FIRST
LANGUAGE

Against stored abstractions: A radical exemplar model of language acquisition

**$SAGE

Ben Ambridge[iD]

University of Liverpool, UK; ESRC International Centre for Language and Communicative Development (LuCiD)

Abstract

The goal of this article is to make the case for a radical exemplar account of child language acquisition, under which unwitnessed forms are produced and comprehended by on-the-fly analogy across multiple stored exemplars, weighted by their degree of similarity to the target with regard to the task at hand. Across the domains of (1) word meanings, (2) morphologically inflected words, (3) *n*-grams, (4) sentence-level constructions and (5) phonetics and phonology, accounts based on independently-represented abstractions (whether formal rules or prototype categories) fail for two reasons. First, it is not possible to posit abstractions that delineate possible and impossible form; e.g. that (1) rule in pool *tables* and data *tables*, but rule out chairs, (2) rule in the past-tense forms *netted* and *bet* but rule out **setted* and **jet*, (3) rule in the bigram *f+t* but rule out (probabilistically) *v+t*, (4) rule in both *John feared Bill* and *John frightened Bill* but rule out **John laughed Bill*, (5) rule in Speaker A but rule out Speaker B as the person who produced a particular word (e.g. *Sa'urday*). Second, for each domain, empirical data provide evidence of exemplar storage that cannot be captured by putative abstractions: e.g. speakers prefer and/or show an advantage for (1) exemplar variation even within word-meaning 'category boundaries', (2) novel inflected forms that are similar to existing exemplars, (3) *n*-grams that have occurred frequently in the input, (4) SVO sentences with *he* as SUBJECT and *it* as OBJECT and (5) repeated productions of 'the same' word that are phonologically similar or, better still, identical. An exemplar account avoids an intractable lumping-or-splitting dilemma facing abstraction-based accounts and provides a unitary explanation of language acquisition across all domains; one that is consistent with models and empirical findings from the computational modelling and neuroimaging literature.

Corresponding author:
Ben Ambridge, Psychological Sciences, University of Liverpool, Eleanor Rathbone Building, Bedford St South, Liverpool L69 7ZA, UK.
Email: ben.ambridge@liverpool.ac.uk

Introduction: a rare, but misplaced, consensus in language acquisition research

Most, perhaps all, mainstream theories of child language acquisition share a common assumption: adult knowledge of language includes stored abstractions such as [VERB] [NOUN] and [SUBJECT], and language acquisition therefore involves forming and/or mastering these abstractions. The goal of the present article is to argue that this assumption is misplaced, and to present the case for a radical alternative: adult knowledge of language consists of nothing but stored exemplars and the ability to analogize across them on the fly in comprehension or production.[1] Child language acquisition therefore involves simply storing these exemplars and developing this ability.

Child language acquisition is often seen as a highly polarized field, dominated by an all-encompassing nature–nurture debate (e.g. Valian, 2014). But, when it comes to the question of stored abstractions, there is widespread agreement. While they differ on the details, all sides agree that adult speakers possess stored linguistic abstractions of some kind. Indeed, though some stored abstractions are controversial (e.g. independently represented passive or *wh*-question constructions), others – such as the category [VERB], and some kind of abstract representation of canonical word order (e.g. English [SUBJECT] [VERB] [OBJECT]) – are agreed upon by virtually all theories, with disagreements revolving solely around timing: are these abstractions present from birth (e.g. Pinker, 1989; Thornton, 2012; Wexler, 1998), formed very quickly (e.g. Gertner, Fisher, & Eisengart, 2006), or formed more gradually (e.g. Tomasello, 2003), as posited by *generativist-nativist, early-abstraction* and *usage-based-constructivist* accounts respectively? The widespread theoretical appeal of such abstractions is obvious; they offer an explanation of the fact that speakers are able to produce and comprehend sentences that they have never heard before (as in Chomksy's famous example *Colourless green ideas sleep furiously*).

But we should not be so easily seduced. In attempting to rebut these stored-abstraction accounts in favour of a radical exemplar account, I develop three lines of argument. The first holds that the apparent explanatory power offered by these stored abstractions is illusory: there is simply no way to formulate a linguistic abstraction – be it the category [VERB], SVO word order, or the word *table* – that rules *in* all the currently-unwitnessed exemplars that are permissible, but rules *out* those that are not (see Ramscar & Port, 2016, for discussion of the difficulties inherent in positing any kind of discrete units in language representation). The second line of argument is that across a wide range of acquisition domains, the evidence for storage of individual exemplars is overwhelming. Consequently, even if it were possible to formulate satisfactory abstractions – which it is not – these abstractions would merely sit alongside, rather than replace, exemplars. The third is that, for every domain, there already exists at least one computational instantiation of an exemplar model, which compares favourably with rival computational

approaches (e.g. connectionism). In many cases, the failure of rival computational approaches is instructive: just like verbal (i.e. non-computational) prototype models, they are unable to form abstract representations that are both sufficiently broad and sufficiently narrow. Visible by their absence are models that explicitly represent the stored discrete abstractions that are assumed (in some cases alongside stored exemplars) by virtually all verbal models of child language acquisition (e.g. the [SUBJECT] [VERB] [OBJECT] transitive construction). Why? They simply don't work. In developing these three arguments, I consider representations at five levels: (1) word meanings, (2) morphologically inflected words, (3) *n*-grams (primarily at the word level), (4) sentence-level constructions and (5) phonetics and phonology.

In place of stored-abstraction models, I argue for an exemplar account under which unwitnessed forms are produced and comprehended by on-the-fly analogy across multiple stored exemplars, weighted by their degree of similarity to the target *with regard to the task at hand*. For example, consider the single exemplar *Mummy kissed her*. If the task at hand is to generate the appropriate phonological form for the past-tense form for *wish*, phonological similarity is relevant (e.g. *kiss→kissed*, so *wish→wished*). If the task at hand is to express who did what to whom in a hugging scenario (e.g. out of Bill and Jane), one relevant dimension is semantic + structural similarity (e.g. if KISSER-KISS-KISSEE → *Mummy kissed her*) then HUGGER-HUG-HUGEE → *Jane hugged Bill*).

Before proceeding any further, two caveats are in order. First, while I am arguing against the view that abstractions like SUBJECT, NOUN and VERB are represented by speakers and acquired by learners, I am not denying their utility in theoretical linguistics. Such abstractions are often useful for capturing patterns and regularities both within and across languages. The mistake, I argue, is to make the leap from abstractions as useful descriptive shorthands to abstractions stored and represented by speakers and learners. To borrow an analogy from MacWhinney (2001), no adequate description of a honeycomb could fail to mention its hexagonal structure, but this structure is not represented in the brains of individual bees, or anywhere else; it is simply emergent from the process of honeycomb formation. Similarly, no adequate description of English could avoid making reference to NOUNs and VERBs but, I argue, these structures are not represented in the brains of English speakers, or anywhere else; they are simply emergent from the processes of language comprehension and production.

The second caveat is that the ideas I am setting out here are not new; the position I take with respect to the adult grammar is similar – perhaps identical – to that set out by Croft (2000, 2001) and Bybee (1985, 2010).[2] My goal in this article is to protest to my colleagues in the field of child language acquisition research that we have all (myself included) failed to take sufficient notice of this work and its implications for all domains of acquisition research, and, as a result, have been working on the basis of an incorrect, if often implicit, assumption: that acquisition involves building (or mastering innately-given) abstractions such as the [SUBJECT] [VERB] [OBJECT] transitive construction. Of course, many child language acquisition researchers have argued for – and provided empirical evidence for – the claim that rote-learned exemplars (or *holophrases* or *frozen/ formulaic phrases*) are important in both children's and adults' grammars. But almost universally, these theories go on to claim that children also form abstractions, such as Tomasello's (2003, p. 308) 'item-based construction[s] . . . *I wanna —*', 'formed as

abstractions across individual word combinations' (p. 123); Pine and Lieven's (1997, p. 132) 'slot and frame patterns' (e.g. *want a + X*) reflecting 'some kind of regularity abstracted from the input' (p. 133); Dąbrowska's (2000, p. 83) ' "lexically based patterns" or "formulas" . . . formulaic frames like *Where's___?* and *What's ___ doing?*'; Theakston, Ibbotson, Freudenthal, Lieven, and Tomasello's (2015, p. 1369) 'verb frames, such as *__want__*, *__see__*, and *__get__*'; Tomasello's (2003, p. 104) 'totally abstract construction' or – as recently as 2016 – my own 'semantic construction prototype' (Ambridge, Bidgood, Pine, Rowland, & Freudenthal, 2016, p. 1455). Even current chunk-based approaches assume that the 'language system must "eagerly" recode and compress linguistic input' (Christiansen & Chater, 2016, p. 1), that 'the resulting compressed representations . . . provide only an abstract summary of the input, from which the rich sensory input cannot be recovered'.

Now, it is possible that these *item-based constructions, slot-and-frame patterns, lexically-based patterns, abstract constructions* and so on are not meant to be taken literally, but are meant merely as metaphors or descriptors for the kind of emergent online generalizations made by an exemplar model (if so, I of course have no quarrel with these proposals[3]). But, at least in most cases, I don't think so. This is certainly not what I meant in my 2016 paper; I really meant to say that children form and store a *semantic construction prototype* (a position with which – needless to say – I now disagree completely). Similarly, the formation and storage of some kind of abstraction seems to be at the very least implicit in the idea of linguistic 'representations becoming fully abstract' (Dittmar, Abbot-Smith, Lieven, & Tomasello, 2008, p. 581), of 'the development of a more schematic and abstract inventory of conventionalised constructions' (Lieven, Salomo, & Tomasello, 2009, p. 505), in the claim that children 'construct . . . abstract syntactic representations in the course of development' (McClure, Pine, & Lieven, 2006, p. 718), that their 'linguistic representations . . . become abstract and productive' (Savage, Lieven, Theakston, & Tomasello, 2006, p. 29) or that 'the development of abstract (lexically-independent) representations occurs as the child develops more and more verb-specific patterns with experience of the language, until she eventually generalizes across them on the basis of commonalities in form and meaning' (Rowland, Chang, Ambridge, Pine, & Lieven, 2012, p. 51).[4] And, lest I be accused of straw-manning, note that the storage of abstractions above and beyond individual exemplars is entirely *explicit* in a paper co-authored by the most highly cited (by quite some distance) child language acquisition researcher, Mike Tomasello,[5] which posits a 'hybrid model comprising both abstractions and the retention of the exemplars of which those abstractions are composed' (Abbot-Smith & Tomasello, 2006, p. 276; see also Goldberg, 2006; Langacker, 1988, for linguistic approaches that posit both stored exemplars and stored abstractions).

If, as seems to me to be the case, these theories[6] really are proposing stored abstractions (e.g. a *want a + X* slot-and-frame pattern; an abstract SVO transitive construction) above and beyond individual exemplars, then I think they are mistaken, and over the next 20,000 words or so, I will do my best to explain why. Do speakers *generalize*? Yes, of course; on the fly, in the moment of comprehension or production. Do they *store* generalizations? No.

First, it will be useful to consider a highly oversimplified ('toy') example that illustrates the principles of an exemplar account, and contrasts it against rival stored-abstraction approaches of both the generativist-nativist and usage-based-constructivist variety

(formal computational implementations of an exemplar account are set out in subsequent sections). Suppose that an English-speaking 4-year-old is taught a novel verb ('Look, tamming!') to describe the motion of a toy block bouncing and spinning on a suspended rope (e.g. Brooks & Tomasello, 1999), and then describes a scene in which a doll bounces and spins on the rope by saying *She's tamming*. Stored-abstraction accounts offer the following explanations.

Generativist-nativist accounts (e.g. Wexler, 1998) assume that children are born with the syntactic categories [SUBJECT] and [VERB], and with the knowledge that these two categories can be combined to yield a phrase or sentence. Furthermore, a 4-year-old will have long ago set the SV/VS parameter (technically known as the specifier-head parameter), which determines (amongst other things) whether the ordering of these categories in the target language is SV (e.g. English) or VS (e.g. Welsh, Maori). (We ignore here the additional complication introduced by the auxiliary *is*.) This knowledge, together with the knowledge that *She* and *tamming* are respectively a SUBJECT and – based on its meaning and morphology – a VERB, allows the speaker to generate *She's tamming*.

Usage-based-constructivist accounts (e.g. Tomasello, 2003) assume that children generalize across input sentences (e.g. *She's dancing, She's playing, She's laughing*) to form, first, slot-and-frame patterns (e.g. *She's [ACTION]ing*) and, later, fully abstract sentence-level constructions (e.g. [SUBJECT] [VERB]). A 4-year-old producing *She's tamming* would normally be assumed to have generated the utterance by inserting the relevant lexical items into the abstract construction, though a much younger child (around 2;0, say) would normally be assumed to have used the lower-level slot-and-frame pattern.

Both of these accounts assume that speakers store free-standing linguistic abstractions. Indeed, although the acquisition processes are very different, the mature abstractions – [SUBJECT] [VERB] phrase structure and a [SUBJECT] [VERB] intransitive construction – are almost indistinguishable. An exemplar account, on the other hand, assumes no such stored abstractions. Rather, learners store concrete exemplars, each including the surface form along with its understood meaning and contextual details, and produce (and comprehend) novel utterances in real time by analogizing across these stored exemplars on the basis of similarity. The basis of this similarity and the precise form of the analogy depend on the particular implementation under consideration; and, as we will see later, for a few circumscribed acquisition problems, there exist fully-specified computational implementations of an exemplar account. For the purposes of the present toy example, however, it will suffice to say that the child generates *She's tamming* to express the message 'Discourse-old female undergoes spinning action' by analogy across utterances with very similar semantic/functional properties (e.g. 'Discourse-old female undergoes bouncing action' = *She's bouncing*; 'Discourse-old female undergoes spinning action' = *She's spinning*), more distant analogy across utterances with somewhat similar semantic/functional properties (e.g. 'Discourse-new female undergoes spinning action' = *Sue's spinning*; 'Discourse-old male undergoes bouncing action' = *He's spinning*), and the utterance in which the novel verb was trained ('Toy block undergoes bouncing+spinning action' = *Look, tamming!*), as summarized in Table 1.

It is important to stress from the outset that, as this toy example illustrates, an exemplar account does not posit that speakers are restricted to producing words, phrases or sentences from a fixed repertoire: at all levels, novel combinations can be generated by

Table 1. Production of the novel utterance *She's tamming* by analogy across stored exemplars with similar semantic/functional properties ('message') to the target.

Stored message	Stored utterance
Toy block undergoes bouncing+spinning action	*Look, tamming!*
Discourse-old male undergoes bouncing action	*He's bouncing*
Discourse-new female undergoes spinning action	*Sue's spinning*
Discourse-old female undergoes bouncing action	*She's bouncing*
Discourse-old female undergoes spinning action	*She's spinning*
Target: Discourse-old female undergoes spinning action	**Output: *She's tamming***

analogy as soon as the learner has stored, in principle, a single relevant exemplar[7] (e.g. *go*→*goed by phonological analogy with *show*→*showed*). Hence, an exemplar model (at least in some possible implementations) actually predicts earlier comprehension and production of novel forms than a constructivist account under which abstractions (e.g. the [SUBJECT] [VERB] construction discussed above) are acquired only when children reach a critical mass of exemplars across which to generalize (e.g. Marchman & Bates, 1994).

It is also important to stress from the outset that an exemplar approach does not entail abandonment of constituent structure. An important insight captured by traditional generative linguistics is that linguistic generalizations operate on constituents, rather than particular words. For example, whatever generalization one posits for forming a question with *Bill* must also apply for *the man, the man who is tall* and so on (e.g. Chomsky, 1980; Crain & Nakayama, 1987):

Bill is happy → Is Bill happy?

The man is happy → Is the man happy?

The man who's tall is happy → Is the man who's tall happy?

Generativist and constructivist accounts alike capture constituent structure by positing the existence of stored abstractions like [SUBJECT] or [NOUN PHRASE] (e.g. *Bill; the man; the man who is tall*). Exemplar accounts do not posit stored abstractions; rather, constituent structure is inherent in the exemplars across which analogies are drawn. The same analogical processes that allow the system to analogize across semantically-similar utterances with *He, She and Sue* (Table 1) allow it to analogize across semantically-similar utterances with *Bill, the man* and *the man who's tall*.

Specifying exactly what is meant by 'similarity' is not straightforward. In fact, semantic similarity alone will almost certainly not suffice. For example, while a phrase like *the man* shares clear semantic similarity to *Bill, the man who's tall*, and so on, it also patterns similarly to semantically-unrelated, abstract phrases (e.g. *the situation*). Yet, although not semantically similar, *the man* and *the situation* share a degree of functional similarity

in that similar properties, actions and events can be predicated of them (e.g. *the man/the situation is bad; the man/the situation made us sad*). The analogical processes assumed by an exemplar account (and any computational instantiation thereof) will therefore need to take account of this type of functional-semantic similarity (we will meet one such model in the section 'Sentence-level constructions'). Although specifying the dimensions of similarity is extremely challenging, this challenge is not unique to exemplar accounts. Accounts which assume that children form stored abstractions by generalizing across the input must also specify the basis of the similarity across which these generalizations are formed. Generativist-nativist accounts are exempt from this challenge, since they assume the relevant abstractions are innately given. However, they face the potentially even more daunting challenge of explaining how these abstractions are linked to the language that learners actually hear (Ambridge, Pine, & Lieven, 2014).

Word meanings

Probably the simplest form of abstraction that is posited by most current theories of language acquisition is the monomorphemic word (e.g. *table*). Indeed, this abstraction is – it would seem – so simple and self-evident that it is easy to forget that it is an abstraction at all. But, of course, an abstraction it is. At the phonetic level, the idealized form *table* is an abstraction across all the different pronunciations that have been witnessed (e.g. by different speakers, by the same speaker on different occasions, and so on). Discussion of this level of abstraction will be generally reserved for the later section on 'Phonetics and phonology'. Here, we focus mainly on the semantic level, where the word *table* is an abstraction that 'maps onto' or somehow 'stands for' some entities in the physical world, other entities that share certain salient properties (e.g. the geological water table), photographs of those entities (or renderings in abstract art), toy versions for use in a doll's house, a talking table in a children's cartoon, and so on.

Notice that no distinction is drawn here between learning *table* the *word*, and *table* the *concept*. This is because there is no meaningful distinction to draw (Ramscar & Port, 2015). As the examples show, the wide variety of entities that English speakers refer to as *tables* share no defining characteristic, other than being referred to as *tables* by English speakers (even if we set aside more obviously problematic cases like data *tables* or multiplication *tables*). Already, then, the cracks in the idea of an abstract meaning representation for *table* are starting to show. There is no way to define the abstraction that rules in everything that an English speaker could conceivably refer to as a *table*, and rules out everything that she could not.

Cognitive psychologists abandoned long ago (e.g. Rosch & Mervis, 1975) the idea that categories in the world are 'rule-based' (Smoke, 1932) (although of course humans are *capable* of making rule-based categorizations in experiments or, for example, the legal system). Any conceivable rules for defining a *table* (e.g. 'has legs'; 'used for eating'; 'made of wood, metal or plastic'; 'waist height') can be easily dismissed with counterexamples (e.g. an empty beer barrel used as a table at a bar; a pool table; an origami paper table glued to the ceiling as part of an art exhibit, and so on). And even if we could define a *table* in these terms (e.g. 'has legs and a flat surface'), this only shifts the problem elsewhere (how do we define a *leg*, a *surface, flat*?; Ramscar & Port, 2015).

A *prototype category* meaning for *table* fares better, but is still highly problematic. On this view, speakers average across every inferred referent of *table* that they have encountered to form an abstract, fuzzy, probabilistic representation of the category labelled by the word *table* (in the same way that they are argued to abstract across instances of VERBs or SVO constructions to form other linguistic abstractions). New uses of *table* are interpreted (in comprehension) or coined (in production) with reference to this abstraction; i.e. people are faster and more willing to accept novel, previously unseen items as *table*s if they are very similar to the *table* prototype. Indeed, a non-linguistic categorization study with dot patterns (Posner & Keele, 1970) went further in demonstrating that prototypical but novel patterns were more likely to be assigned to the relevant category than were patterns that were less typical of the category (though still consistent with it), and that had been shown during training.

However, the prototype category approach suffers from three problems. The first is an empirical problem. In laboratory categorization studies, if the stimuli to be categorized are matched on prototypicality, participants show an advantage for previously-seen items (e.g. Zaki, Nosofsky, Stanton, & Cohen, 2003). This is a problem for prototype models, which store only the prototype, not the individual items. Furthermore, prototype effects for unseen items, of the type demonstrated by Posner and Keele (1970), are also yielded by all but the very simplest type of exemplar model (e.g. Mack, Preston, & Love, 2013; Medin & Schaffer, 1978; see Smith & Minda, 2000; Zaki et al., 2003, for reviews). This is because a prototypical item – even one that has not been seen before – is highly similar to a large number of stored exemplars. Consequently, exemplar models that analogize to new items on the basis of similarity across a number of stored exemplars (e.g. *k*-nearest-neighbour models) yield prototypicality effects for unseen items, without a stored representation of the prototype. (A maximally simple exemplar model that analogizes on the basis of the *single* nearest neighbour cannot yield this effect.) In sum, as Love (2013, p. 348) puts it, 'By and large, exemplar models can mimic all the behaviors of prototype models, but the opposite is not true', including – as we will see in more detail later – when accounting for fMRI data obtained from participants performing categorization tasks (e.g. Mack et al., 2013; Nosofsky, Little, & James, 2012). It is therefore surprising that virtually all accounts of word learning (and child language acquisition more generally) have – whether implicitly or explicitly – opted for the latter.

The second is an in-principle problem. For real-word putative word categories like *table* (as opposed to dot-pattern or abstract-shape categories in a lab-based classification task), there is no way to define the prototype: do speakers just have a single prototype meaning for *table*, that includes domestic dining tables, beer-barrel bar tables and fold-down aeroplane tables (lumping), or a separate prototype for each (splitting)? The lumping approach is unworkable because some (would-be) categories have internal structure. For example, spoons are generally small and metal or large and wooden; but nobody would define a prototypical spoon as one that is of intermediate size and made out of an intermediate wood-metal material (example from Love, 2013). The splitting approach is unworkable, because there is no principled way to stop splitting. Do we have a single prototype of a domestic dining table, or subtypes of wooden and metal tables, or of vintage and modern tables (or, for that matter, of data tables and of multiplication tables)?[8]

The third problem for prototype approaches is an in-practice problem. Prototype categories might be useful for dot-pattern classification tasks, but they are useless in the real world. Suppose a listener is asked if a (beer-barrel bar-room) *table* is free, or to put away her (aeroplane tray) *table*. The likelihood of communicative success is determined not by the extent to which a general prototype invoked for the listener by the word *table* overlaps with the tables present (which is probably not by much). Rather, it is determined by the listener's inference regarding the speaker's most likely meaning of *table, in that particular context*. Which dimensions are relevant when deciding whether or not something can be called a *table* are not fixed, but depend on the speaker's goals. If the relevant dimension is function – I need somewhere to put my glass – I am quite happy to call an upturned beer barrel a *table*. If the relevant dimensions are aesthetic, for example when looking at a picture in a gallery, I am quite happy to refer to a particular arrangement of paint on canvas as a *table*; but I would not put my drink on it.

In an important sense, the three problems are just different ways of saying the same thing: a word like *table* does not have a standalone, prototypical or central meaning that is devoid of context. For any individual language learner, a word like *table* has as many different meanings as the leaner has been in situations in which she has interpreted a speaker's meaning of *table*. That is, for any individual language learner, a word like *table* has not one meaning, but thousands – one for each *table* that she has encountered.[9] This is what an exemplar account of word meaning looks like.

An advantage of an exemplar account of word meaning is that it sidesteps altogether a problem that has given rise to a whole sub-discipline of language acquisition research: how children disambiguate homophones (e.g. river *bank* vs money *bank*), particularly noun/verb homophones (e.g. John likes *fish*; John can *fish*; Pinker, 1987) (see e.g. Conwell, 2018, for a review). But the problem only arises because researchers have assumed (in most cases presumably implicitly) a stored-abstraction, prototype-based model of word meaning, under which a word like *bank, fish* or *table* 'should' have a single meaning. The problem simply does not arise assuming an exemplar model, under which every heard exemplar of the word *bank, fish* or *table* is stored with its meaning, *as understood in that individual situation* (e.g. Elman, 2009; Erk & Padó, 2010). An exemplar model also avoids drawing an unprincipled distinction between 'true' homophones (e.g. river *bank* vs money *bank*) and extensions of a 'prototypical sense'. Is a water *table* the 'same kind' of table as a bar-room beer-barrel *table* and a journal article results *table*? Who can say? The advantage of an exemplar model is that we don't have to. *Table* has not one meaning, or two meanings that somehow require disambiguation (e.g. *table* the noun and *table* the verb), but thousands.

Regier's (2005) Lexicon as Exemplar (LEX) model is a computational model that has most of the properties of the informal exemplar account of word-meaning acquisition that I sketched above, including (a) exemplar storage at the level of both form and meaning and (b) the lack of stored abstractions at the level of either form (i.e. 'words') or meaning (i.e. 'concepts'). The model's task is to predict a form given a meaning (in production) or vice versa (in comprehension). A form (e.g. the word *bat*) is presented to the model as a 50-bit vector; i.e. a string of 50 1s and 0s, each – in principle – denoting the presence or absence of some phonetic/phonological feature of the form. In practice, since LEX uses an artificial training set, these features do not correspond to anything at

all. But in a real-world example, they correspond to phonetic/phonological features that are relevant for distinguishing individual words (e.g. voicing, *bat* vs *pat*) and features that are irrelevant (e.g. the pitch [fundamental frequency]) with which the exemplar was spoken (25 of each type).[10] Alongside each form, a meaning is presented to the model, also as a 50-bit vector. These 50 features are again evenly split between those that are relevant for discriminating between different meanings (e.g. size, colour, shape) and those that are not (e.g. time of day, whether or not the speaker is wearing glasses).

When a form–meaning pair is presented to the model, new 50-bit nodes are created on a hidden form-exemplar layer and a hidden meaning-exemplar layer. The model learns by modifying the associative weights between the nodes on these hidden layers. Importantly, when a form–meaning pair is presented, the model not only creates these nodes, but also strengthens the associative weights between existing form exemplars and meaning exemplars that have sufficient overlap with the current form–meaning pair (to the extent that analogy is happening at the moment of encoding rather than retrieval, this assumption represents something of a departure from a 'pure' exemplar model). When calculating overlap, the model does not weight all features equally. Rather, it learns which of the form features are and are not useful for predicting meaning features and vice versa, and stores this information in form and meaning attention weights, which tell the model which features to weight most strongly when calculating overlap. These attention weights allow LEX to simulate a phenomenon whereby children show a tendency to generalize object names to new objects with the same shape (and/or same function; e.g. Diesendruck, Markson, & Bloom, 2003), rather than (for example) colour (e.g. Landau, Smith, & Jones, 1988).

LEX also simulates three more important phenomena observed in word-learning research. First, after an early period during which children require many exemplars to learn a novel word, they show *fast-mapping* (e.g. Carey & Bartlett, 1978), learning after just one or two exemplars. LEX simulates this phenomenon because, at the early stages, similar exemplars are in competition. A considerable number of input pairs are then required for the model to learn which of the form features are predictive of meaning features, and vice versa. Second, young children struggle to learn similar-sounding words (e.g. *bat* vs *pat*), whereas older children do not (e.g. Stager & Werker, 1997). Again, LEX simulates this phenomenon, because after an early stage in which exemplars with similar form features are in competition, it learns – on the basis of more input exemplars – to attend to the form features that discriminate these words (here, voicing) over those that do not (e.g. pitch). Third, children show an early stage of *mutual-exclusivity*, during which they are unable to learn a second label for an object, before subsequently being able to do so (e.g. Liittschwager & Markman, 1994). LEX simulates this phenomenon because any given set of meaning exemplars is already linked to a set of form exemplars, making it more difficult to associate that set of meaning exemplars with a new set of form exemplars. LEX's success in simulating these phenomena is a result of competition between (or, for the shape bias, recognizing similarity across) individual exemplars in memory, and so could not be captured by non-exemplar models that collapse individual exemplars of 'the same' form or meaning.

LEX also has two further desirable features that are not mentioned by Regier (2005). First, although no such test is reported, LEX would presumably be easily capable of

learning homophones, since there is nothing to stop two similar sets of form features being associated with two dissimilar sets of meaning features. As for the mutual-exclusivity effect, competition would make these homophones harder for the model to learn – particularly lower frequency ones – exactly as we see for human learners (e.g. Rubenstein, Lewis, & Rubenstein, 1971). Second, as we will see in more detail in the section on 'Phonetics and phonology', learners retain considerable phonetic detail that allows them to recall (for example) the identity of the speaker of a particular word. Again, although no such test is reported, it would presumably be easy to simulate this phenomenon in LEX by adjusting its attention weights to focus on features that discriminate speakers (e.g. pitch) rather than meaning (e.g. voicing). This reflects a more general advantage of exemplar models: the dimensions along which similarity is computed are not fixed, as they have to be under a prototype model (else there is no way to define the prototype). Rather, the dimensions along which similarity is computed – on the fly in real time – depend on the task at hand (as in Barsalou's, 1983, p. 4 ad hoc categories, such as 'things on my desk with which I can pound in a nail'). As we will see in more detail in the final section, we have no idea what properties will turn out to be relevant for some future goal.

Morphologically inflected words

English, as a language with relatively impoverished inflectional morphology, marks verbs for tense (e.g. *plays* [present] vs *played* [past]) and person/number (*She plays* [3rd person singular]/*They play* [3rd person plural]), and marks nouns for number (e.g. *one dog* [singular], *two dogs* [plural]). Many languages have much richer systems of inflectional morphology (e.g. noun and verb marking in languages such as Polish and Finnish, which we will consider shortly).

We begin, however, with a system that has attracted a great deal of research attention, due, in part, to its apparent simplicity: English past-tense marking. Setting aside, for a moment, irregular verbs (e.g. *sing/sang*), regular forms appear, at first blush, to be created by a 'regular past-tense rule, which adds the suffix *-ed* to the end of a verb to indicate that the event referred to by the verb took place before the speech act (e.g. *walk-walked*)' (Prasada & Pinker, 1993, p. 2). This rule, which can be summarized informally as *[VERB]+ed*, is a classic example of the type of abstraction posited by abstraction-based theories; one that is 'capable of operating on any verb, regardless of its sound' and that therefore 'affords unlimited productivity' (Prasada & Pinker, 1993, p. 2). Although this particular formulation sits squarely within a *generativist-nativist* framework (and corresponds to a *rule-based* model of categorization), *usage-based-constructivist* accounts also discuss, at least informally, 'the formation of a . . . schema for regular inflection' (Maslen, Theakston, Lieven, & Tomasello, 2004, pp. 1332–1333), which would seem to correspond to a *prototype* model.

Whether one formulates *[VERB]+ed* as an abstract rule or an abstract prototype, it is again impossible to formulate the abstraction in such a way as to account for descriptive facts about the system (i.e. to rule in all, and only, possible forms), let alone empirical data from studies with adults and children. With regard to descriptive facts about the system, apparent 'regular *-ed*' forms are not all created equal, but cluster into what Albright and Hayes (2003, p. 127) call phonological 'islands of reliability'. For example,

every English verb that ends in a voiceless fricative (*f, th, s* or *sh*) has a past-tense form ending in *-t* (e.g. *missed, hissed, wished*). Similarly, except for irregulars, verbs that end in *-t* or *-d* have a past-tense form ending in *schwa+d* (e.g. *tended, needed, voted*), while those that end in *-b, -g* or *-n* have a past-tense form ending simply in *-d* (e.g. *rubbed, sagged, planned*). This is clearly a problem for a *generativist-nativist* abstract rule 'capable of operating on any verb, regardless of its sound' (Prasada & Pinker, 1993, p. 2), but it is equally problematic for a *usage-based-constructivist* abstract 'schema for regular inflection', since the system requires not one schema but at least three (we are back to the *lumping or splitting* problem raised in the previous section).

With regard to empirical data, acceptability judgement and production studies with adults and children (Albright & Hayes, 2003; Ambridge, 2010; Blything, Ambridge, & Lieven, 2018) demonstrate that both the acceptability and production probability of 'regular' past-tense forms for a given novel verb (e.g. *wiss, bredged, chooled, daped*) are predicted by the verb's phonological distance from existing stored 'regular' past-tense forms (e.g. *wissed* is similar to *missed, hissed* and *wished; bredged* to *wedged*, and so on); and likewise for 'irregulars' (e.g. *flept* is similar to *slept, wept* and *crept*).[11] These findings alone are enough to rule out accounts that posit a regular rule (or regular schema), but what should we put in its place? There is no shortage of options: the literature boasts three broad categories of exemplar model, at least 13 connectionist (neural network) models (see Kirov & Cotterell, 2018 for a summary) and the multiple-rules model of Albright and Hayes (2003).[12]

Because exemplar models originate in the cognitive psychology literature on categorization (e.g. the dot-pattern classification studies discussed earlier), the problem must be set up as a classification task. All three classes of exemplar model discussed here work in the same basic way: they store present-tense/stem forms along with some kind of marker denoting the category of the past-tense form (e.g. Keuleers, 2008, p. 41, 'transformation labels'). For example, *sing* is classified as an $i{\rightarrow}a$ verb (along with *ring-rang, sit-sat, spit-spat, swim-swam*, etc.), and not as $i{\rightarrow}u$ verb (e.g. *cling-clung, dig-dug, sting-stung*), a $+d$ verb (e.g. *moved, rained, plunged*), a $+t$ verb (e.g. *miss-missed, hiss-hissed, wish-wished*), and so on. At test, the model is given a novel stem (e.g. *spling*) and asked to predict its class (or to produce a probability distribution over all classes), which allows the model to be evaluated directly against human judgement and production data. The three varieties of exemplar model differ in exactly how this is done.

Nosofsky's (1990) Generalized Context Model (GCM) compares the test item (e.g. *spling*) feature-by-feature to each member of a given class (e.g. for the $i{\rightarrow}a$ class to *ring, sit, spit, swim*, etc.) and sums the number of shared features (e.g. Table 2)

It then compares the test item feature-by-feature to *all* stored exemplars (regardless of class) and again sums the number of shared features. It then divides the first figure by the second, to yield the probability that the test item is a member of that class. Exactly what the features are (e.g. phonemes vs phonetic features) and how they are aligned differs from instantiation to instantiation (the example in Table 2 is based on Keuleers, 2008), and – to complicate matters – does affect performance (Eddington, 2004). For example, *spling* and *swim* do not share a final feature in a phoneme representation, but do so in a phonetic-feature representation which represents both /ŋ/ and /m/ as nasals). Importantly many versions of the GCM use feature weighting (like the attention weights in Regier,

Table 2. Feature comparison in an exemplar model of English past-tense morphology.

Spling	s	p	l	ɪ	ŋ
Ring	r	r	r	ɪ	ŋ
	0	0	0	1	1
Sit	s	s	s	ɪ	t
	1	0	0	1	0
Spit	s	p	ɪ	ɪ	t
	1	1	0	1	0
Swim	s	w	ɪ	ɪ	m
	1	0	0	1	0
Sum = 9	3	1	0	4	1

2005, LEX model) in order to capture the intuition that (for example) later features in the stem are more predictive of past-tense forms. Most versions of the GCM use a decay function, such that the influence of exemplars decreases exponentially with distance (in terms of shared features) from the target. Most also include a memory-strength function, which usually corresponds to the token frequency of each item (here, verb form) in the input; unlike, for example, Regier's (2005) LEX, the GCM considers instances of a verb form (e.g. *swim*) produced with – for example – different pitch to be the same exemplar.

The Tilburg Memory Based Learner (TiMBL; Daelemans & van den Bosch, 2010; Keuleers, 2008) is similar in its use of a form of feature weighting (information gain) and a decay function. It differs in that it considers not all stored exemplars, just *k* nearest neighbours (set as a model parameter) and, in most implementations, does not represent token frequency in any way (though the English past-tense model of Van Noord & Spenader, 2015, is an exception). Skousen's (1989) Analogical Model (AM) does not use feature weighting (arguing that weights are unprincipled since they vary depending on the precise makeup of the corpus) or a memory-strength function, though it does instantiate imperfect memory such that any exemplar has only a 50% chance of being recalled on any test trial. Like TiMBL, AM does not represent token frequency (Skousen, 1989, p. 54 argues that although, in principle, token frequency should be represented, adding this information seems only to harm the model's performance; see Eddington, 2004, for evidence[13]). When classifying a target form, AM considers all exemplars (except those forgotten due to imperfect memory), with no decay function. It identifies all exemplars that share features with the target, but selects for the final analogical set only those that do not increase uncertainty regarding classification. For example, if the target is the novel verb *chool* (from Albright & Hayes, 2003), the analogical set contains *choose* (→*chose*) and *chew* (→*chewed*), which narrow down the choice of classification (to either *chool*→*chole* or *chool*→*chooled*). It does not include, for examples, *cheat, check, cheer, poop, puke* or *boot* because, although each shares one or more feature with the target, they serve only to increase uncertainty regarding classification.

Why all the detail? Two reasons. First, while connectionist models of the English past tense (and inflectional morphology more generally) are well known amongst child

language acquisition researchers, most – at least in my experience – seem to be largely unaware of exemplar models and how they work. Second, the details matter; not so much the choice of GCM, TiMBL or AM *per se*, none of which consistently outperforms the others (see Chandler, 2017, for a review), but the various implementational decisions that must be taken regarding feature representation, feature weighting, decay functions, the use of types versus tokens, and so on. When given the task of predicting the judgement and production data from Albright and Hayes' (2003) novel verb study, each of these exemplar models – depending on the particular instantiation – equals or betters both a state-of-the-art connectionist model (cf. Chandler, 2010, Table 1; Kirov & Cotterell, 2018, Table 5; though see Corkery, Matusevych, & Goldwater, 2019, for concerns regarding the stability of these simulations) and Albright and Hayes' (2003) own model which constructs an explicit micro-rule for each and every sub regularity; an approach which shows no regard for psychological plausibility, in contrast to many exemplar models which have their origins in models and findings from the non-linguistic categorization literature.

Although the English past tense has attracted particular attention, exemplar models have been used to successfully simulate noun plural marking in German (Daelemans, 2002; Hahn & Nakisa, 2000), Arabic (Dawdy-Hesterberg & Pierrehumbert, 2014; see also Ravid & Farah, 1999, for a study with children) and Dutch (Keuleers et al., 2007), novel noun-compound formation in Dutch and German (Krott, Schreuder, & Baayen, 2002; Krott, Schreuder, Baayen, & Dressler, 2007) and word stress in Dutch, including errors typically made by children (Gillis, Daelemans, & Durieux, 2000). Indeed, across morphological systems, when connectionist and exemplar models are compared on a level playing field (i.e. on the same target data, with the same representations), exemplar models usually show better performance (e.g. Mudrow, 2002, for Finnish past tense and Danish noun compounds; Nakisa, Plunkett, & Hahn, 2000, for the English past tense and Arabic plurals).

Why? The explanation is technical but bear with me, as it gets right to the heart of the distinction between exemplar models and connectionist models (which, in many respects, are computational instantiations of verbal prototype models; e.g. Chandler, 2002). As we saw for the *spoon* example in the section 'Word meanings' (from Love, 2013), prototype models cannot account for the learning of a category that is non-linearly separable. For example, suppose a slightly simplified world (not so very different to this one) in which all spoons are either large and wooden or small and metal (see Figure 1). A prototype model cannot represent the category *spoon*, because there is no way to draw a single straight line that separates spoons from things that are not spoons (go ahead, try).

Many linguistic systems also contain non-linearly-separable distinctions. For example, a prototype model of the English past tense cannot represent a category of verbs with stems ending *-et* because – in terms of their past-tense forms – there is no way to draw a single straight line that separates *-et* verbs from things that are not *-et* verbs (see Figure 2). Simple two-layer models with no hidden layer (including Rumelhart & McClelland, 1986, past-tense model) cannot form non-linearly-separable categories (Minsky & Papert, 1969). Such a model cannot form an *-et* verb prototype because – just like our prototype spoon made of a material in between wood and metal – it would have to have a past-tense form in between a no-change form and an /Id/ form. Thus, if it is presented
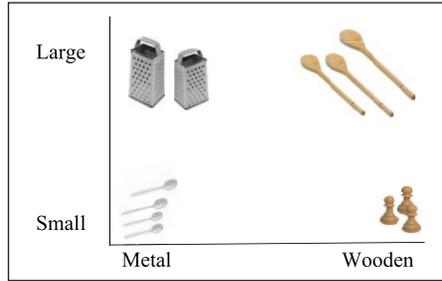
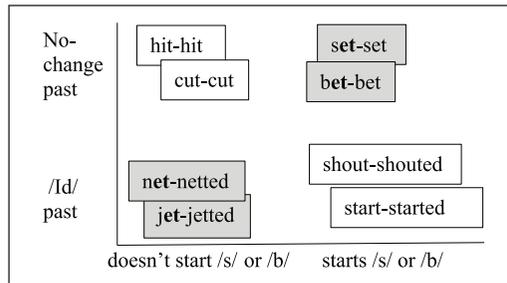**Figure 1.** Spoon as a non-linearly-separable category.

Figure 2. *-et* verbs as a non-linearly-separable category.

with a novel *-et* verb, the model does not know whether to treat it like *set* and *bet* or *net* and *jet*.

Famously, connectionist models with one or more layers of hidden units[14] can learn non-linearly-separable categories, and so show good performance on past-tense learning tasks. What is less well known (but was noted by Rumelhart & McClelland, 1986, p. 210), is that a connectionist model can learn non-linearly-separable categories only if it has at least one hidden unit for each member of the competing categories. But, as noted by Chandler (2002, p. 63), this means that 'the connectionist model must become a de facto exemplar-based model because it must create and maintain a unique representation for each different item presented to it during training'.

Although, up to this point, I have focused on computational modelling, there exists considerable evidence in the child-language literature against accounts based on symbolic rules and in favour of exemplar (and connectionist) accounts (I am not aware of a study that compares these two approaches on their ability to explain data from children). Given that rule-based accounts fail to account for one of the simplest apparent abstractions one could imagine (*[VERB]+ed*), we should not be surprised to learn that they fail for more complex systems, such as systems of verb and noun inflection in languages such as Polish, Finnish, Estonian and Lithuanian. Indeed, for such systems, it is not clear what rules would be possible even in principle, given that the correct 'ending'[15] varies not only across the verb person/number and noun case-marking paradigms, but

according to properties such as gender, conjugation/declension class and phonological properties of the NOUN or VERB 'stem'. A lumping approach does not work, because a generalization such as *[NOUN]+[CASE MARKER]* is at far too high a level of abstraction to explain anything about the system (it is not the case that *any* individual case-marking morpheme can be applied to *any* noun), and furthermore is inaccurate for systems that incorporate changes to the NOUN (or VERB) 'stem'. A splitting approach does not work because, as usual, once you've started splitting, you can't stop. Grammar books for such languages typically list around five conjugation/declension classes that capture broad generalizations, but split many of them into subclasses, some with just a handful of members. For example, Räsänen, Ambridge, and Pine (2016) note that although the Finnish verb paradigm *lacks* conjugation classes *per se*, phenomena such as vowel insertion, vowel harmony and consonant gradation result in descriptive schemes of verb inflection that posit as many as 46 different phonologically-based classes.

We should therefore not be surprised to learn that studies of both verb and noun morphology (Aguado-Orea & Pine, 2015; Dąbrowska, 2004, 2008; Dąbrowska & Szczerbinski, 2006; Engelmann et al., 2019; Granlund et al., 2019; Kirjavainen, Nikolaev, & Kidd, 2012; Kjærbæk, dePont Christensen, & Basbøll, 2014; Krajewski, Theakston, Lieven, & Tomasello, 2011; Kunnari et al., 2011; Leonard, Caselli, & Devescovi, 2002; Maratsos, 2000; Maslen et al., 2004; Räsänen et al., 2016; Rubino & Pine, 1998; Saviciute, Ambridge, & Pine, 2018) yield three findings that constitute evidence for an exemplar (or connectionist) account. The first is an effect of phonological neighbourhood density: the greater the number of phonological 'friends' or 'neighbours' – forms that are phonologically similar to the target and that take the same inflectional ending – the greater the rate at which children produce the target form correctly, and the lower the error rate. The second is an effect of token frequency of the individual target form: the greater the frequency of a particular 'ready inflected' verb or noun form in the input, the greater the rate at which children produce it correctly, and the lower the error rate, in both naturalistic and experimental contexts. Admittedly, most current exemplar models do not incorporate a role for token frequency. But this is only an in-practice problem, not an in-principle one. In my view, a true exemplar model would incorporate token frequency by having each individual exemplar vary slightly on certain features of pronunciation and/or meaning (as in Regier, 2005, LEX model). This would allow exemplar models to explain not only frequency effects, but effects observed in the domains of learning word meanings and phonetics and phonology; e.g. the ability to identify different speakers' pronunciations of the same word. The third is an effect of competition: when children do not produce the correct target form in an experimental study, they generally produce either a more frequent form of the target word (e.g. a 3pl verb form in place of 3sg; an accusative or genitive noun form in place of dative or instrumental), or overgeneralize a higher frequency ending from a different conjugation/declension class.

It is important to note that these frequency effects at the levels of both target and competing forms are not found only in very complex systems, where children have no 'choice' but to store a multiplicity of individual exemplars. They are found also for systems that are virtually exceptionless, such as English 3sg *-s* marking (Räsänen, Ambridge, & Pine, 2014) and Japanese past/non-past marking (Tatsumi, Ambridge, & Pine, 2018), where there is no 'need' to store individual ready-inflected forms (e.g. *fits,*

plays, runs, walks) at all, since all could in principle be generated from the bare/non-finite form (e.g. *fit, play, run, walk*). They are also found in studies of infants' production of monomorphemic single words, which require no 'inflection' at all (e.g. Sosa & Stoel-Gammon, 2012).

These findings suggest that there really is no alternative to some form of exemplar account that posits no additional stored abstractions such as English past tense *[VERB] ed* or plural *[NOUN]s*. The only debate that we should be having is exactly what type of exemplar account is correct.

N-grams

At the word level, bigrams are two-word sequences, trigrams are three-word sequences and, generally, *n*-grams are *n*-word sequences. For example, the utterance *You have another cookie right on the table* (spoken to Brown, 1973, Eve in her first recording) includes the bigrams *you+have, have+another, another+cookie* (etc.), and the trigrams *you+have+another, another+cookie+right* (etc.). As we will see in more detail below, speakers show effects of *n*-gram frequency in production and comprehension. Yet, unlike the structures discussed in the other sections, mainstream abstraction-based accounts of child language acquisition – on both sides of the Chomskyan theoretical divide – include little-to-no role for *n*-grams, at least in the adult system.

Accounts in the generativist, Chomskyan tradition explicitly reject the notion of linguistic rules or representations based on 'strings of words, rather than on their structural representations' or that 'mention only linear relations' (Crain & Nakayama, 1987, p. 522). These accounts do, of course, allow for combinations of *syntactic constituents*, with the order specified by the setting of a head-direction parameter. For example, the DETERMINTER PHRASE (DP) *another cookie* would be formed by combining the DETERMINER *another* and the NOUN PHRASE *cookie* (e.g. Abney, 1987). But the bigram *another+cookie* would not itself be stored; let alone a bigram such as *cookie+right* that violates constituent structure.

Accounts in the constructivist tradition (e.g. Tomasello, 2003) do assume that children store strings such as *another+cookie* (though it is less clear whether they assume that children also store strings such as *cookie+right* that do not constitute a semantic, functional or communicative 'unit'). But, they also assume that – with the exception of high-frequency frozen-phrases or idioms (e.g. *I+dunno*; Bybee & Scheibman, 1999) – stored strings are characteristic of an early rote-learned stage, and are largely replaced, in the adult system, by abstract patterns formed by analogy. For example, a learner might analogize across stored strings such as *the+cat, a+dog* and *another+cookie* to eventually form an abstract *[DETERMINER] [NOUN]* construction, which can be used to form any (semantically appropriate) combination; and that, therefore – in its mature adult state – is virtually indistinguishable from a generativist style DETERMINER PHRASE.

The claim that at least some of these early stored strings are replaced and effaced by these later abstractions is rarely made explicit. But – to repeat just some of the quotations I highlighted in the Introduction – it is not clear what else could be meant by the idea of linguistic 'representations becoming fully abstract' (Dittmar, Abbot-Smith, Lieven & Tomasello, 2008, p. 581), of 'the development of a more schematic and abstract

inventory of conventionalised constructions' (Lieven et al., 2009, p. 505), by the claim that children 'construct . . . abstract syntactic representations in the course of develop-ment' (McClure et al., 2006, p. 718), or that their 'linguistic representations . . . become abstract and productive' (Savage et al., 2006, p. 29).

In sum, with one exception – Abbot-Smith and Tomasello (2006), discussed in detail below – generativist and constructivist accounts alike assume that at least some adult knowledge of language – probably the majority – consists of knowledge of abstract rules, syntactic structures or constructions, rather than individual exemplars, strings or *n*-grams. Therefore, any evidence that children and adults show storage of individual *n*-gram exemplars would be problematic for all abstraction-based accounts of language acquisi-tion. Indeed, *n*-grams constitute a particularly serious problem for these accounts. Unlike the abstract morphological and sentence-level constructions discussed in the previous and following sections, it is not clear what many abstract *n*-gram-level constructions would even look like. *Some* possible abstract two-element constructions would at least be meaningful in terms of the grammar (e.g. *[DETERMINER] [NOUN PHRASE]*). But what kind of abstract *n*-gram construction could encode knowledge of a bigram or tri-gram such as *cookie+right* or *cookie+right+on*? Of course, the storage of such *n*-grams does not, on its own, constitute evidence against the additional storage of some abstrac-tions. But even if it were somehow possible to formulate satisfactory abstractions, it is not clear that they add any explanatory power, given the considerable evidence for knowledge of *n*-grams.

This evidence takes the form of studies show faster processing and/or fewer produc-tion errors for higher than lower frequency *n*-grams for both adults (e.g. Arnon & Snider, 2010; Bybee & Scheibman, 1999; Janssen & Barber, 2012; Jurafsky, Bell, Gregory, & Raymond, 2001; Krug, 1998; Lieberman, 1963; McDonald & Shillcock, 2003; Pluymaekers, Ernestus, & Baayen, 2005; Siyanova-Chanturia, Conklin, & van Heuven, 2011; Sosa & MacFarlane, 2002; Tremblay & Baayen, 2010) and children (Arnon & Clark, 2011; Bannard & Matthews, 2008; Matthews & Bannard, 2010). It is important to note that the *n*-grams in these studies were not part of high-frequency idioms or frozen-phrases in the adult grammar (which have to be stored under any account); neither were they necessarily syntactic constituents or meaningful chunks (e.g. *got any; your truck; of milk*; from Bannard & Matthews, 2008).

The most straightforward way to account for these findings is simply to assume that learners store wholesale strings that they hear (e.g. *You have another cookie right on the table*), paired with their meanings; that is to assume an exemplar model under which learners store each and every individual utterance, even those that are 'the same' at some abstract level (e.g. two instances of *You have another cookie right on the table* produced on different occasions, and/or by different speakers). This assumption of token storage (cf. Eddington, 2004; Keuleers, 2008; Skousen, 1989) is necessary to explain not only *n*-gram frequency effects, but also the speaker and context effects that we will meet in the section on phonetics and phonology.[16] Note that maintaining an inventory of the individual bigrams (*you+have, have+another, another+cookie, cookie+right, right+on, on+the, the+table*), trigrams (*you+have+another, another+cookie+right, right+on+the, on+the table*), four-grams (*you+have+another+cookie, cookie+right+on+the, right+on+the+table*), and so on would require not less storage, but many times more

(see Perruchet, 2018, for evidence that performance on statistical-learning tasks is better explained by storage of chunks than of transitional probabilities; i.e. *n*-grams). And even if it were somehow possible to store *You have another cookie right on the table* as some form of abstraction, this abstraction would efface the *n*-gram frequency information that is evidenced in these studies.

Unless, that is, learners store both; as in Abbot-Smith and Tomasello's (2006, p. 276) 'hybrid model comprising both abstractions and the retention of the exemplars of which those abstractions are composed'. But this risks giving us the worst of both worlds: the profligacy of an exemplar model – presumably the reason that most language acquisition theories favour abstraction-based accounts in the first place – plus the poor-data-coverage of a prototype model, coupled with more profligacy: the abstract prototype must be stored in addition to the exemplars, even though it adds no explanatory power.

All of the arguments and findings set out above also hold at the syllable level, where learners demonstrate sensitivity to the transitional probabilities of individual syllables (e.g. *pre+tty+ba+by*; Aylett & Turk, 2006; Saffran, Aslin, & Newport, 1996). They also apply at the level of individual phonemes (e.g. learners of English know that *f+t* is more common than *v+t*; Mattys & Jusczyk, 2001). The only way to explain these findings would be to assume that, in addition to word-level *n*-grams (*pretty+baby*), learners are also storing syllable-level and phoneme-level *n*-grams (and, under a hybrid account, more abstractions too). Again, this would require not *less* storage than simply storing whole utterances, but many times *more*.

Sentence-level constructions

While they disagree with regard to the timing and technical details, if there is one thing on which all mainstream accounts of child language acquisition agree, it is that by – at the latest – around 3;0 (Tomasello, 2000, 2003) children have abstract knowledge of word order. Constructivist, early-abstraction and generativist-nativist accounts claim, respectively, that learners of English have acquired 'some kind of abstract, verb-general, SVO transitive construction' (Tomasello, 2000, p. 216), 'detected the abstract word-order pattern of English transitive sentences' (Gertner et al., 2006, p. 686) and 'set parameters correctly at an extremely early age . . . includ[ing] word order' (Wexler, 1998, p. 29). Indeed, in many of my previous writings, I have been no exception, concluding for example (in Ambridge & Lieven, 2011, p. 239) that 'by 2;0, children almost certainly have at least some abstract, verb-general knowledge of the basic word-order rules of English (i.e. the SUBJECT VERB OBJECT transitive construction)'. (Though, as hinted at by the caveat 'at least some', I was already beginning to doubt the meaningfulness of the concept of a standalone abstract SVO construction.)

Yet, once again, the notion of an abstract SVO construction, schema or rule falls at the first hurdle of accurately describing the adult grammar, before we even consider data from empirical studies. It is transparently not the case that any verb can appear in the VERB position (e.g. *The comedian **laughed** her* [cf. *The comedian made her laugh*]), nor any OBJECT in the OBJECT position (cf. *The comedian laughed **a hearty laugh***), nor any SUBJECT in the SUBJECT position (cf. ***The magician** disappeared the rabbit* vs **Stalin** *disappeared his enemies*). We cannot solve the problem by simply appending

labels such as 'intransitive' or 'transitive' to verbs' lexical entries, because the acceptability of such sentences depends on the other arguments, and – in adult judgement studies – is finely graded (Ambridge, Pine, Rowland, Jones, & Clark, 2009; Ambridge, Pine, Rowland, & Young, 2008; Bidgood, Pine, Rowland, & Ambridge, submitted). The dilemma again is one of lumping versus splitting; and the notion of a single abstract SVO transitive construction is perhaps the most extreme and inaccurate example of lumping that we have encountered so far. Consider the various different sentence types that must be subsumed by a unitary SVO transitive construction (adapted from Ambridge & Lieven, 2015):

Contact (non-causative)	[AGENT] [ACTION] [PATIENT]	John hit Bill
Causative	[CAUSER] [ACTION] [UNDERGOER]	John broke the plate
Experiencer-Theme	[EXPERIENCER] [EXPERIENCE] [THEME]	John feared Bill
Theme-Experiencer	[THEME] [EXPERIENCE] [EXPERIENCER]	John frightened Bill
'Weigh' Construction	[THING] [MEASURE/COST/WEIGH] [AMOUNT]	John weighed 100lbs
'Contain' Construction	[CONTAINER] [CONTAIN] [CONTENTS]	The tent sleeps four people

What is particularly problematic in this case is that certain concrete instantiations of this would-be unitary construction are not just radically different, but polar opposites (e.g. *John feared Bill* vs *John frightened Bill*). Semantically, a putative SUBJECT prototype would have to encompass both a frightener and one who is frightened; a chaser and a fleer; a giver and a receiver. As we saw earlier with the example of large (wooden) and small (metal) spoons, a prototype category structure cannot represent these types of non-linearly-separable distinctions. If a *frightener* is a prototypical SUBJECT, then a *frightenee* is as non-prototypical a SUBJECT as one can imagine.[17] Formal linguists (in the sense of, e.g., Newmeyer, 2003) may object that syntactic subjecthood is entirely independent of semantics; but as well as being empirically false (as noted above, participants rate some transitive subjects as better than others), this objection misses the point. The very function of word order in morphologically impoverished languages such as English is to convey semantics (cf. *The dog bit the man; The man bit the dog*). If those semantics can flip entirely depending on the identity of the verb (*frighten/fear, chase/flee, give/receive*), then what is conveying the meaning simply cannot be an abstraction that is insensitive to the identity of the verb (i.e. some type of SVO construction, rule or schema).

The only solution is to admit the identity of the verb into the representation; but this is to jump out of the frying pan of lumping, and into the fire of splitting. Suppose that we posit a separate construction for each of the six types outlined above. This does not solve the problem because, to take just one example, the [EXPERIENCER] [EXPERIENCE] [THEME] construction inappropriately lumps across a variety of events (e.g. *John heard / saw / spotted / noticed / recognized Bill*) that differ considerably with regard to the nature of the interaction (who is the one 'doing something' in each of these cases? Sometimes John, sometimes Bill). Indeed, individual instantiations of the (putative) [EXPERIENCER] [EXPERIENCE] [THEME] construction vary continuously in their grammatical acceptability, speed of processing and production probability (see Ambridge, Bidgood, Pine, Rowland, & Freudenthal, 2016; Bidgood et al., submitted, who also find the same for the (putative) [THEME] [EXPERIENCE] [EXPERIENCER] and [AGENT]

[ACTION] [PATIENT] constructions, as well as their passive equivalents; see also Rissman & Majid, 2019, for some more general difficulties with the notion of these type of thematic role categories).

Indeed, Chang (2002) showed that, when faced with the task of acquiring basic syntax, abstraction-based connectionist models fail for the same reasons as the prototype-based verbal models to which they are conceptually related. A simple recurrent network model trained to produce sentences such as *The dog chased the cat* learned the training set very well (99% accuracy), but showed only 6% accuracy for sentences that required it to generalize a previously-seen item into an unwitnessed position. For example, if *dog* appears in training in SUBJECT position, but never in OBJECT position, the OBJECT prototype effectively excludes *dog*, making generalization into that role all but impossible. Chang's (2002) own Dual-Path model (see also Chang, Dell, & Bock, 2006) succeeds at this task, by virtue of its separate semantic (message) and syntactic pathways that allow it to map a single representation in the former (e.g. DOG) onto multiple representations in the latter (e.g. both SUBJECT and OBJECT position). However, in order to do so, it requires semantic role representations that represent a particularly extreme example of lumping; for example, the PATIENT slots collapses PATIENTs, THEMEs and EXPERIENCERs, even though – as discussed above – these are often polar opposites (e.g. *John feared Bill* vs *John frightened Bill*). Thus, while the model shows excellent performance with artificial grammars, it would not – I contend – be able to simulate difference in grammatical acceptability both between and within EXPERIENCER-THEME and THEME-EXPERIENCER passives (e.g. Ambridge et al., 2016; Bidgood et al., submitted). On the generativist side, computational parameter-setting models such as that of Sakas and Fodor (2012) operate at a higher level again, investigating only how different possible orderings of categories such as [SUBJECT] [VERB] and [OBJECT] could be triggered; the categories themselves are already known.

In general, then, we would expect the experimental data to be problematic for the idea of an abstract representation of word order, whether a generativist-nativist style formal rule that is insensitive to the identity of the verb and its arguments, or a constructivist style SVO construction with prototype structure. And this is exactly what we find. Even though virtually all child studies have given abstraction-based theories a head start by equating SVO with [AGENT] [ACTION] [PATIENT] or [CAUSER] [ACTION] [UNDERGOER] – ignoring all the other possible types listed above – there exist a large number of findings that are explained naturally by an exemplar account, but not one that replaces those exemplars with stored abstractions.

Ambridge and Lieven (2011, p. 221) summarized 14 elicited-production studies in which novel verbs were elicited in an SVO transitive construction (e.g. *He's tamming it*), having been presented solely in non-transitive forms during training. Across these studies, the majority of arguments – particularly for the younger children – were pronouns (e.g. *He's tamming it*) (e.g. 90% in Dodson & Tomasello, 1998). On its own, this finding could simply reflect discourse tendencies of English. However, Childers and Tomasello (2001) found that training children on overlapping exemplars with English verbs (e.g. *He's pushing it*) increased the proportion of 2-year-olds who produced an SVO utterance with novel verbs (e.g. *He's tamming it*) from 45% to 85%. Similarly, Akhtar's (1999) weird word-order study found that children used pronouns for around 50% of

all arguments when producing SVO transitives (e.g. *He's tamming it*), but *never* when imitating a weird word order produced by the experimenter (e.g. **Elmo the car gopping* not *He it gopping*) (see also Abbot-Smith, Lieven, & Tomasello, 2001; Matthews, Lieven, Theakston, & Tomasello, 2004, 2007; Savage, Lieven, Theakston, & Tomasello, 2003). This advantage for pronoun-based over full-noun based SVO transitives (e.g. *He's meeking it* vs *The dog's meeking the car*) is also seen in comprehension studies (e.g. Childers & Tomasello, 2001, Study 2).

These findings (which constitute clear evidence against generativist-nativist style context-free word-order rules) are usually taken as evidence for constructivist style *slot-and-frame patterns* or *lexically-based schemas* (e.g. *He's ACTIONing it*). The question is whether this is simply a metaphor for the kinds of on-the-fly analogical generalizations posited by exemplar accounts, or whether such abstractions are somehow stored and represented independently. Most constructivist accounts do not address this question directly, but seem to at least hint implicitly at the latter (e.g. Chandler, 2010, attributes such a position to Croft & Cruse, 2004; Goldberg, 2006; Langacker, 2009). Indeed, for my own part, I have often written about children acquiring slot-and-frame patterns without stopping to think which of the two possibilities I intend or imply. As we have already seen, the paper by Abbot-Smith and Tomasello (2006, pp. 281–282) is an exception in explicitly advocating for the latter view [emphasis added]:

> In view of the prevalence of item-based effects and frequency effects in syntactic acquisition (and which remain to some degree in adult language usage, see Dąbrowska, 2004), exemplar models of categorization are more attractive than a 'pure' prototype-abstraction model in which the extraneous details of original instances are completely lost. Furthermore, such exemplar-learning models are perhaps better able than pure prototype models to explain patterns of family resemblance in syntactic and morphological categories where there is no central tendency (e.g., Bybee, 1995). However, we would question the assumption that more abstract prototype categories are only generalized online and leave no permanent representational change. Even in exemplar models every time an exemplar is comprehended, its representation must change in some way, even if this merely involves registering frequency. If the comprehension or production of a novel utterance involves 'summing over' similar sets of exemplars, the frequency with which a set is called upon probably also leaves a trace. *Therefore, if the mutual similarities of a particular collection of exemplars (such as transitive sentences) are 'summed over' regularly, we believe this is highly likely to permanently change the user's linguistic representations in some way equivalent to the formation of some kind of more abstract representation.* A resolution to the drawbacks of both 'pure' prototype and 'pure' exemplar learning models is *a hybrid in which much of the extraneous details of original instances are retained but where some kind of more abstract schema is gradually formed on the basis of these.*

As Abbot-Smith and Tomasello (2006, p. 282) themselves note, 'It may of course prove difficult to empirically differentiate such a hybrid model from certain "pure" exemplar-learning models'. Here, then, are three non-empirical, theoretical arguments for a pure exemplar-learning model. The first is simply Occam's Razor. If we are positing exemplar storage anyway, we should not posit some additional abstractions on top unless they add explanatory power; and to our knowledge, neither Abbot-Smith and Tomasello

(2006) nor any other paper has made the case for a phenomenon that cannot be captured by a pure exemplar model, whether in the domain of language, or of learning and categorization more generally.

The second argument is the familiar lumping or splitting problem. We have already seen that a high-level abstraction such as SVO cannot accommodate differences between *frighten/chase/give* type sentences and *fear/flee/receive* type sentences (or, indeed, verb-by-verb differences within these two classes). The same is true for lower-level abstractions such as the putative slot-and-frame patterns *He's [ACTION]ing him/it*, since utterances constructed using such a template can again have opposite meanings (e.g. *He's chasing it* vs *He's fleeing it*). In general, as we have seen in all the domains covered in the present article, it is never possible to posit exactly the right abstractions; those that rule in all possible sentences and rule out all impossible ones.

The third reason to favour a pure exemplar-learning based account over the hybrid account posited by Abbot-Smith and Tomasello (2006) is that the additional abstractions posited raise more questions than they answer (if, as per the first point above, they answer any). Which slot-and-frame patterns do learners abstract, and why these ones (token frequency? type frequency? communicative function?)? How do learners move from semi-abstract slot-and-frame patterns (e.g. *He's [ACTION]ing it?*) to a fully abstract SVO construction? Are the various abstractions assumed independent or linked? For example, is the SUBJECT category in the fully abstract adult transitive construction (SVO; *She's dancing*) the same or different to the SUBJECT category in the intransitive construction (SV; e.g. *She danced*), the dative construction (SVOO; *She gave him a book*) and the passive construction (SV by O; *She was chased by him*). If they are 'linked', what exactly does this mean in terms of representation and processing? What about when these constructions are combined (e.g. the dative and passive, to yield SVO by O; *She was given a book by him*). And are the answers to these questions the same or different when we are talking about lower-level, less abstract slot-and-frame patterns (e.g. *She's [ACTION]ing it; She [ACTION]ed; She gave [PERSON] [THING]; It got [ACTION]ed by it*, etc.)?

A pure exemplar account bypasses all of these difficulties. A novel utterance is produced by analogy with all the stored exemplars that are sufficiently close to the target meaning. This might be an entire stored utterance (e.g. *She's dancing*), a set of exemplars with high semantic overlap (e.g. *She's running, She's jumping, She's dancing*; equivalent to the slot-and-frame metaphor) or – failing both of these options – a set of exemplars with lower, but still sufficient, overlap (e.g. *The girl danced; Sue danced; Jim is dancing; He's dancing*). Note that analogy across what would normally be considered 'different constructions' provides a ready-made explanation of *construction conspiracy* phenomena (Abbot-Smith & Behrens, 2006); for example, the finding that the acquisition of the German *sein* passive (e.g. *Der Reis war gekocht*, 'the rice was cooked') is boosted by experience with the *sein* copula construction (e.g. *Der Reis war Schwarz*; 'the rice was black'); see also Ninio (2018) for evidence of transfer-based analogy in children's spontaneous speech.

Of course, the devil is in the detail, in the need to explain the basis for analogy (one computational account is set out below); but the exemplars+abstractions account is not immune here, as it must similarly explain the basis for analogy that leads to the stored

slot-and-frame patterns and higher-level abstractions. If the basis for analogy can be figured out, there is no need to posit the additional stored abstractions: the same analogies can be used by a pure exemplar model to generate utterances as they are needed, bypassing altogether problems regarding links between constructions, combining constructions and so on.

In fact, there is one study that provides direct evidence for an exemplar account over a prototype account with regard to acquisition of the English SVO transitive construction (although its authors describe it as evidence for a prototype account, and do not discuss exemplar accounts). In a training phase, Ibbotson, Theakston, Lieven, and Tomasello (2012) exposed adults and children (aged 3–5 years) to sentences that all deviated in some way from the putative SVO transitive prototype, which they defined as 'an agent intentionally instigating an action that affects a patient' (p. 1272), such as *He sliced the bread* (p. 1288). That is, instead of an intentional AGENT, the SUBJECTS of the training sentences were INSTRUMENTS (e.g. *The key unlocked the door*), EXPERIENCERS (e.g. *John sees Sophie*) or involuntary FORCES (e.g. *The sun warmed the flowers*). Participants then completed a recognition test involving previously-seen sentences, various types of foil and – crucially – unseen sentences which were consistent with the putative SVO transitive prototype (e.g. *He sliced the bread*). Echoing the findings from the classic dot-pattern studies (e.g. Posner & Keele, 1970), adults claimed, with relatively high confidence (median rating 4/5 on a Likert scale), to recognize these in-fact-new items from the training session; a finding which the authors take as evidence for the existence of an SVO transitive prototype. Yet, echoing the findings from more recent dot-pattern studies (e.g. Zaki et al., 2003), adults recognized the previously-seen exemplars with even *greater* confidence (median 5/5), even though (unlike in Zaki et al., 2003) these were *not* matched to the new items for prototypicality, but were *less* prototypical. This is clear evidence for exemplar storage. Children showed even more straightforward evidence for an exemplar account over a prototype account, readily distinguishing the old and new exemplars, and showing no evidence of the false-lure effect shown by adults.

Computational modelling is less advanced in the domain of sentence-level syntax than in most of the other domains considered in this review (presumably because the problem space is so much larger). However, the work that does exist not only supports an exemplar account, but also provides evidence against abstraction-based accounts. Walsh, Möbius, Wade, and Schütze (2010) outline an exemplar model of the acquisition of basic syntax.[18] The model stores utterances as ordered, but otherwise unstructured, lists of words[19] plus boundary markers (b) (e.g. *b_you_like_tea_b; b_people_love_chocolate_b*). At test, the model's task is to classify each input utterance as grammatical or ungrammatical, which it does by evaluating its word-by-word similarity to all stored exemplar utterances (like most models, the assumption here is that essentially all input utterances are grammatical).

The metric on which utterances are compared is a distributional vector representing the one preceding and one following word in all corpus utterances (this assumption is not intended to be psychologically realistic; the distributional vector is simply a convenient way of operationalizing the semantic and functional properties stored by human learners). For example, assume that the model is asked to evaluate the grammatical

acceptability of the input utterance b_I_love_coffee_b, which it does by comparing it against (amongst all the other stored utterances) b_You_like_tea_b. First the model compares the left-hand distributional similarity of *I* and *You*. For example, they are similar in that both are frequently preceded by *do* and *can*, but differ in that *You*, but not *I*, is frequently preceded by *are* (e.g. in questions). Next, the model compares the right-hand distributional similarity of *I* and *You*. For example, they are similar in that both are frequently followed by *want* and *think*, but differ in that *You*, but not *I*, is frequently followed by *are*. This procedure is repeated for the remaining two pairs (*love+like, tea+coffee*), and for the two boundary pairs; i.e. how similar are *I* and *You* with respect to the frequency with which they begin a sentence; and how similar are *tea* and *coffee* with respect to the frequency with which they end a sentence? These five similarity scores are then summed to give an overall similarity score for b_I_love_coffee_b and b_You_like_tea_b. If this similarity score meets some predetermined threshold, then a match is declared. Walsh et al. (2010) implement a simple nearest-neighbour model in which an input utterance is deemed grammatical if it is sufficiently similar to any single exemplar; but the model could easily be adapted to compute similarity across *k* nearest neighbours with a decay function (e.g. Keuleers, 2008).

Walsh et al. (2010) showed that when trained on the Manchester corpus (Theakston, Lieven, Pine, & Rowland, 2001) the exemplar model was able to classify held-out utterances as grammatical or ungrammatical with around 98% accuracy. They also showed that an abstraction-based model which uses the same distributional vectors to assign words to clusters achieved significantly lower accuracy. Although the abstraction-based model still showed good performance in absolute terms (90% accuracy), this should not be taken as evidence for traditional linguistic categories such as VERB and NOUN or MASS NOUN vs COUNT NOUN; the abstraction-based model was not constrained to posit only such categories, but formed whatever clusters yielded best performance.

Finally, what about syntactic priming studies? Doesn't 'priming evidence support the existence of abstract syntactic representations' (Branigan & Pickering, 2017, p. 9)? Syntactic priming is a phenomenon whereby if presented with one exemplar of a particular syntactic construction (e.g. *The vase was broken by the hammer*), speakers are more likely than they would otherwise have been to re-use the same construction when asked to describe a subsequent picture or animation (e.g. to say *The bricks were pushed by the digger*, rather than *The digger pushed the bricks*). The claim is that syntactic priming effects of this type reflect 'repetition of aspects of abstract linguistic structure' (Branigan & Pickering, 2017, p. 2); i.e. that a stored abstract representation of (for this example) the passive construction is activated by the prime, and then used by the speaker to produce her own utterance.

The existence of syntactic priming effects for both adults and children is uncontroversial (see Branigan & Pickering, 2017; Mahowald, James, Futrell, & Gibson, 2016, for comprehensive and up-to-date reviews). What is controversial is the claim that the phenomenon 'is best characterized in terms of the presence or absence of (re)activation of purely abstract syntactic representations' (Feldman & Milin, 2017, p. 22; see other responses to Branigan and Pickering's target article by Kempsona & Gregoromichelaki, 2017, and O'Grady, 2017, as well as rival accounts based on the notion of implicit learning; e.g., Chang, Dell, Bock, & Griffin, 2000).

Indeed, the exemplar approach advocated in the present article can, in principle, easily accommodate syntactic priming effects (e.g. Snider, 2008; see also Goldwater, Tomlinson, Echols, & Love, 2011, for a different but compatible account of syntactic priming as exemplar-driven analogy). A prime sentence such as *The vase was broken by the hammer* activates in memory not an abstract representation of the passive, but stored concrete exemplars that meet some threshold for both surface similarity (phonological and suprasegmental) and semantic similarity, both passive (e.g. *The man got run over by a bus; The window was smashed by a ball*) and active (e.g. *The man broke the vase*). When generating a subsequent utterance, the speaker does so by means of on-the-fly analogy across relevant stored examples, exactly as she does when generating *any* utterance. The only difference is that a number of concrete passive exemplars, having been recently activated by the prime, are more available for retrieval than they would otherwise have been.

An exemplar account of priming can explain a number of findings that are problematic for abstraction-based accounts. First, the 'lexical boost' is a finding whereby the priming effect is boosted if lexical material is shared between the prime and the target (e.g. between ***The*** *vase* ***was broken by the*** *ball* and ***The*** *window* ***was broken by the*** *hammer*, or even – as in Savage et al., 2003 – between ***It got*** *pushed* ***by it*** and ***It got*** *broken* ***by it***). This finding is extremely reliable: a recent meta-analysis of syntactic priming studies found that the average of 63% production of the primed construction jumps to 77% with lexical overlap. Yet abstraction-based accounts can explain this boost only by positing some form of add-on, such as 'a representation that encodes a binding between constituent structure and the lemma (syntactic component) of the lexical entry for the head' (Branigan & Pickering, 2017, p. 12) or 'short-term activation of explicit memory traces' (Rowland et al., 2012, p. 53). Such add-ons have no independent motivation; they are posited solely to explain the lexical boost. In contrast, an exemplar account explains the lexical boost naturally, with no need for additional assumptions: the prime preferentially activates stored exemplars with which it shares surface (lexical, phonological, suprasegmental) and semantic overlap, with these exemplars therefore more available for use – via analogy – in subsequent production. An exemplar account also explains why the lexical boost appears to increase with development (Rowland et al., 2012): the more exemplars have been stored in memory, the greater the probability of lexical overlap between the prime and a stored exemplar.

Second, an exemplar account naturally explains why priming is seen in cases where there is substantial lexical overlap between the prime and the target, but only debatable overlap in terms of abstract syntax. Bock and Loebell (1990) famously showed that passive sentences such as *The construction worker was hit by the bulldozer* prime intransitive locative (i.e. *non*-passive) sentences such as *The 747 was landing by the airport's control tower*. This is entirely expected on an exemplar account, on the basis of overlap between these utterances in terms of overlap in lexical/phonological material (e.g. Ziegler, Goldberg, & Snedeker, 2018) and suprasegmental stress patterns. It is entirely unexpected on a standard linguistic analysis which views these two abstract constructions as having very different underlying syntactic structure. Again, then, the abstraction-based approach favoured by, amongst others, Branigan and Pickering (2017, p. 9) requires a special workaround; here the assumption that 'abstract syntactic

representations . . . are shallow and monostratal' (i.e. [NOUN PHRASE] [VERB PHRASE] by [NOUN PHRASE], as opposed to [SUBJECT] BE [VERB] by [OBJECT]). This solution is unsatisfactory, as it fails to account for the important semantic differences between the two constructions. For example, in *The worker was hit by the boss* (passive) and *The worker was sitting by the tree* (intransitive), 'the worker' is playing very different semantic roles (PATIENT and ACTOR respectively), which is precisely why traditional syntactic theories posit different syntactic structures in the first place. Even with this unsatisfactory workaround, the abstraction-based approach does not accommodate the findings of a recent modified replication of Bock and Loebell (1990), which suggest that this passive priming effect was driven solely by the lexical item *by* (Ziegler et al., 2018). Ziegler et al. found that *by* was *both necessary and sufficient* for priming passive sentences: no priming of passives occurred following locatives that lacked *by* (e.g. *The 747 was landing next to* [cf. *by*] *the airport's control tower*). Conversely, priming of passives *did* occur following *active* locative sentences with *by* (e.g. *The pilot landed the 747 by the control tower*).

Third, an exemplar account naturally explains why priming is boosted if the prime and target share not only lexical overlap, but also overlap in the ordering of their semantic roles. Chang, Bock, and Goldberg (2003) found that if syntactic structure (in the monostratal sense) is held constant, priming is boosted if the prime and target overlap in terms of their semantic roles. For example, a target sentence such as *The farmer heaped straw* [THEME] *onto the wagon* [LOCATION] is better primed by another THEME-LOCATION sentence (e.g. *The maid rubbed polish onto the table*) than by a similar LOCATION-THEME sentence (e.g. *The maid rubbed the table with polish*). Similarly, Ziegler and Snedeker (2018) observed priming at the level of the ordering of semantic roles, even without syntactic or lexical overlap. Again, this is entirely expected on an exemplar account under which the prime preferentially activates stored exemplars in proportion to the degree of both phonological and semantic overlap. And, again, it is highly problematic for an abstraction-based account. Branigan and Pickering (2017, p. 10) suggest that these findings 'could have reflected a tendency to repeatedly assign thematic roles (e.g., Location) to grammatical functions (e.g., direct object) or to word-order positions (e.g., immediately following the verb)'. But this tendency is not incorporated into their theoretical model; and, indeed, overlap in terms of particular semantics and particular word-order positions is more characteristic of an exemplar account than one based solely on overlap of abstract syntactic structure.

Phonetics and phonology

For many speakers of British English, the *t* sound in words such as *Saturday* can be pronounced as either *t* (/t/ in IPA) or as a glottal stop (/ʔ/ in IPA), roughly *Sa'urday*, with the former used in more formal contexts. In many working-class accents, such as Cockney and Estuary English, the latter form predominates; to the extent that high-ranking British politicians who speak with received pronunciation sometimes – to widespread mockery – affect a glottal stop when trying to appeal to down-to-earth voters, for example when touring a factory (e.g. Shariatmadari, 2015); see also Dick Van Dyke's Cockney chimney sweep in *Mary Poppins*). These types of sociolinguistic effects are uncontroversial, and

have been well known since at least Labov's (1963) famous Martha's Vineyard study (see Hay, 2018, for a recent review), and are perhaps why exemplar accounts have had more influence in phonetics and phonology than in other areas of language acquisition (Pierrehumbert, 2001, 2002; see also Port & Leary, 2005; Ramscar & Port, 2016 on the impossibility of a discrete inventory of phonemes, and Schatz, Feldman, Goldwater, Cao, & Dupoux [submitted] for a computational model that simulates findings from infant phonetics research without making use of phonetic categories). Yet these effects are neglected entirely by mainstream accounts of word learning, inflectional morphology and syntax (see earlier sections) – and, indeed, by most of the computational exemplar models reviewed so far – which start from the assumption that learners represent an idealized word form (e.g. *Saturday*), from which the phonetic details of individual exemplars have been abstracted away. British speakers who switch between *Saturday* and *Sa'urday* depending on the sociolinguistic context therefore present a familiar dilemma for non-exemplar accounts: lump or split.

The lumping approach posits that learners have a single prototype representation of each word (in this case, a form somehow in between *Saturday* and *Sa'urday*), and some kind of high-level model which specifies how each of these idealized forms is phonetically realized in particular cases. This claim rapidly starts to look implausible, when it is borne in mind that learners would need a different realization model for every different type of speaker and accent they can recognize (e.g. older vs younger; working- vs middle-class vs upper-class; male vs female; British vs American English; Scottish vs English; Glasgow vs Edinburgh; Manchester vs Liverpool).

The splitting approach posits that, for each word, learners have a number of *different* prototypes for each word (e.g. *Saturday* and *Sa'urday*). But, again, once you start splitting, you can't stop. These two prototypes are certainly not sufficient to explain all of the possible pronunciations of *Saturday* that a British speaker might encounter and recognize as characteristic of a particular group, context or person. Are we then to posit separate *Saturday* prototypes for (a) older working-class women from Glasgow speaking in a formal context, (b) middle-class teenage boys from Manchester chatting informally, and so on all the way down to the level of individual speakers who we can recognize by their voices?

In fact, even this would not solve the problem. Within a given speaker, the pronunciation of a 'the same word' differs depending on the context. Hay and Bresnan (2006) show that, in New Zealand English, the /æ/ sound in *hand* is pronounced differently in idioms (e.g. *give me a hand*) and referals to actual hands (e.g. *my hand hurts*). Similarly, the /i/ sound in *give* is pronounced differently when the speaker is referring to literal giving (e.g. *give me a pen*) versus metaphorical giving (e.g. *give me a chance*). These findings demonstrate not only that the lumping-or-splitting problem holds within individual speakers but, furthermore, that not only single words but entire phrases are stored with phonetic detail.

The exemplar approach accounts straightforwardly for these findings, and sidesteps the lumping-or-splitting dilemma. Every utterance heard is simply stored as an episodic memory that contains both phonetic detail and factors such as the speaker identity and context. If the assumption of such rampant storage seems implausible, then consider the alternative: is it *more* plausible that we represent and store a realization model for every

combination of speaker, social context and lexical context that we encounter, and update these models in real time? And, if not, then what – other than an exemplar model – is the alternative?

As for the other domains of acquisition considered here, an exemplar account is again well supported by empirical data. Over 40 years ago, Craik and Kirsner (1974) found that adult listeners showed better recognition of words presented in the same than in different voices (male vs female) at training and test. Palmeri, Goldinger, and Pisoni (1993) found that this effect was undiminished when as many as 20 different voices were used, or when the gender of the speaker was held constant. Goldinger (1996) further demonstrated that, when *different* voices are heard at training and test, listeners show an advantage for *similar* over dissimilar voices, exactly as would be predicted under an exemplar model which assumes that learners are averaging on the fly across multiple stored exemplars (but not a simple nearest-neighbour exemplar model). Both the same- and similar-voice advantage held when training and test were separated by one day, but not one week, indicating that, as we have already seen for various computational instantiations, exemplar storage does not necessarily entail perfect memory; a point to which we will return in the final section.

Similar findings have also been observed for infants. Houston and Jusczyk (2000) found that 7.5-month-olds showed a listening preference for previously-familiarized words, but only when the speaker was of the same gender at training and test. Although this effect had disappeared by 10.5 months, the adult findings above suggest older infants are not incapable of detecting speaker differences, but that their various stored exemplars of (say) *cup* are sufficiently similar to support recognition (which must be the case for adult speakers, since we are clearly capable of recognizing a familiar word produced by a new speaker). Neither are the exemplars rapidly lost. Houston and Jusczyk (2003) found that 7.5-month-olds' preference for previously-familiarized words held a day later, but *only* when the speaker was the same at training and test.

As a consequence of these well-known and long-standing findings, exemplar accounts have become particularly well established in phonetics and phonology (e.g. Hay & Bresnan, 2006); more so than for the other domains considered in this review. For example, Johnson (1997) outlines a *k*-nearest-neighbour model with learned attention weights (based on Nosofsky, 1990, GCM) which simulates the identification of both male vs female speakers and different individual speakers of each sex (see also Pierrehumbert, 2001). Particularly interesting is the exemplar model of Walsh et al. (2010), which we encountered in the section on sentence-level constructions, but which has also been applied in the present domain, where it learns how segments are combined into syllables (as opposed to words into utterances). This work is particularly interesting because it shows how the same architecture can explain exemplar effects at different levels of analysis (albeit via separate models, not simultaneously). Further examples of computational exemplar models in this domain – all designed to simulate the types of speaker and context effects outlined above – include Jusczyk's WRAPSA (1993), Goldinger's (1998) MINERVA, Singh, Bortfeld, and Morgan's (2001) DRIBBLER, and Johnson's (2007) XMOD. The technical details of these models are not important for our purposes. But what is important is the fact that abstract-prototype models such as Kuhl's (1991) perceptual magnet approach (see Guenther & Gjaja, 1996, for a computational implementation),

as well as rule-based models such as generative phonology (Chomsky & Halle, 1968) and Optimality Theory (Prince & Smolensky, 2008) cannot simulate these types of well-documented effects, because they operate at a level of abstraction that is far removed from individual exemplars. In summary, the advantages of exemplar models are generally well understood in phonetics and phonology, and it is time that other domains of language acquisition followed this lead.

Conclusion

In summary, in the domains of (1) word meanings, (2) morphologically inflected single words, (3) *n*-grams, (4) sentence-level constructions and (5) phonetics and phonology, abstraction-based accounts fail for two reasons. First, it is not possible to posit abstractions that delineate possible and impossible form; e.g. that (1) rule in pool *tables* and data *tables*, but rule out chairs, (2) rule in the past-tense forms *netted* and *bet* but rule out **setted* and **jet*, (3) rule in the bigram *f+t* but rule out (probabilistically) *v+t*, (4) rule in both *John feared Bill* and *John frightened Bill* but rule out **John laughed Bill*, (5) rule in Speaker A but rule out Speaker B as the person who produced a particular word (e.g. *Sa'urday*). Second, for each domain, empirical data provide evidence of exemplar storage that cannot be captured by putative abstractions: e.g. speakers prefer and/or show an advantage for (1) exemplar variation even within word-meaning 'category boundaries', (2) novel inflected forms that are similar to existing exemplars, (3) *n*-grams that have occurred frequently in the input, (4) SVO sentences with *he* as SUBJECT and *it* as OBJECT and (5) repeated productions of 'the same' words that are phonologically similar or, better still, identical.

   An exemplar account not only (1) avoids the problems facing abstraction-based accounts (all some versions of the 'lumping or splitting?' problem) and (2) explains the exemplar effects observed in every domain but (3) provides a unifying account of language acquisition *across* all of these domains. Whether we are looking at the level of individual phonemes, single words, multi-word phrases or full sentences, an exemplar accounts offers exactly the same explanation: storage, and subsequent retrieval with analogy, of episodic memory traces including the phonetic detail of what was heard, the speaker's intended meaning, sociocultural details (e.g. the setting, others present), and so on. In contrast, abstraction-based accounts offer an explanation of a simplified, idealized version of a phenomenon at one particular level, leaving crucial details to the morphologists, the syntacticians, the phonologists and so on. For example, an account of (1) word meaning might ask how children learn the action label *run*, ignoring (2) different morphological forms (*run+ran*), (3) *n*-gram phrases (*run+a+bath, run+for+office*), (4) sentence level combinatorial properties (*The coach ran the race* vs **The coach ran the sprinter*) and (5) phonological realizations produced by different speakers.

   The fact that an exemplar account gives a unifying explanation of language acquisition across all domains is advantageous not solely for reasons of parsimony; the account also offers a ready-made explanation for correlations that are observed across different acquisition domains. For example, a number of studies (see Dąbrowska, 2018, for an up-to-date review and novel adult data) have found correlations across learners on measures of vocabulary and morphosyntactic knowledge, even when looking at adults. This finding

does not sit comfortably with generativist accounts under which the lexicon and the grammar are separate modules (e.g. Pinker's, 1999, 'words and rules'), or with constructivist accounts under which speakers form a morphosyntactic abstraction when they have reached a critical mass of lexical exemplars (e.g. Marchman & Bates, 1994); particularly given that the finding holds for adults, who presumably reached any would-be critical mass long ago. However, they are expected under an exemplar account, since exactly the same processes – storing episodic memories and generalizing across them in comprehension/production – are operational across domains and across the lifespan. The same point can be made for the observed bidirectional relationship between phonological and lexical development (see Stoel-Gammon, 2011, for a review). Although few researchers would find this relationship surprising, the traditional division of theoretical labour means that accounts of phonological and lexical development rarely make reference to one another. Again, an exemplar approach offers a unifying account of acquisition across these domains, and explains the observed correlations as reflecting the same learning process. More generally, an exemplar approach unites language acquisition with the acquisition of other cognitive skills (e.g. Chater & Christiansen, 2018) such as non-linguistic categorization (e.g. Mack et al., 2013) and recognition (e.g. Nosofsky, 2016). As we have seen, there even exists at least one proposal for how a generalized exemplar theory could be implemented in the brain, in terms of 'connectivity between striatal neurons and neurons in sensory association cortex' (Ashby & Rosedahl, 2017, p. 472).

Despite the theoretical and empirical advantages of exemplar models, they have made very little headway in theories of language acquisition. Partly this is because many theorists seem not to have considered in any detail the question of representation, and exactly what abstractions that they informally posit and discuss would entail. But when exemplar models are raised, many theorists quickly dismiss them as intuitively implausible. I therefore end by considering some of the most commonly raised objections to exemplar models.

Objection: it's not plausible that we store all this information

This is perhaps the major sticking point for exemplar models. The idea that we store in memory every linguistic episode that we have encountered (and, for that matter, every non-linguistic episode too) is intuitively implausible. Partly this is because it doesn't feel that way subjectively (i.e. we cannot recall memories at will); partly because we are used to thinking about memory using the finite-storage metaphor of our times (a store house, a tape recorder, computer memory).

But even if memory capacity is finite, it is very large. Although estimating this capacity is not straightforward, at least one memory expert estimates the total at around a million gigabytes; enough to record 300 years' worth of HD video (www.scientificamerican. com/article/what-is-the-memory-capacity/). And this is based on a rather conservative estimate of 1 billion neurons. Perhaps the best recent estimate (Lent, Azevedo, Andrade-Moraes, & Pinto, 2012) puts the number at 86 billion. Capacity is not the problem. A more reasonable objection is that it would take an unfeasibly long time to retrieve and analogize across all instances stored in memory. For this reason, some exemplar modellers have turned to massively parallel quantum computing (e.g. Skousen, 2005); the

outstanding question is whether the brain operates in this manner (e.g. Hameroff & Penrose, 2014).

Objection: but we do forget things, right?

Yes and no. Certainly the vast majority of exemplars are not available for conscious recall; i.e. they have been 'forgotten'. Accordingly, some exemplar approaches (e.g. many instantiations of Nosofsky, 1990, Generalized Context Model) incorporate a decay function at the level of whole exemplars and/or individual details within exemplars. For example, Hintzman's MINERVA2 simulation of human performance in dot-pattern classification tasks incorporates a bout of forgetting in which 75% of features (i.e. details within individual exemplars) are set to zero. But some exemplar approaches (see Chandler, 2002, for a review) argue that 'forgetting' does not reflect the weakening or decay of a memory trace. Rather, interference in memory might make a given exemplar difficult to recall, but it is still there.

Evidence for this view comes from experiments which show that we can recognize (and therefore must have stored) experiences that are not available via free recall. Standing (1973) found that participants who were shown 10,000 photographs could, two days later, subsequently categorize a randomly-chosen subset of 160 as seen or unseen with an average success rate of 83%. Furthermore, there is reason to believe that the failures represent not decay but interference: pictures chosen to be particularly vivid (e.g. a dog with a pipe in its mouth) showed better recognition than normal pictures (e.g. a dog without a pipe), precisely because interference is minimized.

Accordingly, some exemplar models (e.g. LEX, TiMBL) do not incorporate an explicit forgetting mechanism at all (except indirectly in the form of interference in memory, which arises simply as a consequence of storing more and more exemplars). Skousen's (1989) AM adopts a halfway house: exemplars are not forgotten as such, but – in most instantiations – have only a 50% chance of being retrieved on a given test trial.

Objection: but isn't forgetting details of an exemplar a type of 'abstraction'?

Not necessarily. Suppose that, after a delay of one week, a participant can recognize *cup* as a previously-presented word, but shows no same-/similar-speaker advantage; i.e. no retention of phonetic detail. One possible explanation is that she has somehow merged the *cup* that she previously heard into idealized representations of *cup* at both the phonological/phonetic level and the semantic level (see earlier sections). However, this explanation is problematic for reasons that we have already explored in detail. It is not possible to posit an idealized phonetic representation of *cup* that rules in all – and only all – possible pronunciations, or an idealized semantic representation that rules in all – and only all – possible entities that might be referred to as a *cup* (including, for example, the World Cup, which is either a tournament or a statuette). In short, there is no plausible idealized *cup* into which this particular exemplar can be merged. The exemplar account offers a much more straightforward explanation: the participant is able to access only some details of the episodic memory (e.g. the semantics invoked for her by *cup* during

the training session); others (i.e. fine-grained phonetic details) have been subject to either memory decay or memory interference.

The weakness of the abstraction-as-forgetting account becomes even more apparent when we consider other domains. Take, for example, the finding discussed in the section on inflectional morphology that, when asked to judge novel English past-tense forms, participants show sensitivity to the fact that verbs ending *f, th, s* or *sh* form their past tense with *-t*. It would seem very odd to argue that English speakers witness forms like *missed, hissed, wished* and *kissed*, and form an abstract *VERB-FRICATIVE+t* template by simply forgetting the beginnings of the individual words (which is what, after all, convey the different meanings). More importantly, this claim is at odds with the considerable evidence for the storage of such forms; i.e. frequency effects at the level of individual ready-inflected forms. Abstraction-as-forgetting is even more of a stretch at a sentence level, which would involve a speaker forgetting all of the individual words of a sentence – again at odds with findings of widespread lexical and *n*-gram effects – and retaining simply a highly abstract configuration of them.

Objection: but how do we 'know' what to store? After all, we can't know in advance that – for example – the colour of the speaker's eyes or the shape of her face is linguistically irrelevant; we must be constrained to ignore such details and encode only linguistically-relevant ones

Exemplar accounts assume that we do not know in advance what to store (e.g. Chandler, 2010). We take in as much information as we can from each episodic snapshot, albeit modified by attention (e.g. if our primary goal is to determine whether or not we recognize the speaker's face, we might pay more attention to features in the visual than auditory modality). As language acquisition researchers, we generally assume that listeners store linguistic details – or even a narrow subset of 'grammatically relevant' details – and disregard others. But if we were researchers of facial processing (a domain that has its own exemplar versus abstractions debate; e.g. Quiroga, 2017) we would focus instead on people's ability to record eye colour and face shape, ignoring auditory information. And if we were researchers of body language, we would be interested in yet another set of features. Indeed, latent learning (as it is usually termed in animals), implicit learning and perceptual learning (in humans) are well-known phenomena whereby participants show evidence of learning, and improved performance on subsequent tasks, even when they are viewing or processing stimuli with no particular goal – or even a goal that later turns out to be irrelevant – in mind (e.g. Kellman, 2002; Reber, 1969, 1989; Thorpe, 1956).

Even within the domain of language acquisition, what is grammatically irrelevant in one language (e.g. speaker gender, evidentiality, aspect, absolute position in geographical space) is critical in others. We cannot know what features will be important in advance, because the features that turn out to be relevant are determined retrospectively by some future goal; whether that is to produce a novel utterance of one's own, recall the gist of what the speaker was saying, recognize her voice or face, or determine her attitude towards the other people present or the topic under discussion. At the linguistic level, it is exactly these 'irrelevant' properties such as the precise sequence of words used (*n*-gram

effects), the phonetic detail (speaker-effects) and the particular meaning that the speaker seemed to have in mind at that instant (sidestepping the homophony problem when learning word meanings) that are key to the exemplar effects summarized throughout this article.

Objection: abstractions and exemplars are not the only two possibilities. Connectionist models undergo changes in response to each individual exemplar, but do not actually store them

As Chandler (2010) notes, connectionism is an approach, not a single model. Many different types of connectionist models are possible. At one extreme, a connectionist model with a sufficiently large number of hidden units to represent each exemplar – particularly if supplemented with a mechanism for representing the order in which the exemplars were presented (e.g. Elman, 1990) – would be what Chandler (2002, p. 63) calls 'a de facto exemplar-based model': each individual exemplar would be represented by a unique pattern of activation in the model's hidden layer. This is entirely compatible with exemplar accounts, which make no particular claims about the way in which exemplar storage is actually implemented. Indeed, unless one believes in Grandmother cells (e.g. Fodor & Pylyshyn, 1988), everyone agrees that the storage in the brain of any particular exemplar, concept, etc. must be distributed across neurons in a way that is analogous to the distributed storage employed by connectionist (or 'neural-net') models. At the other extreme, a connectionist model with very few hidden units would be forced, by design, to make abstractions that discard huge amounts of information, and so would fail to yield *n*-gram effects, speaker-recognition effects, and all of the other exemplar phenomena summarized in the present article. As we have already seen, any connectionist model that does not maintain a unique representation for the relevant items is also unable to learn non-linearly-separable distinctions such as the English past-tense forms *netted* and *bet* versus **setted* and **jet*. Connectionist models, then, are not an alternative to abstraction-based or exemplar models, but merely a computational framework in which either type of model (or, something in between) can be instantiated (e.g. Regier's, 2005, LEX and Kruschke's, 1992, ALCOVE both use connectionist architectures to instantiate exemplar models).

  Discriminative learning models (e.g. Baayen, Hendrix, & Ramscar, 2013; Milin, Divjak, Dimitrijević, & Baayen, 2016; Ramscar, Dye, & Klein, 2013; Ramscar, Sun, Hendrix, & Baayen, 2017) are a special case because – although effectively a form of connectionist model – they eschew hidden layers and, with them, abstract linguistic representations. This makes them of a piece with exemplar models. As an example, consider the discriminative learning model of Ramscar, Dye, and McCauley (2013), which simulates children's acquisition and overgeneralization of English plural marking (e.g. **mouses*). The model works by learning the predictive value of semantic cues (e.g. *multiple items, single item, multiple mouse items, single mouse item, mousiness, stuff*) for linguistic forms (e.g. *dog+s, dog+0, mouse+0, mice+0, mouse+s*). The use of cues to predict categorical outcomes renders this model all but indistinguishable from an exemplar model.

On the face of it, an apparent difference is that, because discriminative learning models do not store a veridical record of each input utterance, it is not possible – as it is for an exemplar model – to turn back the clock and determine the input to which the model was exposed. In fact, however, this is also the case for many exemplar models (e.g. Daelemans & van den Bosch, 2010, TiMBL; Regier, 2005, LEX; many versions of the GCM), which use attentional feature weights to learn to preferentially weight the cues that best discriminate different outcomes, in a manner identical to discriminative learning models. Consequently, a difference between the approaches can be seen only when we consider 'pure' exemplar models that do not posit feature weights (e.g. Skousen's, 1989, Analogical Model). Chandler (2017, p. 67) argues against feature weightings on the basis that they 'must be determined through separate experimental and computational operations applied to a set of stimuli ahead of time' and 'cannot be interpreted theoretically as resident components of the mental representations of the linguistic forms being modelled'. However, neither of these criticisms would seem to straightforwardly apply to feature weightings that emerge over the course of learning, as for discriminative learning models. To sum up, while, in most cases, a discriminative learning model *is* an exemplar model and vice versa, any attempt to differentiate the two will hinge on the theoretical (dis)advantage and/or theoretical necessity of feature weighting (I thank Michael Ramscar for helpful discussion and clarification of this issue).

Objection: arguing that there is a meaningful difference between an exemplar account and a constructivist stored-abstractions account such as Tomasello (2003) is just splitting hairs: the only difference is whether the abstractions are stored or generated on the fly, and what difference does that really make to anything?

In fact, the accounts differ not only with regard to whether abstractions are stored or generated on the fly but, much more importantly, with regard to the nature of those abstractions. Constructivist stored-abstraction accounts posit very large, very abstract generalizations like the transitive [SUBJECT] [VERB] [OBJECT] construction, that can be used to produce novel utterances like *He kicked the ball* (see earlier section on sentence-level constructions). An exemplar account differs in two important respects.

First, the emergent (rather than stored) generalization is at an extremely fine-grained level; more fine-grained than [SUBJECT] [VERB] [OBJECT], than [AGENT] [ACTION] [PATIENT], even than [KICKER] *kick* [KICKEE]. For example, depending on exactly what exemplar utterances have been stored, a speaker might generate *He kicked the ball* by analogy with *He kicked the tree, She kicked the stone* and *He found the ball*.

Second, the generalization that emerges depends on exactly what the learner is trying to do at that moment. Suppose, for example, that, rather than trying to produce the utterance *He kicked the ball*, she is trying to figure out the meaning of an unknown word that she has just heard in the phrase *He kicked the widget*. The learner might recruit exactly the same exemplars – *He kicked the tree, She kicked the stone* and *He found the ball* – but analogize across them in a very different way, placing most weight on semantic similarities between *tree, stone* and *ball*.

Objection: the assumption of representing language as stored episodic memory traces doesn't fit with either cognitive or neuroimaging findings regarding memory

A traditional view is that memory consists of two largely separate systems, underpinned by different brain systems. For example, under Ullman's (2015) *declarative/procedural* model, the *declarative* (or explicit) system is characterized by conscious awareness (i.e. we can deliberately retrieve or 'think about' these memories) and depends on the 'the hippocampus and other medial temporal lobe (MTL) . . . and . . . neocortical regions, particularly . . . the temporal lobes' (p. 955). The *procedural* system (one of a number of implicit systems, alongside – for example – perceptual systems) is characterized by a lack of conscious awareness, and 'involves a network of interconnected brain structures rooted in frontal/ basal-ganglia circuits, including frontal premotor and related regions, particularly BA 6 and BA 44' (p. 956). Declarative memory can be subdivided into *semantic* memory (e.g. the knowledge that Paris in the capital of France) and *episodic* memory (e.g. the memory of a particular trip to Paris).

The exemplar account advocated here blurs these distinctions, because it places the language-learning burden squarely on the shoulders of *episodic* memory, despite the fact that we usually think of (native) linguistic knowledge as mainly *implicit*, though with some *semantic* memory for word meanings and, perhaps, morphosyntactic generalizations that are easily verbalized (e.g. 'add -*s* to make a noun plural').

But this is not necessarily a problem. Although this *memory systems* framework is the one that is best known to researchers of child language acquisition (perhaps due to its popularization by Pinker & Ullman, 2002, in the context of the English past-tense debate), it is far from the only one. The *processing modes* framework (e.g. Roediger & McDermott, 1993, *conceptual-perceptual* theory) argues that dissociations between performance on implicit and explicit memory tasks do not in fact reflect different memory systems, but merely different manners of processing. A third approach, the *component process* framework (e.g. Moscovitch, 1992), proposes that memory consists not of 'a handful of memory systems or a couple of processing modes' but rather 'dozens of different processing components . . . associated with different brain regions and . . . recruited in different combinations by memory' (Cabeza & Moscovitch, 2013, p. 50).

Child language researchers cannot be expected to play umpire in a debate in a different field, but the very existence of such a debate means that proposals in the domain of language acquisition should not be dismissed on the basis that they are at odds with a particular model of memory, particularly one that – at least according to some experts in this domain – is no longer the dominant model in the field.

Finally, at least one recent functional neuroimaging study (Mack et al., 2013; see also Nosofsky et al., 2012) has provided direct evidence for an exemplar- over prototype-account of categorization (albeit non-linguistic categorization). Mack et al. (2013) had participants perform a standard categorization task (involving different shapes of different colour, size and position) and used statistical models to generate, for each stimulus, for each participant, categorization predictions from a prototype and exemplar account. Although the behavioural predictions of the prototype and exemplar models were very

similar, and could not differentiate between them, the internal states of the models were very different, and could be related to patterns of neural activation, as observed using fMRI. Mack et al. found that, for 13/20 participants, the exemplar model yielded a better fit to the fMRI data than the prototype model, with only a single participant showing the opposite pattern (and six ties). More generally, Ashby and Rosedahl (2017) outline a proposal for how exemplar accounts could be instantiated neurobiologically. In summary, far from being biologically implausible, exemplar accounts exist as neurobiological models that provide a close fit to imaging data.

Objection: saying that production/comprehension works 'by analogy to stored exemplars' is just a big hand wave; there is no detail about how this analogy works

Certainly, an exemplar account of language acquisition will require details of the analogical process to be spelled out; and spelled out in sufficient detail that the account can be implemented and tested as a computational model. For most domains, this process has only just begun; and my hope in writing this review is that colleagues will be inspired to undertake exactly this kind of work, particularly in the domain of core morphosyntax.

In the meantime, it is important to be clear that stored abstraction accounts do not have any clear advantage here. Constructivist accounts such as that set out by Tomasello (2003) (or, indeed, by Ambridge & Lieven, 2015) also face the challenge of specifying the analogical processes that give rise to abstractions (in their case, stored abstractions). If anything, the challenge is greater, since such accounts need to explain the formation of very high-level abstractions (e.g. [SUBJECT] [VERB] [OBJECT], or, at least [AGENT] [ACTION] [PATIENT]) from input exemplars that have very little in common. Generativist accounts have an advantage in that they do not need to account for the formation of these high-level abstractions (e.g. [SUBJECT] [VERB] [OBJECT]), which have an important innate basis, and so are fully formed as soon as the relevant parameters have been set (e.g. Wexler, 1998). However, with this advantage comes a significant cost: if children are operating with such high-level abstractions from such an early age, it is very difficult to explain input effects, whether at the levels of *n*-grams, frequent sentence-level utterance patterns (e.g. *He's X-ing it*), the ability to identify different speakers, and so on.

In summary, although they are few in number and limited in scope, fully-computationally-implemented exemplar models not only exist for all of the domains covered here, but have been relatively successful at explaining empirical data from adults and children. Of course, they are far from perfect. But where are the alternatives? Where are the computational models of child language acquisition that work by building or bootstrapping into the types of abstractions posited by most linguists and virtually all acquisitionists: syntactic constituent structure or sentence-level constructions? They don't exist because they don't work. Cognitive linguists (e.g. Bybee, 1985, 2010; Croft, 2000, 2001) and computational modellers have learned a lesson that should be heeded by researchers in child language acquisition: analogical exemplar models are the only game in town.

A final word on intuitive (im)plausibility

No doubt many of my colleagues (including many of the commentators to follow) will find it difficult to shake off the intuition that exemplar models are just downright implausible. But if so, tell me this: What alternative explanation of the many exemplar phenomena outlined here would be more plausible? To take just one example, it is not clear how the ability to recognize individual accents and speakers can be explained by a non-exemplar model, other than by assuming that we maintain, for each speaker, a realization model that transforms highly abstract phonological representations into the precise surface forms heard. This is not impossible, of course. But it is difficult to make the case that it is more intuitively plausible than the exemplar-storage alternative. And likewise for all of the other phenomena reviewed here including – for me, a particularly compelling example – knowledge of the frequency of particular *n*-grams (e.g. *drink+of+milk > drink+of+tea*). In conclusion, when compelling empirical data contradict our deeply-held intuitions, it is time to revise those intuitions. It is not intuitively plausible that the earth is round and revolves around the sun, that diseases are spread by invisible microscopic agents, that (almost) all cells contain a full set of instructions for building a human, or that we share a common ancestor with bananas. But we have eventually come to accept each of these theoretical models because, in its own domain, each is the only one that is consistent with the empirical data.

ORCID iD

Ben Ambridge https://orcid.org/0000-0003-2389-8477

Notes

1. A reviewer asked whether, on this definition, computers are also capable of learning language. My answer is 'not yet' because although computers store exemplars and analogize across them, they are not doing so in the service of comprehension or production; i.e. in order to understand a speaker's intended message, or to communicate an intended message to a listener (cf. e.g. Iozzi & Barbieri, 2009, for children).
2. My position is not shared by all usage-based or cognitive-linguistics approaches. For example, Langacker (1988) (in the paper that coined the term 'usage-based') argued for the storage of both exemplars and generalizations. Goldberg (2006) similarly argued for both types of storage. In particular, with respect to the storage of generalizations, Goldberg (2006, p. 62) argued that 'in learning constructions . . . such as relative clauses, questions [and] clefts – learners' representations are likely to abstract quickly over the particular semantics involved. For example, learners' representation of the English relative clause construction may contain almost no information about the semantic content that happens to have appeared in actual

usage events. Upon hearing, *the man who sat on every hill*, we do not retain the phrase *sat on every hill* as part of the representation of the relative clause, let alone the phonetic character of the way the word every was pronounced.' In contrast, my position is that we *never* abstract over the particular semantics involved, and that we *do* store the phrase *sat on every hill* with detail not only about the phonetic character of the phrase, but also (as far as possible) the appearance of the speaker, the chair she was sitting on, etc. See also Culicover, Jackendoff, and Audring (2017) for a current approach that posits both stored exemplars and stored abstractions.

3. I would have a quarrel with them, however, if they assume the stored exemplars that give rise to the slot-and-frame patterns (e.g. *What's Daddy doing? + What's Mummy doing? + What's he doing → What's __ doing?*) are stored just once each, with some kind of representational strength representing frequency of occurrence in the input. Under the account I am advocating, each individual input token of, e.g., *What's Daddy doing?* – in the exact form that it was pronounced by that particular speaker – is stored individually, along with contextual information and understood meaning.

4. I put this question to Elena Lieven, a co-author on all of the papers quoted above. She replied (personal communication) that children 'start to behave AS IF they have stored abstractions – but I recognise that it could be generalisations on the fly on the basis of whatever weightings and associations have been made in the network during development'.

5. My definition of a child language acquisition researcher is someone who designs and conducts experimental studies and/or corpus analysis and/or computational modelling in this domain. 'Pure' theorists play many important roles, but they are not child language acquisition researchers.

6. I refer to my proposal as a *radical exemplar account* (approvingly echoing Croft, 2001, *Radical Construction Grammar*) in order to differentiate it from these merely *exemplar-**based*** accounts, which assume that stored exemplars are replaced and/or supplemented by stored more abstract representations.

7. Exemplar accounts do not necessarily predict that an analogy can be made on the basis of a single exemplar *in practice*; it depends on the precise implementational details of the model. For example, some computational implementations of an exemplar model include a forgetting function; so an exemplar may be forgotten before it can be used for analogy. However, there is some evidence for generalization on the basis of a single exemplar in the domain of word learning, where the phenomenon goes by the term *fast-mapping* (e.g. Carey & Bartlett, 1978). Sometimes, when taught a novel label for a novel object, children are able to use and/or comprehend that label (even in phonologically non-identical form) to refer to another instance of that novel object (e.g. the same object viewed from a different angle, and so hence a different visual exemplar). Since neither the phonological nor the visual form are identical across the two instances, this constitutes generalization/analogy on the basis of a single exemplar.

8. Goldberg (2006, p. 45) argues that 'the constructionist approach to grammar offers a way out of the lumper/splitter dilemma . . . [since] individual patterns that are fully compositional are recorded alongside more traditional linguistic generalizations'. My position is that no approach which posits stored generalizations can explain the lumper/splitter dilemma, as it is not possible to posit a generalization at any level (including the apparently-simple level of word meanings) that does not involve intolerable lumping and/or intolerable splitting.

9. A reviewer asked, 'Does BA really want to say the only meaning for "table" is the set of tables we've actually experienced being labelled as tables?' Essentially, yes. Our only *stored* meanings (note the plural) for *table* are the tables we've experienced being labelled as *tables*. That said, we can of course identify and accept new instances as *tables* if they have sufficient

overlap with a sufficient number of stored exemplars (a phenomenon simulated by Regier, 2005, LEX exemplar model, discussed below).

10. LEX, like all exemplar models, instantiates forms and meanings using sets of features. This is an implementational necessity: the algorithm needs something on which to compute similarity between exemplars. However, it is important not to confuse implementational decisions with theoretical commitments. Few proponents of exemplar models would argue that learners *actually* represent forms and meanings in terms of sets of features (I, for one, would argue strongly against such a view). Features are simply a way to make the problem tractable, by converting learners' continuous sensory experience of the world into 0s and 1s that can be processed by computer algorithms.

11. Similar findings are observed in studies of over-regularization errors with English irregular verbs such as *sitted*, *goed* or *bringed* (e.g. Hartshorne & Ullman, 2006 [though only for girls]; Kidd & Lum, 2008; Marchman, 1997; Marchman, Wulfeck, & Weismer, 1999; Stemberger, 1993).

12. The rule-based fragment grammars model of O'Donnell (2015) outperforms two versions of an exemplar model, but the models are compared only on a highly abstract task which involves learning which of a number of pre-given inflectional transformations (e.g. /I/ → /æ/) applies to a given stem (e.g. *ring→rang*). Crucially, this means that the fragment grammars model (and O'Donnell's exemplar models, which use the same input representation) cannot be evaluated on the standard task of predicting the past-tense form of novel verbs on the basis of their phonological properties; hence I do not discuss it further here. I also exclude Yang's (2016) tolerance principle approach for the same reason.

13. My own view is that token frequency should be incorporated into exemplar models by having each individual exemplar vary slightly on certain features of pronunciation and/or meaning (as in Regier, 2005, LEX model). On this point, I agree with Skousen (1989, p. 54): 'Ultimately, the difference between type and token can be eliminated by specifying enough variables. By increasing the number of variables every token occurrence will also represent a single type.'

14. Hidden units allow for success by transforming the input space. Another way this can be done is by re-representing the input in a higher-dimensional space. For example, the two-layer discrimination learning model of Baayen and Hendrix (2016) successfully solves an abstract non-linearly-separable categorization problem in this way. Suppose black and white dots are arranged pseudo-randomly on a 10×10 X-Y grid, and the model's task is to predict each dot's category membership (i.e. colour) from its position. This is not possible if the model's input consists of one X and one Y unit each set to 1,2,3,4. . .10. However, if the input instead consists of 10 one-hot X units and 10 on-one-hot Y units (e.g. the representation [2,3] would instead be [0100000000,0010000000]) a two-layer discrimination model learns the solution almost perfectly.

15. In fact, some distinctions are marked additionally or solely by changes to the stem, such as vowel or consonant lengthening. The situation is similar for paradigms such as gender marking in Welsh (e.g. Thomas & Gathercole, 2007), which is never indicated using a standalone morpheme, but rather by a system of mutations: lenition (e.g. *p→b, t→d*), nasal mutation (e.g. *p→m, t→n*) and aspirate mutation (e.g. *p→f, t→θ*)

16. Baayen, Hendrix, and Ramscar (2013) describe a discriminative learning model that yields four-word *n*-gram effects without storing either sentence-level lexical strings (e.g. *A British provincial city*) or their component *n*-grams. However, because the model relies on a richly structured – albeit unordered – set of meaning representations (e.g. {A, BRITAIN, ISH, PROVINCE, IAL, CITY}), it is unclear whether it is really eschewing exemplar storage

altogether, or instantiating it at the level of stored semantic representations. Furthermore, eschewing storage at the lexical-phonological level is problematic, given findings (summarized in the section 'Phonetics and phonology') that listeners store fine-grained phonological representations of the utterances that they hear, rather than merely extracting semantic representations (the exemplar model of Walsh et al., 2010, shares this shortcoming).

17. Actually, there is one way in which these apparent opposites can be reconciled. Hartshorne et al. (2016) argue that what unifies SUBJECTS is not thematic roles but semantic prominence. Via a series of rating studies, Hartshorne et al. (2016) provide evidence that sentences like *John frightened Bill* 'encode an episode in which one entity causes another to experience an emotion' while sentences like *John feared Bill* 'encode an experiencer's habitual attitude about a target (leaving aside the question of what caused this attitude)' (p. 283). In other words, what unifies the SUBJECT role in *John frightened Bill* and *John feared Bill* (and, indeed, *John broke the plate*) is that both (/all) sentences are *about* John. This is a very different way of thinking about semantics to that found in current prototype accounts, but – I must acknowledge – one that does open up the possibility of a prototype SVO transitive construction (or, indeed, an account based on semantics–syntax linking rules). Of course, it is also compatible with an exemplar account under which analogy between utterances proceeds on the basis of semantic prominence rather than thematic-role type features or lower-level features such as animacy.

18. I do not discuss the exemplar model of Bod (2006, 2009) because it operates on the basis of fully-parsed sentence trees, and therefore – while interesting in many respects – does not represent a psychologically plausible account of child language acquisition.

19. In this respect, the model differs from a true exemplar model (and Regier, 2005, LEX) in that different instances of a word or utterance produced with (for example) different pitch are considered to be the same exemplar.

References

Abbot-Smith, K., & Behrens, H. (2006). How known constructions influence the acquisition of other constructions: The German passive and future constructions. *Cognitive Science*, *30*, 995–1026.

Abbot-Smith, K., Lieven, E., & Tomasello, M. (2001). What preschool children do and do not do with ungrammatical word orders. *Cognitive Development*, *16*, 679–692.

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

Abney, S. P. (1987). *The English noun phrase in its sentential aspect* (Doctoral dissertation). Massachusetts Institute of Technology, Cambridge.

Aguado-Orea, J., & Pine, J. M. (2015). Comparing different models of the development of verb inflection in early child Spanish. *PLoS ONE*, *10*(3), e0119613.

Akhtar, N. (1999). Acquiring basic word order: Evidence for data-driven learning of syntactic structure. *Journal of Child Language*, *26*, 339–356.

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

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

Ambridge, B., Bidgood, A., Pine, J. M., Rowland, C. F., & Freudenthal, D. (2016). Is passive syntax semantically constrained? Evidence from adult grammaticality judgment and comprehension studies. *Cognitive Science*, *40*, 1435–1459.

Ambridge, B., & Lieven, E. V. M. (2011). *Child language acquisition: Contrasting theoretical approaches*. Cambridge, UK: Cambridge University Press.

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

Ambridge, B., Pine, J. M., & Lieven, E. V. (2014). Child language acquisition: Why universal grammar doesn't help. *Language*, *90*, e53–e90.

Ambridge, B., Pine, J. M., Rowland, C. F., Jones, R. L., & Clark, V. (2009). A semantics-based approach to the 'no negative-evidence' problem. *Cognitive Science*, *33*, 1301–1316.

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

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

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

Ashby, F. G., & Rosedahl, L. (2017). A neural interpretation of exemplar theory. *Psychological Review*, *124*, 472–482.

Aylett, M., & Turk, A. (2006). Language redundancy predicts syllable duration and the spectral characteristics of vocalic syllable nuclei. *Journal of the Acoustical Society of America*, *119*, 3048–3058.

Baayen, R. H., & Hendrix, P. (2016). Two-layer networks, non-linear separation, and human learning. In M. Wieling, G. Bouma, & G. van Noord (Eds.), *From semantics to dialectometry* (pp. 1–11). Groningen, The Netherlands: University of Groningen Press.

Baayen, R. H., Hendrix, P., & Ramscar, M. (2013). Sidestepping the combinatorial explosion: An explanation of n-gram frequency effects based on naive discriminative learning. *Language and Speech*, *56*, 329–347.

Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science*, *19*, 241–248.

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition*, *11*, 211–227.

Bidgood, A., Pine, J. M., Rowland, C. F., & Ambridge, B. (submitted). Abstract knowledge does not efface detailed lexical representations: Evidence from children's and adults' comprehension and primed production of the English passive. *Cognitive Science*.

Blything, R. P., Ambridge, B., & Lieven, E. V. (2018). Children's acquisition of the English past-tense: Evidence for a single-route account from novel verb production data. *Cognitive Science*, *42*, 621–639.

Bock, K., & Loebell, H. (1990). Framing sentences. *Cognition*, *35*, 1–39.

Bod, R. (2006). Exemplar-based syntax: How to get productivity from examples. *The Linguistic Review*, *23*, 291–320.

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

Branigan, H. P., & Pickering, M. J. (2017). An experimental approach to linguistic representation. *Behavioral and Brain Sciences*, *40*, e282.

Brooks, P. J., & Tomasello, M. (1999). How children constrain their argument structure constructions. *Language*, *75*, 720–738.

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

Bybee, J. (1985). *Morphology: A study of the relation between meaning and form* (*Vol. 9*). Amsterdam, The Netherlands: John Benjamins.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, *10*, 425–455.

Bybee, J. (2010). *Language, usage and cognition*. Cambridge, UK: Cambridge University Press.

Bybee, J., & Scheibman, J. (1999). The effect of usage on degrees of constituency: The reduction of don't in English. *Linguistics*, *37*, 575–596.

Cabeza, R., & Moscovitch, M. (2013). Memory systems, processing modes, and components: Functional neuroimaging evidence. *Perspectives on Psychological Science*, *8*, 49–55.

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development*, *15*, 17–29.

Chandler, S. (2002). Skousen's analogical approach as an exemplar-based model of categorization. In R. Skousen, D. Lonsdale, & D. B. Parkinson (Eds.), *Analogical modeling: An exemplar-based approach to language* (pp. 51–105). Amsterdam, The Netherlands: John Benjamins.

Chandler, S. (2010). The English past tense: Analogy redux. *Cognitive Linguistics*, *21*, 371–417.

Chandler, S. (2017). The analogical modeling of linguistic categories. *Language and Cognition*, *91*, 52–87.

Chang, F. (2002). Symbolically speaking: A connectionist model of sentence production. *Cognitive Science*, *26*, 609–651.

Chang, F., Bock, K., & Goldberg, A. E. (2003). Can thematic roles leave traces of their places? *Cognition*, *90*, 29–49.

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

Chang, F., Dell, G. S., Bock, K., & Griffin, Z. M. (2000). Structural priming as implicit learning: A comparison of models of sentence production. *Journal of Psycholinguistic Research*, *29*, 217–230.

Chater, N., & Christiansen, M. H. (2018). Language acquisition as skill learning. *Current Opinion in Behavioral Sciences*, *21*, 205–208.

Childers, J. B., & Tomasello, M. (2001). The role of pronouns in young children's acquisition of the English transitive construction. *Developmental Psychology*, *37*, 739–748.

Chomsky, N. (1980). *Language and learning: The debate between Jean Piaget and Noam Chomsky* (M. Piatelli-Palmarini, Ed.). Cambridge, MA: Harvard University Press.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. Cambridge, MA: The MIT Press.

Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*, 1–72.

Conwell, E. (2018). Token frequency effects in homophone production: An elicitation study. *Language and Speech*, *61*, 466–479.

Corkery, M., Matusevych, Y., & Goldwater, S. (2019). Are we there yet? Encoder-decoder neural networks as cognitive models of English past tense inflection. *arXiv preprint arXiv:1906.01280*.

Craik, F. I., & Kirsner, K. (1974). The effect of speaker's voice on word recognition. *The Quarterly Journal of Experimental Psychology*, *26*, 274–284.

Crain, S., & Nakayama, M. (1987). Structure dependence in grammar formation. *Language*, *63*, 522–543.

Croft, W. (2000). *Explaining language change: An evolutionary approach* (Longman Linguistics Library). London, England: Longman.

Croft, W. (2001). *Radical construction grammar: Syntactic theory in typological perspective*. Oxford, UK: Oxford University.

Croft, W., & Cruse, A. (2004). *Cognitive linguistics*. Cambridge, UK: Cambridge University Press.

Culicover, P. W., Jackendoff, R., & Audring, J. (2017). Multiword constructions in the grammar. *Topics in Cognitive Science*, *9*, 552–568.

Dąbrowska, E. (2000). From formula to schema: The acquisition of English questions. *Cognitive Linguistics*, *11*, 83–102.

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

Dąbrowska, E. (2018). Experience, aptitude and individual differences in native language ultimate attainment. *Cognition*, *178*, 222–235.

Dąbrowska, E., & Szczerbinski, M. (2006). Polish children's productivity with case marking: The role of regularity, type frequency, and phonological diversity. *Journal of Child Language*, *33*, 559–597.

Daelemans, W. (2002). A comparison of analogical modeling of language to memory-based language processing. In D. Skousen, R. Lonsdale, & D. Parkinson (Eds.), *Analogical modeling* (pp. 157–179). Amsterdam, The Netherlands: John Benjamins.

Daelemans, W., & van den Bosch, A. (2010). Memory-based learning. In A. Clark, C. Fox, & S. Lappin (Eds.), *Handbook of computational linguistics and natural language processing* (pp. 154–178). Oxford, UK: Wiley-Blackwell.

Dawdy-Hesterberg, L. G., & Pierrehumbert, J. B. (2014). Learnability and generalisation of Arabic broken plural nouns. *Language, cognition and neuroscience*, *29*, 1268–1282.

Diesendruck, G., Markson, L., & Bloom, P. (2003). Children's reliance on creator's intent in extending names for artifacts. *Psychological Science*, *14*, 164–168.

Dittmar, M., Abbot-Smith, K., Lieven, E., & Tomasello, M. (2008). Young German children's early syntactic competence: A preferential looking study. *Developmental Science*, *11*, 575–582.

Dodson, K., & Tomasello, M. (1998). Acquiring the transitive construction in English: The role of animacy and pronouns. *Journal of Child Language*, *25*, 555–574.

Eddington, D. (2004). Issues in modeling language processing analogically. *Lingua*, *114*, 849–871.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179–211.

Elman, J. L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science*, *33*, 547–582.

Engelmann, F., Granlund, S., Kolak, J., Szreder, M., Ambridge, B., Pine, J., . . .Lieven, E. (2019). How the input shapes the acquisition of verb morphology: Elicited production and computational modelling in two highly inflected languages. *Cognitive Psychology*, *110*, 30–69.

Erk, K., & Padó, S. (2010, July). *Exemplar-based models for word meaning in context*. Paper presented at the Association for Computational Linguistics Conference Short Papers, Uppsala, Sweden.

Feldman, L., & Milin, P. (2017). If priming is graded rather than all-or-none, can reactivating abstract structures be the underlying mechanism? *Behavioral and Brain Sciences*, *40*, E287.

Fodor, J. A., & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*, 3–71.

Gertner, Y., Fisher, C., & Eisengart, J. (2006). Learning words and rules: Abstract knowledge of word order in early sentence comprehension. *Psychological Science*, *17*, 684–691.

Gillis, S., Daelemans, W., & Durieux, G. (2000). Lazy learning: Natural and machine learning of word stress. In P. Broeder & J. Murre (Eds.), *Models of language acquisition* (pp. 76–99). Oxford, UK: Oxford University Press.

Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. New York, NY: Oxford University Press.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1166–1183.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*, 251–279.

Goldwater, M. B., Tomlinson, M. T., Echols, C. H., & Love, B. C. (2011). Structural priming as structure-mapping: Children use analogies from previous utterances to guide sentence production. *Cognitive Science*, *35*, 156–170.

Granlund, S., Kolak, J., Vihman, V., Engelmann, F., Lieven, E. V., Pine, J. M., . . . Ambridge, B. (2019). Language-general and language-specific phenomena in the acquisition of inflectional noun morphology: A cross-linguistic elicited-production study of Polish, Finnish and Estonian. *Journal of Memory and Language*, *107*, 169–194.

Guenther, F. H., & Gjaja, M. N. (1996). The perceptual magnet effect as an emergent property of neural map formation. *The Journal of the Acoustical Society of America*, *100*, 1111–1121.

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

Hameroff, S., & Penrose, R. (2014). Consciousness in the universe: A review of the 'Orch OR' theory. *Physics of Life Reviews*, *11*, 39–78.

Hartshorne, J. K., O'Donnell, T. J., Sudo, Y., Uruwashi, M., Lee, M., & Snedeker, J. (2016). Psych verbs, the linking problem, and the acquisition of language. *Cognition*, *157*, 268–288.

Hartshorne, J. K., & Ullman, M. (2006). Why girls say 'holded' more than boys. *Developmental Science*, *9*, 21–32.

Hay, J. (2018). Sociophonetics: The role of words, the role of context, and the role of words in context. *Topics in Cognitive Science*, *10*, 696–706.

Hay, J., & Bresnan, J. (2006). Spoken syntax: The phonetics of giving a hand in New Zealand English. *The Linguistic Review*, *23*, 321–349.

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers*, *16*, 96–101.

Houston, D. M., & Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 1570–1582.

Houston, D. M., & Jusczyk, P. W. (2003). Infants' long-term memory for the sound patterns of words and voices. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 1143–1154.

Ibbotson, P., Theakston, A. L., Lieven, E. V., & Tomasello, M. (2012). Semantics of the transitive construction: Prototype effects and developmental comparisons. *Cognitive Science*, *36*, 1268–1288.

Iozzi, L., & Barbieri, M. S. (2009). Preschoolers' use of analogies in referential communication. *First Language*, *29*, 192–207.

Janssen, N., & Barber, H. A. (2012). Phrase frequency effects in language production. *PLoS ONE*, *7*(3), e33202.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 145–165). San Diego, CA: Academic Press.

Johnson, K. (2007). Decisions and mechanisms in exemplar-based phonology. In M. Sole, P. Speeter Beddor, & M. Ohala (Eds.), *Experimental approaches to phonology* (pp. 25–40). Oxford, UK: Oxford University Press.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 229–254). Amsterdam, The Netherlands: John Benjamins.

Jusczyk, P. W. (1993). From general to language-specific capacities: The WRAPSA model of how speech perception develops. *Journal of Phonetics*, *21*, 3–28.

Kellman, P. J. (2002). Perceptual learning. In R. Gallistel (Ed.), *Stevens' handbook of experimental psychology: Learning, motivation, and emotion* (3rd ed., Vol. 3, pp. 259–299). New York, NY: Wiley.

Kempson, R., & Gregoromichelaki, E. (2017). Action sequences instead of representational levels. *Behavioral and Brain Sciences*, *40*, E296.

Keuleers, E. (2008). *Memory-based learning of inflectional morphology* (Unpublished doctoral dissertation). Universiteit Antwerpen, Belgium.

Keuleers, E., Sandra, D., Daelemans, W., Gillis, S., Durieux, G., & Martens, E. (2007). Dutch plural inflection: The exception that proves the analogy. *Cognitive Psychology*, *54*, 283–318.

Kidd, E., & Lum, J. A. (2008). Sex differences in past tense overregularization. *Developmental Science*, *11*, 882–889.

Kirjavainen, M., Nikolaev, A., & Kidd, E. (2012). The effect of frequency and phonological neighbourhood density on the acquisition of past tense verbs by Finnish children. *Cognitive Linguistics*, *23*, 273–315.

Kirov, C., & Cotterell, R. (2018). Recurrent neural networks in linguistic theory: Revisiting Pinker and Prince (1988) and the past tense debate. *Transactions of the Association for Computational Linguistics*, *6*, 651–665.

Kjærbæk, L., dePont Christensen, R., & Basbøll, H. (2014). Sound structure and input frequency impact on noun plural acquisition: Hypotheses tested on Danish children across different data types. *Nordic Journal of Linguistics*, *37*, 47–86.

Krajewski, G., Theakston, A. L., Lieven, E. V., & Tomasello, M. (2011). How Polish children switch from one case to another when using novel nouns: Challenges for models of inflectional morphology. *Language and Cognitive Processes*, *26*, 830–861.

Krott, A., Schreuder, R., & Baayen, R. H. (2002). Linking elements in Dutch noun-noun compounds: Constituent families as analogical predictors for response latencies. *Brain and Language*, *81*, 708–722.

Krott, A., Schreuder, R., Baayen, R. H., & Dressler, W. U. (2007). Analogical effects on linking elements in German compounds. *Language and Cognitive Processes*, *22*, 25–57.

Krug, M. (1998). String frequency. A cognitive motivating factor in coalescence, language processing, and linguistic change. *Journal of English Linguistics*, *26*, 286–320.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*, 22–44.

Kuhl, P. K. (1991). Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, *50*, 93–107.

Kunnari, S., Savinainen-Makkonen, T., Leonard, L. B., Mäkinen, L., Tolonen, A. K., Luotonen, M., & Leinonen, E. (2011). Children with specific language impairment in Finnish: The use of tense and agreement inflections. *Journal of Child Language*, *38*, 999–1027.

Labov, W. (1963). The social motivation of a sound change. *Word*, *19*, 273–309.

Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, *3*, 299–321.

Langacker, R. W. (1988). A usage-based model. In B. Rudzka-Ostyn (Ed.), *Topics in cognitive linguistics* (pp. 127–161). Amsterdam, The Netherlands: John Benjamins.

Langacker, R. W. (2009). *The emergence of structure from usage*. Unpublished manuscript, University of California, San Diego.

Lent, R., Azevedo, F. A., Andrade-Moraes, C. H., & Pinto, A. V. (2012). How many neurons do you have? Some dogmas of quantitative neuroscience under revision. *European Journal of Neuroscience*, *35*, 1–9.

Leonard, L. B., Caselli, M. C., & Devescovi, A. (2002). Italian children's use of verb and noun morphology during the preschool years. *First Language*, *22*, 287–304.

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech*, 6, 172–187.

Lieven, E., Salomo, D., & Tomasello, M. (2009). Two-year-old children's production of multi-word utterances: A usage-based analysis. *Cognitive Linguistics*, 20, 481–507.

Liittschwager, J. C., & Markman, E. M. (1994). Sixteen- and 24-month-olds' use of mutual exclusivity as a default assumption in second-label learning. *Developmental Psychology*, 30, 955–968.

Love, B. C. (2013). Categorization. In K. N. Ochsner & S. M. Kosslyn (Eds.), *Oxford handbook of cognitive neuroscience* (pp. 342–358). Oxford, UK: Oxford University Press.

Mack, M. L., Preston, A. R., & Love, B. C. (2013). Decoding the brain's algorithm for categorization from its neural implementation. *Current Biology*, 23, 2023–2027.

MacWhinney, B. (2001). Emergentist approaches to language. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 449–470). New York, NY: John Benjamins.

Mahowald, K., James, A., Futrell, R., & Gibson, E. (2016). A meta-analysis of syntactic priming in language production. *Journal of Memory and Language*, 91, 5–27.

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

Marchman, V. A. (1997). Children's productivity in the English past tense: The role of frequency, phonology and neighborhood structure. *Cognitive Science*, 21, 283–304.

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

Marchman, V. A., Wulfeck, B., & 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, 206–219.

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

Matthews, D., & Bannard, C. (2010). Children's production of unfamiliar word sequences is predicted by positional variability and latent classes in a large sample of child-directed speech. *Cognitive Science*, 34, 465–488.

Matthews, D., Lieven, E., Theakston, A. L., & Tomasello, M. (2004). The role of frequency in the acquisition of English word order. *Cognitive Development*, 20, 121–136.

Matthews, D., Lieven, E., Theakston, A. L., & Tomasello, M. (2007). French children's use and correction of weird word orders: A constructivist account. *Journal of Child Language*, 34, 381–409.

Mattys, S. L., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition*, 78, 91–121.

McClure, K., Pine, J. M., & Lieven, E. V. M. (2006). Investigating the abstractness of children's early knowledge of argument structure. *Journal of Child Language*, 33, 693–720.

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

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, 85, 207–238.

Milin, P., Divjak, D., Dimitrijević, S., & Baayen, R. H. (2016). Towards cognitively plausible data science in language research. *Cognitive Linguistics*, 27, 507–526.

Minsky, M., & Papert, S. (1969). *Perceptrons: An introduction to computational geometry*. Cambridge, MA: The MIT Press.

Moscovitch, M. (1992). Memory and working-with-memory: A component process model based on modules and central systems. *Journal of Cognitive Neuroscience*, *4*, 257–267.

Mudrow, M. (2002). Version spaces, neural networks, and analogical modeling. In R. Skousen, D. Lonsdale, & D. B. Parkinson (Eds.), *Analogical modeling an exemplar-based approach to language* (pp. 225–264). Amsterdam, The Netherlands: John Benjamins.

Nakisa, R., Plunkett, K., & Hahn, U. (2000). Single- and dual-route models of inflectional morphology. In P. Broeder & J. Murre (Eds.), *Models of language acquisition* (pp. 201–222). Oxford, UK: Oxford University Press.

Newmeyer, F. J. (2003). Grammar is grammar and usage is usage. *Language*, *79*, 682–707.

Ninio, A. (2018). Learning to produce complement predicates with shared semantic subjects. *First Language*, *38*, 399–418.

Nosofsky, R. M. (1990). Relations between exemplar similarity and likelihood models of classification. *Journal of Mathematical Psychology*, *34*, 393–418.

Nosofsky, R. M. (2016). An exemplar-retrieval model of short-term memory search: Linking categorization and probe recognition. In B. H. Ross (Ed.), *The psychology of learning and motivation* (*Vol. 65*, pp. 47–84). New York, NY: Academic Press.

Nosofsky, R. M., Little, D. R., & James, T. W. (2012). Activation in the neural network responsible for categorization and recognition reflects parameter changes. *Proceedings of the National Academy of Sciences, USA*, *109*, 333–338.

O'Donnell, T. J. (2015). *Productivity and reuse in language: A Theory of linguistic computation and storage*. Cambridge, MA: The MIT Press.

O'Grady, W. (2017). The syntax of priming. *Behavioral and Brain Sciences*, *40*, E306.

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 309–328.

Perruchet, P. (2018). What mechanisms underlie implicit statistical learning? Transitional probabilities versus chunks in language learning. *Topics in Cognitive Sciences*, *11*, 520–535.

Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency effects and the emergence of linguistic structure* (pp. 137–157). Amsterdam, The Netherlands: John Benjamins.

Pierrehumbert, J. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology VII* (pp. 101–140). Berlin, Germany: Mouton de Gruyter.

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

Pinker, S. (1987). The bootstrapping problem in language acquisition. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 399–442). Hillsdale, NJ: Lawrence Erlbaum.

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

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

Pinker, S., & Ullman, M. T. (2002). The past and future of the past tense. *Trends in Cognitive Sciences*, *6*, 456–463.

Pluymaekers, M., Ernestus, M., & Baayen, R. (2005). Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica*, *62*, 146–159.

Port, R. F., & Leary, A. P. (2005). Against formal phonology. *Language*, *81*, 927–964.

Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology*, *83*(2, Pt. 1), 304–308.

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

Prince, A., & Smolensky, P. (2008). *Optimality theory: Constraint interaction in generative grammar*. Hoboken, NJ: John Wiley.

Quiroga, R. Q. (2017). How do we recognize a face? *Cell*, *169*, 975–977.

Ramscar, M., Dye, M., & Klein, J. (2013). Children value informativity over logic in word learning. *Psychological Science*, *24*, 1017–1023.

Ramscar, M., Dye, M., & McCauley, S. M. (2013). Error and expectation in language learning: The curious absence of mouses in adult speech. *Language*, *89*, 760–793.

Ramscar, M., & Port, R. (2015). Categorization (without categories). In E. Dąbrowska & D. Divjak (Eds.), *Handbook of cognitive linguistics* (pp. 75–99). The Hague, The Netherlands: De Gruyter Mouton.

Ramscar, M., & Port, R. F. (2016). How spoken languages work in the absence of an inventory of discrete units. *Language Sciences*, *53*, 58–74.

Ramscar, M., Sun, C. C., Hendrix, P., & Baayen, R. H. (2017). The mismeasurement of mind: Life-span changes in paired-associate-learning scores reflect the 'cost' of learning, not cognitive decline. *Psychological Science*, *28*, 1171–1179.

Räsänen, S. H. M., Ambridge, B., & Pine, J. M. (2014). Infinitives or bare stems? Are English-speaking children defaulting to the highest-frequency form? *Journal of Child Language*, *41*, 756–779.

Räsänen, S. H. M., Ambridge, B., & Pine, J. M. (2016). An elicited-production study of inflectional verb morphology in child Finnish. *Cognitive Science*, *40*, 1704–1738.

Ravid, D., & Farah, R. (1999). Learning about noun plurals in early Palestinian Arabic. *First Language*, *19*, 187–206.

Reber, A. S. (1969). Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology*, *81*, 115–119.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118*, 219–235.

Regier, T. (2005). The emergence of words: Attentional learning in form and meaning. *Cognitive Science*, *29*, 819–865.

Rissman, L., & Majid, A. (2019). Thematic roles: Core knowledge or linguistic construct? *Psychonomic Bulletin & Review*. Advanced online publication. doi:10.3758/s13423-019-01634-5

Roediger, H. L., & McDermott, K. B. (1993). Implicit memory in normal human subjects. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (Vol. 8, pp. 63–131). Amsterdam, The Netherlands: Elsevier.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573–605.

Rowland, C. F., Chang, F., Ambridge, B., Pine, J. M., & Lieven, E. V. M. (2012). The development of abstract syntax: Evidence from structural priming and the lexical boost. *Cognition*, *125*, 49–63.

Rubenstein, H., Lewis, S. S., & Rubenstein, M. A. (1971). Evidence for phonemic recoding in visual word recognition. *Journal of Verbal Learning and Verbal Behavior*, *10*, 645–657.

Rubino, R. B., & Pine, J. M. (1998). Subject-verb agreement in Brazilian Portuguese: What low error rates hide. *Journal of Child Language*, *25*, 35–59.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. Cambridge, MA: The MIT Press.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-olds. *Science*, *274*, 1926–1928.

Sakas, W. G., & Fodor, J. D. (2012). Disambiguating syntactic triggers. *Language Acquisition*, *19*, 83–143.

Savage, C., Lieven, E., Theakston, A., & Tomasello, M. (2003). Testing the abstractness of children's linguistic representations: Lexical and structural priming of syntactic constructions in young children. *Developmental Science*, *6*, 557–567.

Savage, C., Lieven, E., Theakston, A., & Tomasello, M. (2006). Structural priming as implicit learning in language acquisition: The persistence of lexical and structural priming in 4-year-olds. *Language Learning and Development*, *2*, 27–50.

Savičiūtė, E., Ambridge, B., & Pine, J. M. (2018). The roles of word-form frequency and phonological neighbourhood density in the acquisition of Lithuanian noun morphology. *Journal of Child Language*, *45*(3), 641–672.

Schatz, T., Feldman, N. H., Goldwater, S., Cao, X., & Dupoux, E. (submitted). *Early phonetic learning without phonetic categories—Insights from machine learning*. Retrieved from doi: 10.31234/osf.io/fc4wh

Shariatmadari, D. (2015, April 20). Why have we got it in for the glottal stop? *The Guardian*. Retrieved from www.theguardian.com/commentisfree/2015/apr/30/why-have-we-got-it-in-for-glottal-stop

Singh, L., Bortfeld, H., & Morgan, J. L. (2001). Effects of variability on infant word recognition. In B. Skarabela, S. Fish, & A. H.-J. Do (Eds.), *Proceedings of the 26th Boston University conference on language development* (pp. 608–619). Somerville, MA: Cascadilla Press.

Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. J. B. (2011). Seeing a phrase 'time and again' matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 776–784.

Skousen, R. (1989). *Analogical modeling of language*. New York, NY: Springer.

Skousen, R. (2005). Quantum analogical modeling: A general quantum computing algorithm for predicting language behavior. *arXiv preprint quant-ph/0510146*.

Smith, D. J., & Minda, J. P. (2000). Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 3–27.

Smoke, K. (1932). An objective study of concept formation. *Psychological Monographs*, *42*, 1–46.

Snider, N. (2008). *An exemplar model of syntactic priming* (Unpublished doctoral dissertation). Stanford University, CA.

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

Sosa, A. V., & Stoel-Gammon, C. (2012). Lexical and phonological effects in early word production. *Journal of Speech, Language, and Hearing Research*, *55*, 596–608.

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, *388*, 381–382.

Standing, L. (1973). Learning 10000 pictures. *Quarterly Journal of Experimental Psychology*, *25*, 207–222.

Stemberger, J. (1993). Vowel dominance in over-regularizations. *Journal of Child Language*, *20*, 503–521.

Stoel-Gammon, C. (2011). Relationships between lexical and phonological development in young children. *Journal of Child Language*, *38*, 1–34.

Tatsumi, T., Ambridge, B., & Pine, J. M. (2018). Disentangling effects of input frequency and morphophonological complexity on children's acquisition of verb inflection: An elicited production study of Japanese. *Cognitive Science*, *42*, 555–577.

Theakston, A. L., Ibbotson, P., Freudenthal, D., Lieven, E. V., & Tomasello, M. (2015). Productivity of noun slots in verb frames. *Cognitive Science*, *39*, 1369–1395.

Theakston, A. L., Lieven, E. V., Pine, J. M., & Rowland, C. F. (2001). The role of performance limitations in the acquisition of verb-argument structure: An alternative account. *Journal of Child Language*, *28*, 127–152.

Thomas, E. M., & Gathercole, V. C. M. (2007). Children's productive command of grammatical gender and mutation in Welsh: An alternative to rule-based learning. *First Language*, *27*, 251–278.

Thornton, R. (2012). Studies at the interface of child language and models of language acquisition. *First Language*, *32*, 281–297.

Thorpe, W. H. (1956). *Learning and instinct in animals*. Cambridge, MA: Harvard University Press.

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

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

Tremblay, A., & Baayen, R. H. (2010). Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 151–173). London, England: The Continuum International Publishing.

Ullman, M. T. (2015). The declarative/procedural model: A neurobiological model of language learning, knowledge, and use. In G. Hickok & S. L. Small (Eds.), *Neurobiology of language* (pp. 953–968). Amsterdam, The Netherlands: Elsevier.

Valian, V. (2014). Arguing about innateness. *Journal of Child Language*, *41*(Suppl. 1), 78–92.

Van Noord, R., & Spenader, J. K. (2015). Modeling the learning of the English past tense with memory-based learning. *Computational Linguistics in the Netherlands Journal*, *5*, 65–80.

Walsh, M., Möbius, B., Wade, T., & Schütze, H. (2010). Multilevel exemplar theory. *Cognitive Science*, *34*, 537–582.

Wexler, K. (1998). Very early parameter setting and the unique checking constraint: A new explanation of the optional infinitive stage. *Lingua*, *106*, 23–79.

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

Zaki, S. R., Nosofsky, R. M., Stanton, R. D., & Cohen, A. L. (2003). Prototype and exemplar accounts of category learning and attentional allocation: A reassessment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 1160–1173.

Ziegler, J., Goldberg, A., & Snedeker, J. (2018, March). *Passive priming requires function word overlap*. Poster presented at the 31st Annual Meeting of the CUNY Conference on Human Sentence Processing, Davis, CA.

Ziegler, J., & Snedeker, J. (2018). How broad are thematic roles? Evidence from structural priming. *Cognition*, *179*, 221–240.