam.

Dehaene, S. (2011). *The number sense: How the mind creates mathematics*. Oxford University Press, New York.

Demuth, K. (2003). The acquisition of Bantu languages. In *The Bantu languages*, pages 209–222. Curzon Press Surrey,, United Kingdom.

Derks, P. L. and Paclisanu, M. I. (1967). Simple strategies in binary prediction by children and adults. *Journal of Experimental Psychology*, 73(2):278.

Derwing, B. L. and Baker, W. J. (1977). The psychological basis for morphological rules. In Macnamara, J., editor, *Language learning and thought*, pages 85–110. Academic Press, New York.

Di Sciullo, A.-M. and Williams, E. (1987). *On the definition of word*. MIT Press, Cambridge, MA.

Dowty, D. (1991). Thematic proto-roles and argument selection. *language*, 67(3):547–619.

Dowty, D. R. (1979). *Word meaning and Montague grammar*. Reidel, Dordrecht.

Dye, C., Kedar, Y., and Lust, B. (2019). From lexical to functional categories: New foundations for the study of language development. *First Language*, 39(1):9–32.

Eddington, D. (1996). Diphthongization in Spanish derivational morphology: An empirical investigation. *Hispanic Linguistics*, 8(1):1–13.

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, 48(1):71–99.

Elsen, H. (2002). The acquisition of German plurals. In *Morphology 2000: Selected Papers from the 9th Morphology Meeting, Vienna, 25-27 February 2000*, volume 218, page 117. John Benjamins Publishing.

Ervin, S. M. (1964). Imitation and structural change in children's language. In Lenneberg, E. H., editor, *New directions in the study of child language*, pages 163–189. MIT Press, Cambridge.

Ervin, S. M. and Miller, W. R. (1963). Language development. In Stevenson, H., editor, *Child psychology: The sixty-second yearbook of the National Society for the Study of Education*, pages 108–143. University of Chicago Press.

Fanselow, G. and Féry, C. (2002). Ineffability in grammar. In Fanselow, G. and Féry, C., editors, *Resolving conflicts in grammars: Optimality Theory in syntax, morphology, and phonology*, pages 265–307. Helmut Buske, Hamburg.

Fazly, A., Alishahi, A., and Stevenson, S. (2010). A probabilistic computational model of cross-situational word learning. *Cognitive Science*, 34(6):1017–1063.

Feigenson, L., Dehaene, S., and Spelke, E. (2004). Core systems of number. *Trends in cognitive sciences*, 8(7):307–314.

Feldman, J. (2000). Minimization of boolean complexity in human concept learning. *Nature*, 407(6804):630–633.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C. B., and Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, pages i–185.

Figueroa, M. and Gerken, L. (2019). Experience with morphosyntactic paradigms allows toddlers to tacitly anticipate overregularized verb forms months before they produce them. *Cognition*, 191:103977.

Fillmore, C. J. (1968). The case for case. In Bach, E. and Harms, R. T., editors, *Universals of linguistic theory*, pages 1–88. Holt, Rinehart and Winston, New York.

Frank, M. C., Goodman, N. D., and Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science*, 20(5):578–585.

Fuson, K. C. (1988). *Children's counting and concepts of number*. Springer, Berlin.

Fuson, K. C. and Kwon, Y. (1991). Chinese-based regular and european irregular systems of number words: The disadvantages for english-speaking children. *Language in mathematical education: Research and practice*, pages 211–226.

Fuson, K. C., Richards, J., and Briars, D. J. (1982). The acquisition and elaboration of the number word sequence. In *Children's logical and mathematical cognition*, pages 33–92. Springer.

Gawlitzek-Maiwald, I. (1994). How do children cope with variation in the input? the case of German plurals and compounding. In Tracy, R. and Lattey, E., editors, *How tolerant is Universal Grammar? Essays on language learnability and language variation*, pages 225–266. Niemeyer, Tübingen.

Gelman, R. and Gallistel, C. R. (1978). *The child's understanding of number*. Harvard University Press.

Gergely, G., Nádasdy, Z., Csibra, G., and Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, 56(2):165–193.

Gillette, J., Gleitman, H., Gleitman, L., and Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, 73(2):135–176.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, 1(1):3–55.

Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., and Trueswell, J. C. (2005). Hard words. *Language Learning and Development*, 1(1):23–64.

Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., and Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language*, 14(01):23–45.

Goodman, J. C., Dale, P. S., and Li, P. (2008). Does frequency count? parental input and the acquisition of vocabulary. *Journal of child language*, 35(3):515–531.

Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science*, 306(5695):496–499.

Gorman, K. and Yang, C. (2018). When nobody wins. In Rainer, F., Gardani, F., Luschützky, H. C., and Dressler, W. U., editors, *Competition in inflection and word formation*, pages 169–193. Springer, Berlin.

Graves, M. F. and Koziol, S. (1971). Noun plural development in primary grade children. *Child Development*, pages 1165–1173.

Green, G. M. (1974). *Semantics and syntactic regularity*. Indiana University Press.

Gropen, J., Pinker, S., Hollander, M., Goldberg, R., and Wilson, R. (1989). The learnability and acquisition of the dative alternation in English. *Language*, 65(2):203–257.

Guasti, M. T. (1993). Verb syntax in Italian child grammar: Finite and nonfinite verbs. *Language Acquisition*, 3(1):1–40.

Gxilishe, S., de Villiers, P., de Villiers, J., Belikova, A., Meroni, L., and Umeda, M. (2007). The acquisition of subject agreement in Xhosa. In *Proceedings of the Conference on Generative Approaches to Language Acquisition (GALANA)*, volume 2, pages 114–123.

Hahn, U. and Nakisa, R. C. (2000). German inflection: Single route or dual route? *Cognitive Psychology*, 41(4):313–360.

Hale, K. L. (1975). Gaps in grammar and culture. In Kinkade, M. D., Hale, K. L., and Werner, O., editors, *Linguistics and anthropology: In honor of C. F. Voegelin*, pages 295–315. The Peter de Ridder Press, Lisse.

Halle, M. (1973). Prolegomena to a theory of word formation. *Linguistic Inquiry*, 4(1):3–16.

Halle, M. and Marantz, A. (1993). Distributed morphology and the pieces of inflection. In Hale, K. and Keyser, S. J., editors, *The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger*, pages 111–176. MIT Press, Cambridge, MA.

Harley, H. and Miyagawa, S. (2017). Syntax of ditransitives. In *Oxford Research Encyclopedia of Linguistics*. Oxford University Press.

Harris, J. W. (1969). *Spanish phonology*. MIT Press, Cambridge, MA.

Harris, Z. S. (1951). *Methods in structural linguistics.* University of Chicago Press, Chicago.

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

Hauser, M. D., Chomsky, N., and Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298(5598):1569–1579.

Hay, J. and Baayen, R. H. (2005). Shifting paradigms: Gradient structure in morphology. *Trends in Cognitive Sciences*, 9(7):342–348.

Hayes, B., Zuraw, K., Siptár, P., and Londe, Z. (2009). Natural and unnatural constraints in Hungarian vowel harmony. *Language*, 85(4):822–863.

Herman, L. and Herman, M. S. (2014). *American dialects: A manual for actors, directors, and writers*. Routledge, New York.

Hoff, E. and Naigles, L. (2002). How children use input to acquire a lexicon. *Child development*, 73(2):418–433.

Hudson Kam, C. L. and Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development*, 1(2):151–195.

Huttenlocher, J. (1964). Children's language: Word-phrase relationship. *Science*, 143(3603):264–265.

Ivimey, G. (1975). The development of english morphology: an acquisition model. *Language and speech*, 18(2):120–144.

Jackendoff, R. S. (1990). *Semantic structures*. MIT Press, Cambridge, MA.

Jackendoff, R. S. (1996). *The architecture of the language faculty*. MIT Press, Cambridge, MA.

Jespersen, O. (1922). *Language, its nature, development, and origin*. Allen and Unwin, London.

Jung, Y.-J. and Miyagawa, S. (2004). Decomposing ditransitive verbs. In *Proceedings of SICGG*, pages 101–120.

Keller, F. (2000). *Gradience in grammar: Experimental and computational aspects of degrees of grammaticality*. PhD thesis, University of Edinburgh.

Kim, J. J., Pinker, S., Prince, A., and Prasada, S. (1991). Why no mere mortal has ever flown out to center field. *Cognitive Science*, 15(2):173–218.

Kodner, J. (2019). Estimating child linguistic experience from historical corpora. *Glossa*, 4(1):122.

Kotovsky, L. and Baillargeon, R. (1998). The development of calibration-based reasoning about collision events in young infants. *Cognition*, 67(3):311–351.

Koulaguina, E. and Shi, R. (2019). Rule generalization from inconsistent input in early infancy. *Language Acquisition*, 26(4):416–435.

Kuczaj, S. A. (1977). The acquisition of regular and irregular past tense forms. *Journal of Verbal Learning and Verbal Behavior*, 16(5):589–600.

Labov, W. (1972). *Sociolinguistic patterns*. University of Pennsylvania Press, Philadelphia.

Labov, W. (1989). The exact description of the speech community: Short *a* in philadelphia. In Fasold, R. and Schiffrin, D., editors, *Language change and variation*, pages 1–57. Georgetown University Press, Washington, D.C.

Labov, W., Ash, S., and Boberg, C. (2006). *The atlas of North American English: Phonetics, phonology, and sound change*. Mouton de Gruyter, Berlin.

Landau, B. and Gleitman, L. R. (1985). *Language and experience: Evidence from the blind child*, volume 8. Harvard University Press, Cambridge, MA.

Leslie, A. M. and Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition*, 25(3):265–288.

Levin, B. (1993). *English Verb Classes and Alternations: A Preliminary Investigation*. University of Chicago Press, Chicago.

Levin, B. and Hovav, M. R. (1995). *Unaccusativity: At the syntax-lexical semantics interface*. MIT Press, Cambridge.

MacWhinney, B. (1975). Rules, rote, and analogy in morphological formations by Hungarian children. *Journal of Child Language*, 19(2):65–77.

MacWhinney, B. (1978). *The acquisition of morphophonology*. Monographs of the Society for Research in Child Development. University of Chicago Press.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. Lawrence Erlbaum, Mahwah, NJ, 3rd edition.

Maratsos, M. (2000). More overregularizations after all: New data and discussion on Marcus, Pinker, Ullman, Hollander, Rosen and Xu. *Journal of Child Language*, 27:183–212.

Marchman, V. A. and Bates, E. (1994). Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of Child Language*, 21(2):339–366.

Marchman, V. A., Wulfeck, B., and Weismer, S. E. (1999). Morphological productivity in children with normal language and sli: A study of the english past tense. *Journal of Speech, Language, and Hearing Research*, 42(1):206–219.

Marcus, G., Pinker, S., Ullman, M. T., Hollander, M., Rosen, J., and Xu, F. (1992). *Overregularization in language acquisition*. Monographs of the Society for Research in Child Development. University of Chicago Press, Chicago.

Marcus, G. F. (1995). The acquisition of the English past tense in children and multilayered connectionist networks. *Cognition*, 56(3):271–279.

Marcus, G. F., Brinkmann, U., Clahsen, H., Wiese, R., and Pinker, S. (1995). German inflection: The exception that proves the rule. *Cognitive Psychology*, 29:189–256.

Maslen, R., Theakston, A. L., Lieven, E. V., and Tomasello, M. (2004). A dense corpus study of past tense and plural overregularization in English. *Journal of Speech, Language and Hearing Research*, 47(6):1319–1333.

Mausch, H. (2003). Current alternations in inflection of Polish masculine inanimate nouns in the singular: A pilot study. *Investigationes Linguisticae*, 9:1–21.

Mayol, L. (2007). Acquisition of irregular patterns in Spanish verbal morphology. In Nurmi, V. and Sustretov, D., editors, *Proceedings of the twelfth ESSLLI Student Session*, pages 1–11, Dublin.

McClelland, J. L. and Bybee, J. (2007). Gradience of gradience: A reply to Jackendoff. *The Linguistic Review*, 24(4):437–455.

Medin, D. L. and Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, 85(3):207.

Miller, K. F., Smith, C. M., Zhu, J., and Zhang, H. (1995). Preschool origins of cross-national differences in mathematical competence: The role of number-naming systems. *Psychological Science*, 6(1):56–60.

Mills, A. (1986). *The acquisition of gender: A study of English and German*. Springer, Berlin.

Mitchell, B. (1985). *Old English Syntax: Concord, the parts of speech, and the sentence*. Oxford University Press, USA.

Mitchell, T. M. (1982). Generalization as search. *Artificial Intelligence*, 18(2):203–226.

Mix, K. S., Huttenlocher, J., and Levine, S. C. (2002). *Quantitative development in infancy and early childhood*. Oxford University Press, Oxford.

Molnar, R. A. (2001). "Generalize and sift" as a model of inflection acquisition. Master's thesis, MIT.

Murphy, G. (2004). *The big book of concepts*. MIT Press.

Nagy, W. E. and Anderson, R. C. (1984). How many words are there in printed school English? *Reading Research Quarterly*, 19(3):304–330.

Naigles, L. (1990). Children use syntax to learn verb meanings. *Journal of Child Language*, 17(02):357–374.

Newport, E. (1990). Maturational constraints on language learning. *Cognitive Science*, 14(1):11–28.

Nida, E. A. (1949). *Morphology: the descriptive analysis of words*. University of Michigan Press, Ann Arbor, 2nd edition.

Nida, E. A. (1975). *A componential analysis of meaning: An introduction to semantic structures*, volume 57. Walter de Gruyter, Berlin.

Nosofsky, R. M., Palmeri, T. J., and McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, 101(1):53.

O'Donnell, T. (2015). *Productivity and reuse in language*. MIT Press, Cambridge, MA.

Osherson, D. N. and Smith, E. E. (1981). On the adequacy of prototype theory as a theory of concepts. *Cognition*, 9(1):35–58.

Parducci, A. and Perrett, L. F. (1971). Category rating scales: Effects of relative frequency of stimulus values. *Journal of Experimental Psychology*, 89(2):427–452.

Pica, P., Lemer, C., Izard, V., and Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science*, 306(5695):499–503.

Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In Bybee, J. and Hopper, P., editors, *Frequency and the emergence of linguistic structure*, pages 137–157, Philadelphia. John Benjamins.

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

Pinker, S. (1995). Why the child holded the baby rabbit: A case study in language acquisition. In Gleitman, L. R. and Liberman, M., editors, *An invitation to cognitive science, Vol. 1: Language*, pages 107–133. MIT Press, Cambridge, MA.

Pinker, S. (1999). *Words and rules: The ingredients of language*. Basic Books, New York.

Pinker, S. and Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, 28(1):73–193.

Pizzuto, E. and Caselli, M. C. (1994). The acquisition of Italian verb morphology in a cross-linguistic perspective. In Levy, Y., editor, *Other children, other languages*, pages 137–187. Lawrence Erlbaum, Hillsdale, NJ.

Plunkett, K. and Juola, P. (1999). A connectionist model of English past tense and plural morphology. *Cognitive Science*, 23(4):463–490.

Prasada, S. and Pinker, S. (1993). Generalisation of regular and irregular morphological patterns. *Language and Cognitive Processes*, 8(1):1–56.

Prince, A. and Smolensky, P. (2004). *Optimality Theory: Constraint interaction in generative grammar*. MIT Press, Cambridge, MA.

Pullum, G. K. and Wilson, D. (1977). Autonomous syntax and the analysis of auxiliaries. *Language*, 53(4):741–788.

Quine, W. V. O. (1960). *Word and object*. MIT Press, Cambridge, MA.

Ramscar, M. (2002). The role of meaning in inflection: Why the past tense does not require a rule. *Cognitive Psychology*, 45(1):45–94.

Ringe, D. and Yang, C. (2020). The threshold of productivity and the irregularization of verbs in Early Modern English. In Los, B. and Honeybone, P., editors, *Proceedings of the 20th International Conference on English Historical Linguistics*. John Benjamins, Amsterdam.

Roberts, J. and Labov, W. (1995). Learning to talk Philadelphian: acquisition of short *a* by preschool children. *Language Variation and Change*, 7:101–112.

Rowe, M. L. and Goldin-Meadow, S. (2009). Differences in early gesture explain ses disparities in child vocabulary size at school entry. *Science*, 323(5916):951–953.

Rumelhart, D. E. and McClelland, J. L. (1986). On learning the past tenses of English verbs. In McClelland, J. L., Rumelhart, D. E., and the PDP Research Group, editors, *Parallel distributed processing: Explorations into the microstructure of cognition. Volume 2: Psychological and biological models*, pages 216–271. MIT Press, Cambridge, MA.

Sarnecka, B. W. and Carey, S. (2008). How counting represents number: What children must learn and when they learn it. *Cognition*, 108(3):662–674.

Saussure, F. d. (1916). *Cours de linguistique générale*. Éditions Payot & Rivages ((Wade Baskin 1959 translation), Paris.

Schuler, K. (2017). *The acquisition of productive rules in child and adult language learners*. PhD thesis, Georgetown University, Washington, D.C.

Schuler, K., Yang, C., and Newport, E. (2016). Testing the Tolerance Principle: Children form productive rules when it is more computationally efficient to do so. In *The 38th Cognitive Society Annual Meeting*, Philadelphia, PA.

Schütze, C. T. (2005). Thinking about what we are asking speakers to do. In Kepser, S. and Reis, M., editors, *Linguistic evidence: Empirical, theoretical, and computational perspectives*, pages 457–485. Mouton de Gruyter, Berlin.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, 237(4820):1317–1323.

Shi, R. and Melançon, A. (2010). Syntactic categorization in French-learning infants. *Infancy*, 15(5):517–533.

Siegler, R. S. and Mu, Y. (2008). Chinese children excel on novel mathematics problems even before elementary school. *Psychological Science*, 19(8):759–763.

Sims, A. D. (2006). *Minding the gap: Inflectional defectiveness in a paradigmatic theory*. PhD thesis, Ohio State University.

Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, 61(1):39–91.

Skousen, R. (1989). *Analogical modeling of language*. Kluwer, Dordrecht.

Skousen, R., Lonsdale, D., and Parkinson, D. B., editors (2002). *Analogical modeling: An exemplar-based approach to language*. John Benjamins, Amsterdam.

Sneller, B., Fruehwald, J., and Yang, C. (2019). Using the tolerance principle to predict phonological change. *Language Variation and Change*, 31(1):1–20.

Sommerville, J. A., Woodward, A. L., and Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. *Cognition*, 96(1):B1–B11.

Spelke, E. S. and Kinzler, K. D. (2007). Core knowledge. *Developmental science*, 10(1):89–96.

Stevens, J. S., Gleitman, L. R., Trueswell, J. C., and Yang, C. (2017). The pursuit of word meanings. *Cognitive science*, 41:638–676.

Stevenson, H. W. and Weir, M. W. (1959). Variables affecting children's performance in a probability learning task. *Journal of Experimental Psychology*, 57(6):403.

Szagun, G., Steinbrink, C., Franik, M., and Stumper, B. (2006). Development of vocabulary and grammar in young german-speaking children assessed with a german language development inventory. *First Language*, 26(3):259–280.

Szagun, G., Stumper, B., Sondag, N., and Franik, M. (2007). The acquisition of gender marking by young german-speaking children: Evidence for learning guided by phonological regularities. *Journal of Child Language*, 34(3):445–471.

Tenenbaum, J. B. and Griffiths, T. L. (2001). Generalization, similarity and bayesian inference. *Behavioral and Brain Sciences*, 24(4):629–640.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., and Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *science*, 331(6022):1279–1285.

Tomasello, M. (2003). *Constructing a language*. Harvard University Press, Cambridge, MA.

Trueswell, J. C., Lin, Y., Armstrong, B., Cartmill, E. A., Goldin-Meadow, S., and Gleitman, L. R. (2016). Perceiving referential intent: Dynamics of reference in natural parent–child interactions. *Cognition*, 148:117–135.

Trueswell, J. C., Medina, T. N., Hafri, A., and Gleitman, L. R. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive psychology*, 66(1):126–156.

Ullman, M. T., Corkin, S., Coppola, M., Hickok, G. S., Growdon, J., Koroshetz, W., and Pinker, S. (1997). A neural dissociation within language: Evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. *Journal of Cognitive Neuroscience*, 9(2):266–276.

Visser, F. T. (1963). *An historical syntax of the English language*. Brill Archive.

Weir, M. W. (1964). Developmental changes in problem-solving strategies. *Psychological review*, 71(6):473.

Westfal, S. (1956). *A study in Polish morphology: The genitive singular masculine*. Mouton, 's-Gravenhage.

Wexler, K. and Culicover, P. (1980). *Formal principles of language acquisition*. MIT Press, Cambridge, MA.

Weyerts, H. and Clahsen, H. (1994). Netzwerke und symbolische regeln im spracherwerb: experimentelle ergebnisse zur entwicklung der flexionsmorphologie. *Linguistische Berichte*, 154(4):430–460.

White, L. and Genesee, F. (1996). How native is near-native? The issue of ultimate attainment in adult second language acquisition. *Second language research*, 12(3):233–265.

Wiese, R. (1996). *The phonology of German*. Clarendon, Oxford.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, 69(1):1–34.

Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, 358(6349):749–750.

Xu, F. and Pinker, S. (1995). Weird past tense forms. *Journal of Child Language*, 22(3):531–556.

Yang, C. (2002). *Knowledge and learning in natural language*. Oxford University Press, Oxford.

Yang, C. (2005). On productivity. *Linguistic Variation Yearbook*, 5(1):333–370.

Yang, C. (2016a). The linguistic origin of the next number. Manuscript, University of Pennsylvania https://ling.auf.net/lingbuzz/003824.

Yang, C. (2016b). *The price of linguistic productivity: How children learn to break rules of language*. MIT Press, Cambridge, MA.

Yang, C. (2017). Rage against the machine: Evaluation metrics in the 21st century. *Language Acquisition*, 24(2):100–125.

Yang, C. (2018). A formalist perspective on language acquisition (Target article with commentary from Theresa Biberauer, Cécile De Cat, Laura Domínguez and Jorge González Alonso, Christine Dimroth, Adele E. Goldberg, Stefan Th. Gries, Vsevolod Kapatsinski, Jeffrey Lidz and Laurel Perkins, Silvina A. Montrul, Johanne Paradis, Tom Roeper, Jason Rothman and Noam Chomsky, Caroline F. Rowland, Roumyana Slabakova, Peter Svenonius, Eva Wittenberg and Ray Jackendoff, Noriaki Yusa). *Linguistic Approaches to Bilingualism*, 8(6):665–809.

Yang, C., Lei, M., and Lee, T. H.-T. (2019). Language shapes children's understanding of number. In *The 44th Boston University Conference on Language Development*.

Yip, K. and Sussman, G. J. (1997). Sparse representations for fast, one-shot learning. In *Proceedings of the National Conference on Artificial Intelligence*, pages 521–527.

Yu, C. and Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, 18(5):414–420.

Zwicky, A. M. and Pullum, G. K. (1983). Cliticization vs. inflection: English N'T. *Language*, 59(3):502–513.