Snyder, William (in press) Compound word formation. In Jeffrey Lidz, William Snyder, and Joseph Pater (eds.) *The Oxford Handbook of Developmental Linguistics*. Oxford: Oxford University Press.

CHAPTER 6

Compound Word Formation

William Snyder

Languages differ in the mechanisms they provide for combining existing words into new, "compound" words. This chapter will focus on two major types of compound: synthetic -ER compounds, like English *dishwasher* (for either a human or a machine that washes dishes), where "-ER" stands for the crosslinguistic counterparts to agentive and instrumental *-er* in English; and endocentric bare-stem compounds, like English *flower book*, which could refer to a book about flowers, a book used to store pressed flowers, or many other types of book, as long there is a salient connection to flowers. With both types of compounding we find systematic cross-linguistic variation, and a literature that addresses some of the resulting questions for child language acquisition. In addition to these two varieties of compounding, a few others will be mentioned that look like promising areas for coordinated research on cross-linguistic variation and language acquisition.

6.1 Compounding—A Selective Review

6.1.1 Terminology

The first step will be defining some key terms. An unfortunate aspect of the linguistic literature on morphology is a remarkable lack of consistency in what the "basic" terms are taken to mean. Strictly speaking one should begin with the very term "word," but as Spencer (1991: 453) puts it, "One of the key unresolved questions in morphology is, 'What is a word?'."

Setting this grander question to one side, a word will be called a "compound" if it is composed of two or more other words, and has approximately the same privileges of occurrence

within a sentence as do other word-level members of its syntactic category (N, V, A, or P). A given compound will be called "synthetic" if it contains morphemes corresponding to both a verb and a VP-internal argument of the verb. Thus, *dishwasher* is synthetic because it corresponds to the VP [*wash(es) dish(es)*].[1]

A compound will be called "endocentric" if it contains a "head" morpheme that determines its morphosyntactic features and general semantic type. In *flower book*, for example, the head is *book*. Hence the compound is a noun (like *book*), and names a type of book. Alternatively, a compound will be "exocentric" if there is an understood head that is not pronounced. For example, in the English exocentric compound *pick-pocket*, which means a person who picks pockets, the understood head would be *person*. Still another possibility is that a compound is "doubly headed," as in Spanish *hombre lobo*, literally 'man wolf,' meaning something that is simultaneously a man and a wolf (i.e. a werewolf). In this particular example the plural form is *hombres lobos* 'men wolves,' which makes it especially clear that the two nouns are both functioning as heads. (The word *hombre lobo* is also an example of an "appositional" compound, because it is composed of two Ns that are both possible descriptions of the individuals named by the compound.)

The term "bare stem" will be used for any form that (i) could be used as an independent word (or at least, could be so used after the addition of inflectional morphology), and (ii) is the form that inflectional morphology would combine with, but (iii) does not yet bear any inflection. Thus, the English word *plum* is a bare stem. The regular plural suffix -s can be attached directly to this form to make *plum-s*, which is no longer a bare stem.

---

[1] Note that some authors refer to non-synthetic compounds (including what will here be called "bare-stem compounds") as "root compounds."

An example of a "bare-stem compound" is the English word *worm cans* (e.g. 'tin cans used for storing fishing bait'), where the modifier *worm* is a bare stem (cf. *worm-s*). In contrast to English, many languages require a morphosyntactically more complex structure, involving more than bare stems, in order to express a comparable meaning. For example, French would need to substitute an expression like *boite aux verres* ('can for-the worm-s'), which includes a prepositional phrase. Following Bauer (1978), this type of expression will be called a "compound phrase."

Another example involving a structure more complex than a bare-stem compound is the Hebrew expression *kufsat tulaAim* 'can-of worm.' Here the head N *kufsat* appears in a form known as the "construct state." For each noun in Hebrew there exists an inflected, construct-state form, which is similar to a genitive-case form but involves the word that refers to the (literal or metaphorical) possession, not the possessor. Note that in many cases the construct form happens to be homophonous with the bare-stem form, but unlike bare-stem compounds, construct-state expressions permit the syntax to see and manipulate the words that are contained within them. For example, two nouns in the construct state can share a single modifier in a way that is impossible in bare-stem compounds (cf. English **a black -bird and -board*, for 'a blackbird and blackboard').[3]

6.1.2 Creativity

---

[3] Along with compound phrases and constructs, a number of further alternatives to bare-stem compounding can be seen in Snyder (2001: Appendix A). The seminal work on the Hebrew construct state, with its mixture of phrasal and word-like properties, is Borer (1988). A brief synopsis of some of her key evidence and arguments can be found in Spencer (1991: 449–50.)

The next step will be to highlight a distinction that deserves far greater attention than it currently receives: A process of word formation either is, or is not, creative. Here "creative" is being used in the sense of "the creative aspect of human language": the fact that we can freely create a new sentence, potentially one that has never been used before, and reasonably expect a listener to understand it. A specific type of compound word formation is likewise creative, in a given language, if it is available for automatic, impromptu use whenever a new word is needed to fit the occasion.[4]

In contrast, it can happen that the lexicon of a language includes numerous compound words of a given type, even though there is no corresponding process of compound word formation available for creative use. This is the situation of bare-stem endocentric compounding in French. In a detailed analysis of French-speakers' use of compounds, Bauer (1978: especially 83–84) highlights the fact that French-speakers only create new endocentric compounds when they are deliberately trying to coin a new word. He also notes that endocentric compounds in the Germanic languages, like English *frog man* (in the sense of an underwater diver), sometimes

---

[4] Spencer (1991: 322–4), referring to creativity as "productivity," comments on morphologists' surprising lack of attention to cross-linguistic variation in the creativity of compounding: "In general, the problem of productivity (in its various senses) is not raised in the theoretical discussion of root compounds, . . ." (1991: 322); "One particularly interesting, but largely unexplored, question is what governs differences in root compounding between languages" (1991: 323–4); "Finally, not all the root compound types found in English can properly be said to be productive. The question of what governs productivitity and whether it's necessary to distinguish productive from non-productive compounding types has not been discussed extensively in the theoretical literature" (1991: 324).

give rise to "calques" like *homme grenouille* (lit. 'man frog') in French. Crucially, where *frog man* in English could mean any number of things, depending on the context (e.g. a person who sells frogs), in French it can only mean a diver.

Unfortunately the standard term in the morphology literature, "productive," is highly ambiguous: Some authors use "productive" to mean creative, while others, for example, define the productivity of a given type of compound in terms of the number of instances they find listed in a dictionary. An example of the confusion that can result is discussed in Snyder (2012: 281–282). [6]

6.1.3 Cross-linguistic variation

Thus, for any specific type of compound word formation, a first point of cross-linguistic variation is whether it is creative. A second major point of variation is whether it can apply recursively. For example in English, the endocentric compound *flower book* may be handed back as an input to the process of endocentric compounding, to obtain a new compound like [[*flower*

---

[6] Bauer (2001) provides a book-length treatment of the various notions of "productivity," and argues that a distinction is crucially needed between what he calls "availability" and "profitability." Availability refers to whether a process of word-formation is available to the speaker whenever the need for it might arise. Thus, availability corresponds to what I am calling "creativity," and is likewise taken to be a binary property.

book] *collection*]. Namiki (1994) proposes the generalization that endocentric compounding is creative, in a given language, if and only if it is recursive.[7]

In Snyder (1999) I built on this idea and proposed that evidence of recursion in endocentric compounding may be significant for the child learning a language, as an indication that endocentric compounding is creative. Roeper and Snyder (2004) took the idea further, and suggested that evidence of recursion might be used by the learner much more broadly, as an indication that any particular grammatical process is creative.

Namiki's generalization has held up well in the years since it was proposed, but the details of recursive endocentric compounding have turned out to be a bit more complex. As discussed in Roeper and Snyder (2005), Swedish is a language with endocentric bare-stem compounding similar to that found in English, and much as in English, the process is both creative and recursive. Yet, the recursion in Swedish compounds is subject to a restriction: Any left-branching node that is itself branching must be followed by the "linking element" *-s-* (Josefsson 1997: 60). This can be seen in (1) (based on Roeper and Snyder's example 5):

(1) a. barn [bok klub] 'child [book club]', or 'book club for children'

b. *[barn bok] klub '[child book] club', or 'club for (collectors of) children's books'

c. [barn bok]-s klub '[child book]'s club', or 'club for (collectors of) children's books'

Here we see that morphologically simplex modifiers, as in (1a), can combine with either simplex or complex right branches, and no linker occurs. Yet a morphologically complex (i.e. branching)

---

[7] Cross-linguistic variation in the creativity (and recursivity) of bare-stem endocentric compounding will be taken up in Section 6.3 under the heading of "The Compounding Parameter."

modifier, as in (1b,c), triggers the insertion of an *-s-* between the complex modifier (*barn bok*) on the left and its corresponding head (*klub*).

In (1c) the Swedish element *-s-* has been translated into English as the "Saxon genitive," *-'s*, but the correspondence is only approximate. For one thing the more direct English translation of (1c), namely *children's book club*, has the element in a different location than Swedish. More to the point, there is no general need for a genitive marker in English left-branching compounds, as can be seen from familiar examples like *[student film] committee*.

German likewise has endocentric bare-stem compounding that is both creative and recursive, and like Swedish it sometimes employs a linking element that is realized as *-s-*. Yet the distribution of this element is clearly different than in Swedish, because it is not required in left-branching compounds. For example it is absent from the direct counterpart to (1c), *Kinderbuchclub* (literally 'children book club'), on the relevant interpretation, 'club for people interested in children's books.' The rules governing the distribution of linking elements in the different Germanic languages, and children's acquisition of these rules, would be an interesting pair of topics to investigate in tandem.[8]

---

[8] While the German linker *-s-* is homophonous with the *-s* ending of the German genitive singular, it is clearly distinct. Genitive *-s* is restricted to masculine and neuter nouns, while linking *-s-* also occurs with feminine nouns (e.g. *Hochzeitstorte* 'wedding-s-cake,' where *Hochzeit* 'wedding' is feminine). Nübling and Szczepaniak (2008), who examined a one-billion-word sample of written German, report that (i) in most cases, German compounds do not contain a linker; (ii) German provides a variety of linking elements, but most of them (e.g. *-es-*, *-er-*, *-e-*, *-ens-*) are restricted to specific, lexicalized compounds; and (iii) only *-s-* and *-n-* are "productive" linkers in present-day German. They propose a phonological account: "[T]he occurrence of

Another point of cross-linguistic variation in endocentric compounding is the linear position of the head. The Germanic languages all follow a Right-hand Head Rule (cf. Williams 1981): Whenever the structure of a compound branches, the daughter-node on the right is the head. Williams proposed his version of this rule as a cross-linguistic universal, but its universality appears to be contradicted by languages like Khmer and Thai, which are the mirror image of English: They permit creative, endocentric, bare-stem compounding, but with the head on the left. In other words, exactly where an English speaker might create a compound like [[*fish sauce*] *bottle*], a Khmer speaker can create the compound [*daup* [*tuk trey*]] '[bottle [sauce fish]]' to mean the same thing. Hence, in addition to determining that the target language has creative (and fully recursive), endocentric bare-stem compounding, a child acquiring English or Khmer will need to determine whether an endocentric compound's head appears to the left or the right of a modifier.

Quite interestingly, Beard (1995, 1996) has observed the following pattern across languages: The order of the modifier and the head in an endocentric compound is almost always the same as the (default) order of an attributive adjective and the noun that it modifies within a noun phrase. For example, in English both an attributive adjective in a noun phrase (*the red house*), and the modifier in an (endocentric) N-N compound (*flower book*), appear on the left. In

---

linking elements, above all *-s-*, strongly depends on the quality of the preceding pword: The more distant it is from the ideal pword (i.e. a trochee, the second syllable containing [ə] or [ɐ]), the more probable a linking *-s-* becomes." Note too that *Kinderbuchclub* contains an *-er* that could be a plural marker or a linking element; either way it brings the pword *Kind* 'child' much closer to Nübling and Szczepaniak's ideal. (The *-s-* in *Hochzeitstorte*, however, may be present to prevent adjacency of the two [t] sounds.)

Khmer the attributive adjective (*pteah krahom*, literally 'house red'), like the modifier in an endocentric compound, appears on the right.

Yet, there do exist exceptions: In a survey of more than 50 languages, Beard found that the pattern was fairly robust, but he discovered several apparent counterexamples among the languages of the Americas. Of particular concern are Dakota, Kiowa, Koasti, Navajo, and the Michoacán variety of Nahuatl, all of which require attributive adjectives to appear on the right in a noun phrase, but require the modifier to appear on the left in a compound.[10]

In Snyder (2012: 286–287) I noted that the same problem arises for Basque, and suggested one type of solution: Perhaps Beard's Generalization really only applies to bare-stem

---

[10] Bauer (2001) conducted an independent cross-linguistic survey on the position of modifiers in nominal compounds, and the position of attributive adjectives in noun phrases. Unlike Beard he reports only a weak association, and suggests it might be stronger (though still imperfect) if one looks at possessors rather than adjectives. This outcome may be due to Bauer's methodological choices: (i) heavy reliance on descriptive grammars, rather than native speakers (cf. discussion of West Greenlandic and Turkana, 2001: 698); and (ii) extreme inclusiveness (i.e. when in doubt, call it a compound), as seen in his decision to ignore differences in creativity (cf. discussion of Hebrew, 2001: 698). Bauer arrives at some surprising conclusions: (i) endocentric compounds are all but universal (2001: 697: "This is the majority pattern . . . in the languages of the world, and there are very few languages which do not have compounds of this type"); and (ii) the order of head and modifier is highly unstable (2001: 697: "Although it might be expected that [the ordering of modifier and head in compounds] would be fixed in any individual language, that is the case only in about half of my sample from any of the areas used"). For present purposes I will instead rely on Beard's work.

compounding. In this case Basque would no longer be a counterexample, because the type of endocentric compounding it employs does not in fact contain a bare-stem modifier. While the modifier in a Basque compound is often homophonous with a bare stem, de Rijk (2008: 853–857) has demonstrated that in many other cases the two are not homophonous. Compounding in Basque is thus reminiscent of construct-state expressions in Hebrew.

More precisely, every Basque noun has a special morphological form that we might call "modificational," and for many nouns the modificational form is distinct from the bare-stem form (i.e. the form to which inflectional affixes attach). Moreover, in the Basque version of endocentric compounding, any noun that is not the head must appear in its modificational form. For example, Basque has a noun meaning 'human,' which is *gizon* in its bare-stem form. (Thus the ergative-indefinite form, for example, is *gizon* + e*k* = *gizonek* 'humans.') Yet, the modificational form is *giza-*, and this is the form that must be used in an N-N compound: *giza kuntza* 'human language.' According to de Rijk, the modificational form can usually be derived from the bare-stem form by applying one of a small number of morpho-phonological rules, but exactly which rule will apply to a particular noun is somewhat unpredictable. Hence, Basque does not in fact have a grammatical process of bare-stem compounding, and would not be a counterexample to the amended version of Beard's Generalization.[11]

---

[11] The linking elements of German compounds may be reminiscent of Basque modificational forms, because their likelihood of occurrence depends on the particular modifier, but they differ in important ways: (i) in German the linker is usually optional (except in certain lexicalized compounds, and after specific derivational suffixes); (ii) the linker has no effect on the bare-stem form that precedes it; and (iii) Nübling and Szczepaniak (2008) found that the likelihood of a linker is predictable from the modifier's phonological shape. Where Basque has a process

Interestingly, Beard (1996: 24) observes that in all four of Dakota, Kiowa, Koasti, and Navajo, a possessor noun (unlike an attributive adjective, but just like the modifier in a compound) precedes the head of its noun phrase.[12] This observation raises a possibility that to my knowledge has not been tested: Perhaps what Beard is glossing as bare-stem compounds in each of these "exception" languages will, on closer examination, turn out to be something akin to a construct-state expression in Hebrew, or an "izafet" construction in Turkish, where the head and/or the modifier is actually in a special, possession-related form.

As discussed by Spencer (1991: 314–319 and 449–450), both Hebrew constructs and Turkish izafets (especially the so-called "indefinite" izafets) share a number of properties with bare-stem compounds, but neither type of expression can be constructed out of bare stems. In fact, Spencer indicates that neither Turkish nor Hebrew has any "productive" (i.e. creative) bare-stem compounding whatsoever. Hence, it seems possible that tying Beard's Generalization more closely to bare-stem compounding (or perhaps to creative bare-stem compounding) would exclude the problem cases identified in (Beard 1996).[14]

---

combining a modificational form with a bare-stem head, German has bare-stem compounding plus the possibility of something akin to epenthesis.

[12] Beard does not clearly indicate the position of posessional Ns in Michoacán.

[14] Note that Beard was not specifically looking for bare-stem compounds. Hence, where Beard (1996: 22) glosses the Dakota expression *čhą-luí* as the bare-stem compound "tree skin" (meaning "bark"), for example, this does not necessarily signal a bare-stem compound. In fact, Beard's gloss for another of his nominal compounds contains a possessive form, and strongly resembles a Turkish indefinite izafet: *hociłí im-layki*, 'star its-dung' (= 'meteor'), from Koasati (1996: 21, ex.26b, emphasis mine).

Turning now to synthetic compounds, two important types are exemplified by (i) the English compound *dishwasher*, and (ii) the corresponding French compound *lave-vaisselle*, literally 'wash(es) dishware,' which functions as a noun meaning 'dishwasher' (usually in the sense of a dish-washing machine). Type (i) is creative in English, but unattested in the Romance languages. Type (ii), in contrast, is widely attested in the Romance languages, where it represents a creative process of word formation. English, however, allows only a few, lexicalized examples of this kind, such as *scarecrow* and *killjoy*.[16]

In fact, Beard (1995, 1996) argues that his generalization about head-modifier order in compounds can explain the heavy reliance on V-N compounding in Romance. The reasoning

---

Note also that for these problematic languages Beard sketches a different approach, in which (for example) the position of a nominal modifier relative to a deverbal nominal head in a compound is determined either by the position of a nominal modifier (i.e. argument) in relation to a V (the "input" category), or by the position of a nominal modifier in relation to an N (the "output" category), depending on a parameter-setting. In the problematic languages the "input" setting (unlike the "output" setting) yields appropriate word order. Interestingly, Lardiere (1998) reports that this input–output parameter can account for certain synthetic compounds she obtained from adult English-learners in an elicited production study.

[16] Somewhat closer to English *dishwasher* (type i) is the French phrasal compound *laveur de vaisselle* 'washer of dishware.' The head noun is *laver* 'to wash' plus agentive *-eur*, and a preposition *de* ('of') is inserted, probably for case reasons. The definite article *la* would ordinarily accompany *vaisselle*, but is omitted in *laveur de vaisselle*, just as it is in *lave-vaisselle*. The gender of a French compound like *lave-vaisselle* is consistently masculine (like *laveur*), even when the noun on the right-hand side is feminine (like *vaisselle*).

goes as follows. In Romance languages, the default position for an attributive adjective is normally to the right of the N heading the NP. Theferefore, by Beard's Generalization, the modifier in a compound should also be on the right side.

At the same time, according to Beard, any derivational suffix like English *-er*, or its counterpart *-eur* in French, is normally subject to two positional requirements: It must be affixed to a V, and it must appear at the rightmost edge of its word. A compound like **lav-eur vaisselle* is ruled out in French, because the stem *vaisselle* (which for present purposes is functioning as a "modifier" inside the compound word) needs to follow its head *laveur*, but if it does, the suffix *-eur* is no longer at the word's right edge. (Of course, suffixing *-eur* to *vaisselle* would solve the one problem but create another.) According to Beard, the solution in French and other Romance languages is usually to leave the morpheme *-eur* unpronounced: hence, *lave-vaisselle*. Beard suggests that V-N compounds in French are masculine precisely because the unpronounced suffix *-eur* is masculine.

I would add that this account might also explain why V-N compounding is not a creative process of word formation in English: Perhaps the suppression (i.e. non-pronunciation) of an overt morpheme like *-eur* is allowed only as a last resort. In English one can readily create a compound like *dishwasher* (i.e. the modifier *dish* and the suffix *-er* are on opposite sides of the V), hence there can be no suppression of *-er*, and a compound noun like **dish-wash* or **wash-dish* cannot be created (except perhaps as a conscious coinage).

6.2 The Acquisition of Synthetic –ER Compounding

6.2.1 What Must Be Acquired?

Turning now to acquisition, consider synthetic compounds like English *dishwasher* and French *lave-vaisselle*. What exactly does the child acquire? Based on the material in the preceding

section, the forms of synthetic compounding that are potentially part of the child's target language will be connected to several broader characteristics of the language.

Assuming (for expository purposes) that the specific proposals in the preceding section are entirely correct, some of the things the child will need to determine are whether the target language normally positions modifiers to the left or the right of a head noun, and whether derivational morphology (like an agentive affix) is prefixal or suffixal. If the answers are "Left, Suffixal" as in English, then it becomes possible that the language will have compounds like *dishwasher*, where the modifer *dish* and the suffix --ER are on opposite sides of the V.

Moreover, if suppression of the agentive suffix (as in French *lave-vaisselle*) is grammatically possible only as a last-resort operation, as I speculated at the end of Section 6.1.3, then it should not be available as a creative process of word formation in a Left-Suffixal language. Nonetheless, such a language could have forms like English *washer of dishes*, where the logical object *dishes* is outside the complex word, in a separate phrase. (Note too that a Right-Prefixal language like Swahili should have the option of allowing -ER style synthetic compounds, with a linear order of "ER-Wash-Dish.")

If the broader characteristics are "Right, Suffixal" as in French, compounds like *dishwasher* should be blocked, but V-N compounds like *lave-vaisselle* might exist in the language. Forms like French *laveur de vaisselle* ('wash-er of dishware') might also exist, because in this type of expression the agentive suffix *-eur* is adjacent to both the V and the right edge of its word. This is because the modifier *vaisselle* is now outside the word *laveur*, in a separate phrase. (Note that in a Left-Prefixal language—if such languages exist—an -ER synthetic compound should likewise be disallowed, since both the modifier and the prefix would

be competing for the left edge of the compound word. Such a language might use suppression of the derivational prefix to obtain a form like "Dish-Wash," the mirror image of *lave-vaisselle*.)

6.2.2 Puller-wagons?

To examine children's acquisition of synthetic compounding in English, Clark et al. (1986) ran an Elicited Production (EP) experiment on 48 children, aged 3–6 (12 children in each of four age groups). Suprisingly, they found that 3-year-olds sometimes produced forms like *puller-wagon* in place of the target *wagon-puller*. At first glance this suggests that children might, for a certain period, have a grammar that differs from adult English (and from other languages that have been examined) by allowing a derivational suffix (*-er*) to be separated from the right edge of its word by intervening material (i.e. *wagon*). It also suggests that the child's grammar at this point places "modifiers" (in terms of Beard's system) to the right of the head, at least in the case of synthetic compounds—even though there are no corresponding reports of children putting attributive adjectives after the noun (e.g. *the book blue is over there*), as Beard's proposals would lead us to expect. Indeed, Lardiere (1998: 298) presents these specific child-language findings as an especially serious empirical threat to Beard's Generalization.

Before sounding the alarm, however, we should carefully inspect the nature of Clark et al.'s evidence. Each research method has its own strengths and weaknesses, and these need to be kept firmly in mind when evaluating the results of any study. In the case of EP, one of the weaknesses is that a child may respond to uncertainty about the point of grammar being tested, or to excessive processing demands of the experimental task, by producing forms that are not genuinely permitted by her grammar. In places where we can perform side-by-side comparsions (cf. Snyder 2007: 96-104), this is a far more common occurrence in EP studies than in studies of Spontaneous Speech (SS). In SS it seems the child can much more easily confine herself to the

types of utterances that her current grammar actually allows. (On the other hand, if one is interested in a linguistic structure that is used only rarely, even by children who definitely know it, then EP can provide a vastly greater quantity of data than SS.)

In Clark et al.'s (1996) study, the experimenter asked the child questions of the form, "What could you call a girl who pulls wagons?" This may seem like a simple enough question, but the task of constructing the target form, *wagon-puller*, at the very least requires the child's language-processing system to identify the relevant portion of the prompt's hierarchical structure (i.e. [$_{VP}$ [$_{V}$ *pull-s*] [$_{DP}$ [$_{N}$ *wagon-s*]]]), remove all inflectional morphology (yielding [$_{VP}$ [$_{V}$ *pull*] [$_{NP}$ [$_{N}$ *wagon*]]]), add an agentive, category-changing suffix onto the verbal element (yielding [$_{NP}$ [$_{N}$ [$_{V}$ *pull*] *-er*] [$_{NP}$ *wagon*]]), and then extract the head ([$_{N}$ *wagon*]) of the complement phrase in order to left-adjoin it to the next-higher head in the structure, yielding the complex noun [$_{N}$ [$_{N}$ *wagon*] [$_{N}$ [$_{V}$ *pull*] *-er*]]. If a 3-year-old child were to become fatigued during all this computation, the occasional production of a derivationally intermediate form, like [$_{NP}$ *pull-er* [$_{NP}$ *wagon*]], would be unsurprising (and would constitute a performance error, not a signal that the child's grammar is non-target-like).

On the other hand, if an individual child consistently gave answers of the precise form seen in *puller-wagon*, then it would be quite important to test the hypothesis that this child (and probably other children as well) had temporarily adopted an incorrect grammar, indeed a grammar that Beard's approach to synthetic compounding would not have predicted possible. Evidence that children routinely adopted such a grammar (at least temporarily) during the acquisition process would support Clark's (1993: 150) proposal that this error-type is characteristic of a middle stage along the normal path to the adult-English system.

Yet, Clark et al. found no such children. Of the 48 children in their study, 33 exhibited a consistent response pattern (i.e. at least 75 percent use of a single pattern) on at least one of agent nouns or instrument nouns (12 nouns of each type were elicited from every child). Of the 33 consistent children, the younger ones (3–4 years) predominantly used bare-stem endocentric compounds (e.g. *water person* and *garbage machine*, as well as the less adult-like *hugger man* for a man who hugs, and *feeder machine* for a machine that feeds), while the older ones (5–6 years) predominantly used adult-like synthetic compounds of the form O V-*er* (e.g. *wall-builder*, *box-mover*). Crucially, not a single child met the 75 percent-consistency criterion with compounds of the form V-*er* O. This strongly suggests that the 3- and 4-year-olds who occasionally produced V-*er* O compounds were not doing so because they had temporarily adopted an incorrect grammar, in which "V-*er* O" was the correct way to build an -ER compound. Rather, the error most likely reflected the child's limited processing capacity.

6.2.3 Rat(*s)-eaters

Returning to the issue of exactly what information is acquired during the acquisition of synthetic compounding, note that so far the cross-linguistic generalizations from Section 6.1.3 have suggested strong possibilities, but not certainties, about the types of synthetic compounding that will be allowed if the target language has certain broader characteristics. This is because a type of compounding that is possible in principle might be blocked by independent lexical or grammatical characteristics of the language. In addition, while certain types of information, like the position of attributive adjectives relative to the noun they modify, should be readily available from any number of different sentence-types that occur frequently in child-directed speech, other essential information, like the specific derivational morphemes that can be used to create synthetic compounds, may be impossible to determine except from examples of actual synthetic

compounds. Hence, knowing the broader characteristics of the language may help the child to know what to look for, but cannot provide all the answers.

Given that simple memorization will be a large part of learning the realization of -ER in the child's target language, why should we suppose that a child is ever doing anything more than simple memorization? For example, how can one exclude the possibility that early on, the child is simply "muddling through" with superficial, non-grammatical strategies, like memorizing specific compounds that have occurred in the input, and modifying them only slightly if at all? How can we tell when the child is using the same, grammatical system that we attribute to the adult?

A very elegant answer is provided by Peter Gordon's seminal (1985) "rat-eater" experiment. The study was based on the observation that adult English-speakers usually find it unacceptable to include regular inflectional morphology on the noun inside a synthetic compound. For example, a monster who loves to eat rats can certainly be described as a *rat-eater*, but is quite unlikely to be called a *rats-eater*.[17] This contrast seems to disappear, however, if we use a noun whose plural form is irregular: a monster who loves to eat mice could be either a *mice-eater* or a *mouse-eater*.

Crucially, the child acquiring English seldom if ever hears irregular plurals inside compounds, and therefore receives little or no evidence to suggest that irregular plurals, in contrast to regular plurals, are generally acceptable in compounds. Gordon searched for five specific high-frequency nouns with irregular plurals (*mouse*, *man*, *tooth*, *foot*, *goose*) in the Kučera and Francis (1967) word count, which was based on a carefully constructed sample of

---

[17] Note, however, that regular plurals are sometimes acceptable within compounds, when certain conditions are met. This point will be discussed below.

approximately one million words of printed English. Within the sample, these particular nouns were used a total of 153 times as non-head elements in compounds of various types, but 150 of those uses employed the singular form. Hence, if young children do not yet have the adult grammar of English synthetic compounding, and are simply modeling their speech on frequent surface characteristics of their input, they should strongly prefer the singular form of both regular nouns and irregular nouns when used inside a synthetic compound. In contrast, if young children already follow the same grammar of synthetic compounding as do adult English-speakers, then it may be possible to elicit irregular nouns (in contrast to regular nouns) as either a singular or a plural inside a synthetic compound.

To test these predictions, Gordon ran an EP study on 33 3- to 5-year-olds (divided by age into three groups of 11). The task used a Cookie Monster puppet and a variety of small objects. Elicitation of *rat/rats-eater* or *mouse-/mice-eater* went roughly as follows:

Experimenter: Do you know who this is? . . . It's the Cookie Monster. Do you know what he likes to eat?

Child: Cookies.

Experimenter: Yes—and do you know what else he likes to eat?—He likes to eat all sorts of things . . . . [Experimenter shows a single toy rat / mouse.]

Experimenter: Do you know what this is?

Child: A rat. / A mouse.

Experimenter: [Showing four rats / mice] Here we have a bunch of . . . what?

Child: Rats. / Mice.

Experimenter: What do you call someone who eats X? [where X is the child's plural form]

Child: A rat-eater. / A rats-eater. // A mouse-eater. / A mice-eater.

Experimenter: Do you think Cookie Monster is a X-eater? [where X is whatever the child said]

Child: Yes!

Notice that the experimenter's prompt, "What do you call someone who eats rats/mice?," contains a plural form of the noun, and can be expected to bias the child very strongly towards using a plural in his or her response, whenever this is possible.

The experiment used the same five irregular nouns mentioned above (*mouse*, *man*, *tooth*, *foot*, *goose*), five matched regular nouns (*rat*, *baby*, *bead*, *hand*, *duck*), four pluralia tantum (*clothes*, *pants*, *(sun)glasses*, *scissors*) and another four matched regular nouns (*toy*, *shirt*, *shoe*, *knife*). Pluralia tantum, like irregular plurals, are acceptable as non-heads inside compounds.

The children at all ages made a clear distinction between the regular plurals on the one hand, and the irregular plurals and pluralia tantum on the other. Out of 297 opportunities (33 children × 9 regular items), the children produced in total only six regular plurals inside compounds (2.02 percent of the opportunities). Interestingly, five of these six were produced by 5-year-olds, 1 was produced by a 4-year-old, and none were produced by the 3-year-olds. Gordon speculates that the older children's metalinguistic awareness may have played a role.

In contrast, out of 165 opportunities (33 children × 5 irregular items), the children produced a total of 36 irregular plurals inside compounds (21.8 percent of the opportunities). With pluralia tantum, out of 132 opportunities (33 children × 4 pluralia tantum items) the children produced a total of 68 pluralia tantum inside compounds (51.5 percent of the opportunities). In most other responses to pluralia tantum items, and especially in responses to *glasses* and *scissors*, the childen used a morphologically singular form, as in *scissor-eater*, which is also possible for adult speakers. Thus, the findings provide powerful evidence that very young children are already using the same grammar of synthetic compounding as adults.[20]

---

[20] The contrasts between regulars versus irregulars and pluralia tantum were found to be statistically reliable using chi-square tests, where each subject was classified according to his or

Where then does the "*mice-eater* / **rats-eater*" contrast come from? The general answer is that it must somehow follow from Universal Grammar. The best specific answer available in the early 1980s, when Gordon began his research, was that of Paul Kiparsky, who had been the first to point out the phenomenon. Kiparsky's (1982) account was based on Level Ordering (therefore building on earlier work of Siegel 1979 and Allen 1978), within the framework of Lexical Phonology. Roughly speaking, the idea was that there are three distinct, ordered levels of word formation processes in the lexicon. Level 1 is the access-point for stored forms, including those with irregular inflection. Compounding and regular derivational morphology (i.e. derivational affixation that does not alter the phonological form of the base) occur at Level 2. Regular inflectional morphology is inserted at Level 3. Thus, Level 1 can supply *eat* and any of the forms *mouse*, *mice*, and *rat* to Level 2. There, the verb *eat* can be combined with the

---

her dominant response pattern. Gordon (1985: 85) indicates there was only a single error of the *puller-wagon* type in his entire study. (The child's age is not specified.) Hence, even the youngest group of children in his study (N = 11, age 3;02 to 4;00, 18 items per child) made fewer such errors (at most 1; i.e. 0.5% of opportunities) than did the 3-year-olds (N = 12, age 3;00 to 3;10, 24 items per child) in Clark et al.'s (1986) study (10 errors; i.e. 3.5% of opportunities). This may result from Gordon's holding the "_-eater" frame constant, and thereby reducing the processing demands on the child.

Clahsen et al. (1992) successfully replicated Gordon's study in German, where nominal inflection is more complex than in English. For adults, possible plural markers include *-s*, *-en*, *-er*, and *-e*. Clahsen et al. argue that adults take the "regular" plural ending to be *-s*, but certain children in their study took it to be the more frequent ending *-en*. Either way, each child avoided his or her "regular" ending inside compounds, and allowed the others.

derivational morpheme *-er*; and then any of the compounds *rat-*, *mouse-*, and *mice-eater* can be constructed and passed onward. At Level 3 the stem *rat* could in principle be combined with the regular plural-marker *-s*, but not if it is already contained inside a larger word. Hence, **rats-eater* cannot be generated.

Gordon, in his 1985 paper, was careful to point out that the theoretical importance of his findings would remain, even if Kiparsky's Level-Ordering story came to be replaced by another account of the regular / irregular contrast. Indeed, the general Level Ordering thesis has since fallen out of favor. (See Spencer 1991: 179–183 for a brief synopsis of the principal reasons.) One challenge to the thesis comes directly from plurals-within-compounds.

On the one hand, Senghas et al. (1991) have demonstrated that adult English-speakers judge irregular plurals to be far more acceptable than regular plurals in the modifier position of a novel compound. Futhermore, as we have just seen, children as young as 3 years old exhibit the same pattern in Gordon's EP task. On the other hand, there exist a great many exceptions to this pattern, and they belong to several different categories. Two cases were actually addressed in Kiparsky's original proposals. First, as discussed above, pluralia tantum include morphologically regular plural marking, but are necessarily lexically listed and therefore must be handled at Level 1. As expected, this type of plural-marking is unproblematic within compounds.

A second type of exception concerns examples like *Human Resources Department*. Here Kiparsky proposed an explanation in terms of recursion. The idea is that a morphologically complex output of the word-formation system (e.g. the plural compound *human resources*) sometimes comes back and gets listed as a single unit in the lexicon. When it does, it can be included within another compound in the same way as a pluralia tantum. Also, as a consequence of being lexically listed, the form can have either a more specialized meaning, or an entirely

different meaning, than its compositional semantics would indicate (cf. *human resources*, meaning "personnel").[21]

Alegre and Gordon (1996) discuss a number of additional types of exception. One concerns modification by an inherently quantificational noun. For example, a *week-long seminar* can last only one week, while a *weeks-long seminar* must last longer than one week. Another type of exception, which the authors term the "heterogeneous" type, includes examples like *publications catalog*, where the use of plural-marking seems to indicate that the publications listed in the catalog are heterogeneous in nature (i.e. that multiple publication-types are represented). This intuition is brought out fairly sharply by the observation that a mineralogist, who presumably studies many different kinds of rocks and is interested in the differences among them, can be called a ?*rocks expert* much more readily than a simple pile of rocks can be called a **rocks pile*.

The main type of exception that Alegre and Gordon investigate is a case where the plural modifier in a nominal compound crucially needs to be modified by an adjective: *[new book<u>s</u>] shelf* versus **book<u>s</u> shelf*. The authors' analysis of this case is as follows. First, they argue that Kiparsky's proposal of a recursive loop is on the right track, but needs to allow full-on syntactic objects to be stored as lexical items. This position is supported by examples like *the how-can-he-be-a-seat-of-the-pants-executive-if-he-needs-experience absurdity*.

Second, they propose that the human sentence-processing system prefers not to posit the type of structure that is required in order to have a <u>syntactic</u> object inside a <u>morphological</u> object inside a <u>syntactic</u> object. Hence, faced with a choice between [$_{N}$ [$_{NP}$ *red rat*] *eater*] (= 'eater of

[21] A crucial assumption is that this type of recursion is constrained in some way, because otherwise forms like **rats-eater* would be (or at least, would rapidly become) fully acceptable, as a result of feeding the NP *rats* back into the lexicon as a stored expression.

red rats') and [$_{NP}$ *red* [$_{N}$ *rat eater*]] (= 'red eater of rats'], it favors the latter parse. Yet, if the modifier bears regular plural marking, then a parse involving the compound noun **rats-eater* (i.e. [$_{NP}$ *red* [$_{N}$ *rats eater*]] (= 'red eater of rats']) is strongly dispreferred, and the parse that interleaves syntactic and morphological structure becomes the best available option: [$_{N}$ [$_{NP}$ red rats] eater] (= 'eater of red rats').[22]

Alegre and Gordon argue that far from undermining the work in (Gordon 1985), the discovery of these apparent exceptions to "Kiparsky's Generalization" simply indicates that the contributions of UG must be even richer than previously thought, in order to account for the complex system (whatever it is exactly) that adults are using. Moreover, the discovery of even finer-grained patterns in the adult data, such as the interaction of regular plural-marking with the presence/absence of an attributive adjective, creates wonderful new opportunities to assess how early the adult system is actually present in the child.

To this end the authors carried out a new child-language experiment, this time checking children's preferred interpretation of structurally complex compounds like *red rat eater* and *red rats eater*. The method was picture-selection, where the child chose between two side-by-side images. For example, in one image there might be a red monster eating a blue rat, while in the other there would be a blue monster eating a red rat. The subjects were 36 children (12 each) at

---

[22] As noted above, a third proposal is needed to block structures like [$_{N}$ [$_{NP}$ *rats*] *eater*]. If these were allowed we would have the unwanted prediction that **rats-eater* is fully acceptable (or at least as acceptable as *red rats eater*, given that the derivation would be essentially the same). The authors' proposal on this point is not entirely clear to me, but seems to involve the idea that an unmodified nominal modifier with plural marking is necessary given the "heterogeneity" interpretation (which would then be anomalous in the case of a rat-eater).

the ages of 3, 4, and 5 years. Each child judged four test items. The child received either four items with plural marking (*red rats-eater*), or four items without it.

The main findings were as follows. When there was no plural marker in the compound, children in all three age groups preferred the "non-recursive" interpretation (e.g. red eater of blue rats), as expected if children have the adult system in place very early. When the plural marker was present, the children's preference reversed, again as expected if children master the system very early. At age 3, the children on average selected the "recursive" interpretation 1.0 times (out of four opportunities) when the modifier was singular, and 3.33 times when it was plural. For 4-year-olds the corresponding averages were 0.67 and 2.5, and for 5-year-olds they were 1.5 and 2.5. (The contrast between the singular and plural conditions was robustly significant by ANOVA, while there was no significant effect of Age nor a significant interaction of Age with Condition.)

In sum, the work of Gordon and Alegre has provided powerful evidence that English-learning 3-year-olds already know a great deal about the morphology and syntax of synthetic compounds—indeed, far beyond what they could realistically have inferred from their input. Clark et al.'s well-known (1986) finding that English-learning preschoolers occasionally produce compounds of the form *puller-wagon* (section 6.2.2) seems best explained in terms of task demands, rather than erroneous decisions about the grammar of -ER compounds, given that none of the children in the study produced the error consistently. On the other hand, Clark et al. obtained a robust finding that when English-learning children are asked to create a novel word to name a type of person or a type of physical instrument, younger children (3- and 4-year-olds) mainly employ bare-stem compounds like *water-person*, while older children (5- and 6-year-olds) mainly employ adult-like synthetic compounds (e.g. *box-mover*).

6.2.4 Synthetic Compounds in Romance Languages

Turning to languages other than English, we have seen that French (together with the other major Romance languages) does not allow a morpheme-by-morpheme counterpart to English synthetic compounds like *dishwasher*. Recall that this fact can be explained in terms of Beard's Generalization. If the "unmarked" position for French adjectives is post-nominal, then modifiers in compounds should likewise follow the compound's head, and will end up competing with the derivational suffix *-eur* for the right edge of the word.

Hence, as soon as the child recognizes that the target language puts modifiers to the right of a head noun, and that derivational morphology is suffixal, he or she should (ideally) be on the lookout for exocentric V-N compounds like *lave-vaisselle*, and for phrasal compounds like *laveur de vaisselle*, both of which are in principle possible in this language type. Hence we might expect one or both of these forms to be relatively early acquisitions.

Clark (1993: ch.10) reviews studies eliciting novel agent nouns and/or instrument nouns that had been conducted with children acquiring English, Icelandic, Hebrew, French, or Italian. For French there were no EP data available for children younger than 5, but for Italian there was work by Lo Duca (1990) with children ranging from 3;03 to 7;10. Clark (1993: 194–195) reports that Lo Duca used an elicitation prompt of the form, "Com[e] si chiama quello che fa le pizze?" ('What do you call a person who makes pizzas?'), and that more than two-thirds of the responses from 3-year-olds used exocentric V-N compounds. As one looks at progressively older children, one finds increasing use of derivational suffixes, which is consistent with adult practice. (Adults reportedly find V-N compounds more appropriate for instruments than for agents.)

Thus, Italian V-N compounding appears to be well-established by around age 3 (and quite posssibly earlier) as a fully creative process of word-forrmation. As the child gradually

acquires the derivational morphemes that Italian provides for creating novel agent nouns, it seems the use of V-N compounds for agents is increasingly preempted by the more specific terms that become available. The earliness of V-N compounding, and the fact that children initially use it even more extensively than adults do, are both consistent with the idea that broader characteristics of the language may have "primed" the child to acquire V-N compounding very early.

6.3 The Acquisition of Bare-Stem Endocentric Compounding

Turning now to bare-stem compounds, some of the questions that a language learner will need to answer are the following:

(2) a. Is bare-stem endocentric compounding a creative process in my target language?
b. If so, can it be used recursively?
c. Is the head on the left or the right side?
d. Are there any linking elements that occur inside the compounds?
e. If so, what determines their distribution?

Here I will focus on (2a,b). Given that Namiki's Generalization (discussed in section 6.1.3) has held up well over the years since it was proposed, I will also assume that creativity and recursivity can be treated as a package.

6.3.1 Origins of The Compounding Parameter

Beginning with Snyder (1995), and in subsequent work up to the present, I have been investigating what I refer to as "The Compounding Parameter" (TCP). In its current formulation (e.g. Snyder 2012), TCP concerns the availability of a mechanism that is essential for, among other things, the semantic interpretation of novel endocentric compounds. To a first approximation, however, we might think of TCP as a simple yes/no specification of whether bare-stem endocentric compounding is creative.

In Snyder (1995) I first proposed that certain "complex predicates," including adjectival resultatives (e.g. *wipe the table clean*) and separable-particle constructions (*pull the lid off*), are possible only in [+TCP] languages. This hypothesis was suggested by work on Dutch (Neeleman 1994) and Afrikaans (LeRoux 1988) arguing that in those languages, both adjectival resultatives and verb-particle combinations often have the morphological status of compound words. Even though the same is not true in English, I began to explore the possibility that there is nonetheless, even in English, a more abstract connection between complex predicates and compounding.

A comparison of Germanic languages with Romance languages suggested that creative, bare-stem endocentric compounding (which is available in all the Germanic languages but none of the major Romance languages) might be a relevant type of word formation, because all the Germanic languages have adjectival resultatives and separable-particle constructions that are comparable to the ones in English, while none of the Romance languages do. A small-scale survey of the world's languages lent plausibility to the idea, because for the languages sampled, there was at least a one-way implication: Every language with adjectival resultatives and/or separable particles had bare-stem endocentric compounding as a creative process of word formation.[23]

My next step was to check for a connection between complex predicates and compounding in language acquisition. I decided to focus on the acquisition of English, because at the time the CHILDES database (MacWhinney and Snow 1990) already included more than a dozen longitudinal corpora for English, while for any other language there were considerably fewer. I also decided to focus on separable particles with transitive verbs, because unlike adjectival resultatives they are used frequently by both adults and older children; and because

---

[23] For details, see Snyder (2001) and the update regarding Basque in Snyder (2012).

with transitive verbs there is often a direct object intervening between the verb and the particle, which reduces the likelihood that a child's verb-particle combination is simply an unanalyzed "chunk."

Given the high frequency of particle constructions, and as it turned out, novel N-N compounds, it was possible to identify a fairly precise point in each child's corpus where the child went from never using the given structure, to using it frequently, correctly, and with a variety of lexical items. (In Snyder 2007 this point is referred to as the age of FRU, for "First use, followed soon after by Regular Use with varied lexical items.") In the case of novel compounds, there was a worry that the frequency might be too low to see a sharp change at the point of acquisition, but fortunately the children went through what Brown and Hanlon (1970: 33) termed a "brief infatuation": when they discovered bare-stem endocentric compounding, they treated it like a new toy. Hence there was no difficulty in identifying any child's FRU.

The result was an extremely tight correlation, with a best-fit line that closely approximated an identity function: the FRUs for compounding and particles were consistently very close together in time, and often occurred during the same recording session. This pattern has held up quite well as more longitudinal corpora have become available from children acquiring English. In Snyder (2007: 92–-93) I reported an updated version of the analysis, based on the 19 highest-quality longitudinal corpora available at that time. The ages of FRU for compounding ranged from 1.85 to 2.59 years, and for particles ranged from 1.85 to 2.56 years. Pearson's r was .937, indicating that 88 percent of the variability in ages of FRU for either compounds or particles was predicted by the ages of FRU for the other ($t(17) = 11.1$, $p < .001$).

Of course, one needs to be careful when interpreting this type of correlation, because children go rapidly from knowing very little about their target language to knowing a great deal.

As a consequence, many different measures of language ability will show some degree of correlation. I therefore obtained a quantitative measure of general linguistic development, and applied the statistical technique of partial correlation. More precisely, I determined that the children, at the point of their FRU for verb-particle constructions, had an average MLU (mean length of utterance) of 1.919 morphemes. For each child I then determined the age at which the MLU had first reached 1.919 morphemes.

Thus armed with a general developmental predictor of when each child would begin using verb-particle combinations, I applied the statistical method of partial correlation. This allowed me to see what would happen if I removed all the variability in the ages for particles and compounds that could be explained in terms of the general developmental predictor, and just looked for a correlation in whatever variability was left over. The correlation between the MLU-based measure and the ages of FRU for verb-particle combinations was quite strong ($r = .8690$), which indicated that the MLU measure was a good control. Nonetheless, after partial correlation, when all the variation that could be explained by the MLU-measure had been "partialed out," there was still a robust correlation between compounds and particles: $r_{partial} = .799$, $t(17) = 5.31$, $p < .001$. Hence, the association of particles with creative, endocentric compounding goes well beyond what one would expect on general developmental grounds, and instead seems to be a deeper, grammatical connection.

6.3.2 Further Tests of the Syntax–Compounding Link: Japanese and German

The proposed link of certain complex predicates to creative endocentric compounding has been tested acquisitionally in both Japanese and German. Japanese is a [+TCP] language, with creative bare-stem endocentric compounding as well as adjectival resultatives, but no separable-particle construction. Japanese-learning children acquire endocentric compounding considerably

later, on average, than English-learning children. This fact enabled Sugisaki and Isobe (2000) to use EP and truth-value judgment (TVJ) tasks (both of which work best when the child is at least 36 months old) in their study testing predictions for the acquisition of Japanese.

Specifically, Sugisaki and Isobe (2000) used a cross-sectional approach, and tested each of 20 children (age 3;04 to 4;11) in (i) an EP task for novel compounds, and (ii) a comprehension task for adjectival resultatives. The prediction was that children who understood adjectival resultatives would be able to produce novel compounds. Each child received four items on the EP task, and six items (three for each of two verbs) on the TVJ task. A child was classified as "passing" the EP task if she answered at least three out of four items successfully, and as passing the TVJ task if she answered all three items correctly for at least one of the two verbs.

The results were that ten children passed on both, and six children failed on both. Of the remaining four children, two passed on compounding but failed on resultatives, and two passed on resultatives but failed on compounding. Curiously, these four children with discordant patterns were also the oldest children in the study. Sugisaki and Isobe report that they were clearly more interested in the laptop that was used to present the stimuli than they were in the actual stimuli. Nonetheless, the association between compounding and resultatives reached statistical significance, $p = .019$ by two-tailed Fisher Exact Test, and thus supported the proposed link between resultatives and compounding.[24]

---

[24] Given the noisiness of the data, and the difficulties in holding some children's attention, it would be highly desirable to do more research with Japanese-learning children. One possible change would be to use a narrower age range, so that the experimental materials can be age-appropriate for all the subjects. Another possible change would be to avoid analytic methods that require imposing pass/fail cut-off points on the data.

Turning to German, Hanink and Snyder (2014) have analyzed longitudinal corpora from ten children acquiring German as their first language. German has creative endocentric compounding much like that in English, and in addition to adjectival resultatives, has a separable-particle construction that is broadly speaking comparable to that found in English, although many of the details are different. In particular, German is a V2 language with underlying SOV order. Particles are preverbal, but are left behind if the V raises into V2 position: *Jan hat das Buch aufgehoben*. (literally) 'Jan has the book up-lifted.' *versus Jan hob das Buch auf.* 'Jan lifted the book up.'

Children's FRUs of novel compounds ranged in age from 1;11,02 (*Schlümpf Buch* 'Smurf book') to 2;06,14 (*Babytuhl* 'baby chair'). Due to concern about children's possible production of unanalyzed chunks, particles in immediately preverbal position were set aside. The FRUs of separated particles (including cases where the particle immediately followed the verb) ranged from 1;10,14 (*des kippt um* 'that falls over') to 2;08,00 (*pa(sst) das nich hier hin* (lit.) 'fits that not here in'). The ages of FRU for compounds and particles were strongly intercorrelated ($r = .97$, $t(8) = 11.34$, $p < .001$). This correlation remained robust after partialling out the age of each child's first lexical compound ($r = .95$, $t(7) = 8.41$, $p < .001$), and after partialling out the age of each child's first attributive adjective modifying a noun ($r = .81$, $t(7) = 3.64$, $p = .008$).

Lexical compounds were chosen as a control because they are closely matched to novel compounds in terms of factors such as length, stress pattern, and articulatory difficulty. Attributive adjectives were chosen because the relationship between an attributive adjective and its head noun is a closely matched control for the conceptual difficulty associated with the modifier–head relationship in a novel endocentric compound. Given that the correlation between

the ages of FRUs for separated particles and novel bare-stem endocentric compounds had a best-fit line approximating an identity function, and remained robust even when closely matched developmental controls were removed by partial correlation, the German findings are highly similar to those in English, and lend strong support to the hypothesis that separable-particle constructions are closely connected to the availability of creative endocentric compounding.

6.3.3 Setting TCP

A few brief remarks are in order about how a child might arrive at the positive or negative setting of TCP. First, given that the negative setting does not appear to add any new types of utterances to the repertoire of a speaker, it very plausibly serves as an initial and/or default setting. Thus, in the case of a [–TCP] language it is possible that there is never any moment when the learner specifically decides that the [–TCP] setting is the correct one. Alternatively, it is also possible that the learner allows a specific window of opportunity (e.g. in terms of the number of utterances encountered so far, in the input for that particular language), and waits until the end of that period to make a determination either way.

Interestingly, in the case of a [+TCP] language, it seems there must be at least two possible routes. In the case of children acquiring English there is an extraordinarily close association between the onset of novel endocentric compounding and the onset of V-particle combinations, but it could have turned out differently. This is because there exist languages like Japanese, where novel compounding is available but separable particles are not. Hence, it could have turned out that children acquiring English went through a Japanese-like stage along the way. The fact that they do not (at least in the cases of the children examined so far) strongly suggests that they discover compounding by way of the particle constructions. In other words,

separable particles are frequent in the child's input and have a highly distinctive surface form. As soon as the child works out what they are, [+TCP] is a necessary consequence.

In contrast, if the child were to rely on endocentric compounds directly, there would be a problem. Even the child hearing French or Spanish is likely to hear a non-trivial number of these, because they exist as frozen, lexicalized forms. It would be quite difficult for the learner to decide whether a specific N-N compound in the input was lexical or novel. Yet, there exists an alternative, given Namiki's Generalization (from section 6.1.3): If the input includes a robust number of recursive endocentric compounds (e.g. [[*Christmas tree*] *cookie*]), the child can safely conclude that the language allows endocentric compounding as a creative process. Such recursive compounds definitely exist in child-directed English (in Snyder 1999 they were found in ten out of ten CHILDES corpora examined), although they have a considerably lower frequency than the particles.

For a child acquiring Japanese, recursive compounds might very well be the best available indication that Japanese is a [+TCP] language, given that Japanese lacks separable particles. Indeed, aside from recursive compounds, the only cue available to the learner might be adjectival resultatives, which are also likely to have a low freqency, given that they do in adults' English. Moreover, while endocentric bare-stem N-N compounding is certainly a creative process in adult Japanese, it competes with another form that has a very similar semantic range, namely the N-*no*-N construction. The genitive marker *no* can be used to create something highly similar to the "N of N" phrasal compounds used in the Romance languages.

Hence, it seems the child acquiring Japanese will have to wait a good bit longer than the English-learning child in order to obtain conclusive evidence that the target language is [+TCP]. This is plausibly why the sample of 3- and 4-year-olds studied by Sugisaki and Isobe still

included a considerable number of children who failed to produce novel compounds. Furthermore, there is an interesting implication for models of language acquisition. If Japanese-learning children set TCP using recursive compounds (or perhaps adjectival resultatives), while English-learning children set TCP using V-particle combinations, it speaks against acquisition models in which each parameter is innately connected to a single trigger.

6.3.4 The Nature and Scope of TCP

At present my working hypothesis is that TCP is a parameter of the syntax–semantics interface, and that the [+TCP] setting makes available an interpretive rule that I term "Generalized Modification" (GM), as shown in (3).

(3) The Compounding Parameter (TCP):
The language (does / does not) permit Generalized Modification.

The proposal is that GM plays an essential role in the semantic interpretation of a novel compound, and also in the interpretation of a complex predicate (such as a V-Particle combination or an adjectival resultative) in which two syntactically autonomous constituents are jointly characterizing a single event.

The proposed role of GM in the interpretation of compounds is very similar to proposals of Jackendoff (2002) and Kratzer (2010), among others, to the effect that the relationship between the head and the modifier in an English endocentric compound is extremely flexible and context-dependent. A formal definition is given in (4).

(4) Generalized Modification (GM)
If α and β are syntactic sisters under the node γ, where α is the head of γ, and if α denotes a kind, then interpret γ semantically as a subtype of α's kind that stands in a pragmatically suitable relation to the denotation of β.

For example, I assume that a noun can be viewed as denoting a kind of individual, and that when GM applies to a novel N_1-N_2 compound in English (i.e. with N_2 as the head), the resulting

meaning will be a subtype of N_2's kind that stands in some contextually appropriate relation to the denotation of N_1. This is extremely flexible, and is meant to be, so as to capture the fact that for an English-speaker, in the right context, a compound like *frog man* could mean the kind of man who looks like a frog, eats frogs, studies frogs, sells frogs, or wishes to be buried in a frog-shaped casket, for example.

The definition in (4) is also intended to encompass verbal predicates, which I assume serve to specify a kind of event (or more precisely, a kind of eventuality—which might be either an event or a state). My proposal is that GM might be one of the only mechanisms available to combine two separate event descriptions into the description of a single event. For example, in the case of an adjectival resultative like *wipe the table clean*, we might assume that *wipe* has undergone movement, and that originally it was part of the constituent *wipe clean*. There GM could apply, taking the V *wipe* to be the head, and returning the meaning "a wiping event of the kind associated with a state of cleanliness." Along the lines of Levin and Rappaport Hovav (1995: 54), I assume that the conceptual calculus of events is highly restricted, and that this denotation would have to be interpreted as "an accomplishment event, with wiping (of the patient) as its development subpart, and cleanliness (of the patient) as its culmination."

A more detailed exposition of the proposed semantics can be found in Snyder (2012). Acquisitional evidence that links a number of additional syntactic structures to the availability of creative bare-stem endocentric compounding can be found in the following works: Beck and Snyder (2001), for English telic path PPs with manner-of-motion verbs; Goodrich and Snyder (2013), for English atelic path PPs with manner-of-motion verbs; and Snyder (2001) for English *make*-causatives, perceptual reports, and *put*-locatives, though for a contrasting view see Eom and Snyder (2012). For a new acquisitional test of the English particle-compound link, using the

Intermodal Preferential Looking paradigm, see Naigles et al. (2013) and Snyder et al. (2014). For evidence that children acquiring French never produce novel N-N compounds, but do produce novel phrasal compounds (N *de* N), see Snyder and Chen (1997).[25]

References– this list will removed before the chapter is typeset and the entries all moved to a collated list to appear at the end of the book. However, please note that there is a query below.

Alegre, Maria A., and Peter Gordon. "Red rats eater exposes recursion in children's word formation." *Cognition* 60, no. 1 (1996): 65-82.

Allen, Margaret. 1978. *Morphological investigations*. Doctoral Dissertation, University of Connecticut, Storrs.

Bauer, Laurie. 1978. *The Grammar of Nominal Compounding, with Special Reference to Danish, English and French*. Odense, Denmark: Odense University Press.

Bauer, Laurie. (2001) *Morphological Productivity*. Cambridge, UK: Cambridge University Press.

Beard, Robert (1995). *Lexeme-morpheme base morphology: a general theory of inflection and word formation*. Albany, NY: SUNY Press.

Beard, Robert (1996, January) Base rule ordering of heads in nominal compounds. Paper presented at the 1996 Linguistic Society of America Meeting, San Diego, California. (Circulated ms., Bucknell University).

Beck, Sigrid and William Snyder (2001) Complex predicates and goal PP's: Evidence for a semantic parameter. In A. H.-J. Do, L. Dominguez, and A. Johansen (eds.) *Proceedings of the 25th Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press.

Borer, Hagit (1988) On the parallelism between compounds and constructs. *Yearbook of Morphology* 1:45-66.

---

[25] While the Romance languages are clearly [-TCP], there do seem to be a few types of compounding that are creative, and these merit greater attention from acquisitionists. These include the exocentric V-N compounds to name instruments and (occasionally) agents, which were discussed in Section 6.2 (cf. French *le mange-souris* lit. 'the eat(s)-mice' for 'the mouse-eater'); and the doubly-headed, appositive compounds like Spanish *la mujer araña*, literally 'the woman spider,' for 'the spider woman.'

Brown, Roger and Camille Hanlon. 1970. Derivational complexity and order of acquisition in child speech. In John R. Hayes (ed.) *Cognition and the Development of Language*, 155-207. New York: Wiley.

Clahsen, Harald, Monika Rothweiler, Andreas Woest, and Gary F. Marcus. "Regular and irregular inflection in the acquisition of German noun plurals." *Cognition 45*, no. 3 (1992): 225-255.

Clark, Eve V., Barbara Frant Hecht, and Randa C. Mulford. "Coining complex compounds in English: Affixes and word order in acquisition." *Linguistics 24*, no. 1 (1986): 7-30.

Clark, Eve (1993) *The Lexicon in Acquisition*. Cambridge, UK: Cambridge University Press.

Eom, Soyoung and William Snyder (2012) Children's acquisition of English datives: Competing parametric accounts. In Y. Otsu (ed.) *Proceedings of the 2012 Tokyo Conference on Psycholinguistics*, 41-59. Tokyo: Hituzi Shobo.

Goodrich, Mary and William Snyder (2013) Atelic paths and the Compounding Parameter: Evidence from Acquisition. In N. Goto, K. Otaki, A. Sato, and K. Takita (eds.) *Proceedings of GLOW in Asia IX 2012*, 19-30. Mie, Japan: Mie University.

Gordon, Peter (1985) Level ordering in lexical development. *Cognition, 21, 73-93.*

Hanink, Emily and William Snyder. (2014) Particles and compounds in German: Evidence for the Compounding Parameter. *Language Acquisition* 21(2):199-211.

Jackendoff, Ray (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.

Josefsson, Gunlög (1997) *On the Principles of Word Formation in Swedish*. Lund: Lund University Press.

Kiparsky, Paul. (1982) "From cyclic phonology to lexical phonology." *The structure of phonological representations* 1:131-175.

Kratzer, Angelika (2010) What grammar might contribute to meaning composition. Slides from talk presented at the Contextualism & Compositionality Workshop, Paris, France, 18 May 2010.

Kučera, Henry, and Winthrop Nelson Francis. (1967). *Computational Analysis of Present-Day American English*. Providence, RI: Brown University Press.

Lardiere, Donna (1998) Parameter-resetting in morphology: Evidence from compounding. In Maria-Luise Beck (ed.) *Morphology and its Interfaces in Second Language Knowledge*, 283-305. Amsterdam: John Benjamins.

LeRoux, Cecile. 1988. On the interface of morphology and syntax. Stellenbosch Papers in Linguistics #18. South Africa: University of Stellenbosch.

Levin, Beth and Malka Rappaport Hovav. 1995. *Unaccusativity: At the Syntax-Lexical Semantics Interface*. Cambridge, MA: MIT Press.

Lo Duca, Maria G. (1990) *Creatività e regole: studio sull'acquisizione della morfologia derivativa dell'italiano*. Bologna: Il Mulino.

MacWhinney, Brian and Snow, Catherine. (1990). The Child Language Data Exchange System: An update. *Journal of Child Language* 17:457-472.

Naigles, Letitia, Katelyn Guerrera, Vanessa Petroj, José Riqueros Morante, Diane Lillo-Martin, and William Snyder (2013) The Compounding Parameter: New evidence from IPL. Paper presented at the Boston University Conference on Language Development, 1 November 2013.

Namiki, Takayasu (1994) Subheads of compounds. In Shuji Chiba (ed.) *Synchronic and Diachronic Approaches to Language: A Festschrift for Toshio Nakao on the Occasion of his Sixtieth Birthday*, 269–285. Tokyo: Liber Press.

Neeleman, Ad. 1994. *Complex Predicates*. Utrecht: Onderzoekinstituut voor Taal en Sprak (OTS).

Nübling, Damaris & Szczepaniak, Renata (2008) On the way from morphology to phonology: German linking elements and the role of the phonological word. *Morphology* 18:1–25.

Rijk, Rudolf P.G. de (2008) *Standard Basque: A Progressive Grammar*. Cambridge, MA: The MIT Press.

Roeper, Thomas and Snyder, Willlliam (2004) Recursion as an analytic device in acquisition. In *Proceedings of GALA 2003 (Generative Approaches to Language Acquisition)*, 401-408 . Utrecht: LOT Publications.

Roeper, Thomas and Snyder, Willlliam (2005) Language learnability and the forms of recursion. In Anna Maria Di Sciullo (ed.) *UG and External Systems: Language, Brain and Computation*, 155-169. Amsterdam: John Benjamins.

Senghas, Ann, Kim, John J., Pinker, Steven., & Collins, Christopher. (I991). Plurals-inside-compounds: Morphological constraints and their implications for acquisition. Paper presented at the Sixteenth Annual Boston University Conference on Language Development, Boston, MA, October 1991.

Siegel, Dorothy. 1979. *Topics in English morphology*. Garland Press, New York. (Doctoral dissertation, MIT, Cambridge, Massachusetts, 1974)

Snyder, William Brandon. (1995) *Language Acquisition and Language Variation: The Role of Morphology*. Doctoral dissertation, MIT, Cambridge, Massachusetts.

Snyder, William (1999) Compounds and complex predicates in English, Basque, and Romance. Poster presented at IASCL (International Association for the Study of Child Language), San Sebastian, Spain, 16 July 1999. [Text available at http://web.uconn.edu/snyder/Snyder_1999_IASCL.pdf]

Snyder, William (2001) On the nature of syntactic variation: Evidence from complex predicates and complex word-formation. *Language* 77:324-342.

Snyder, William (2007) *Child Language: The Parametric Approach*. Oxford, UK: Oxford University Press.

Snyder, William (2012) Parameter theory and motion predicates. In Violeta Demonte & Louise McNally (eds.) *Telicity, Change, and State: A Cross-categorial View of Event Structure*, 279-299. Oxford, UK: Oxford University Press.

Snyder, William and Deborah Chen (1997) The syntax-morphology interface in the acquisition of French and English. In K. Kusumoto (ed.) *Proceedings of NELS 27 (North East Linguistic Society)*. Amherst, MA: GLSA.

Snyder, William, Diane Lillo-Martin, and Letitia Naigles. (2014) The Compounding Parameter: New evidence from IPL. Ms., University of Connecticut, Storrs, 1 September 2014.

Sugisaki, Koji and Miwa Isobe. 2000. Resultatives result from the Compounding Parameter: On the acquisitional correlation between resultatives and N-N compounds in Japanese. Billerey & Lillehaugen (eds.) *WCCFL 19 Proceedings*, 493-506. Somerville, MA: Cascadilla Press.

Spencer, Andrew (1991) *Morphological Theory.* Oxford: Blackwell.

Williams, Edwin (1981) On the notions 'lexically related' and 'head of a word'. *Linguistic Inquiry* 12: 255-274.