Productivity and Creative Use of Compounds in Reduced Registers: Implications for Grammar Architecture

Anastasia Smirnova (smirnov@sfsu.edu)
Department of English Language and Literature, 1600 Holloway Ave
San Francisco, CA 94132 USA

Abstract

Reduced registers – search queries, print and TV ads, and navy messages – are characterized by an unusually high number of novel compounds. Results of a production study reported here reveal combinatorial patterns not attested in the standard language and allow us to establish the range of possibilities. We argue that productivity and creative use of compounds in reduced registers is not coincidental but follows directly from the grammar that generates expressions in these registers. We adopt an analysis couched in the Parallel Architecture framework (Jackendoff, 1997; Jackendoff & Audring, 2016) and demonstrate how productivity and idiosyncrasy of compounds in reduced registers can be explained.

Keywords: compounds; Parallel Architecture; linear grammar; reduced registers; search queries

Introduction

At first encounter, expressions such as *invent toy school project* or *tall tree soil composition* are difficult to comprehend. Yet, such compounds – multiword expressions that function as a single lexical unit – are fairly common. They enjoy high productivity, despite their syntactic and semantic idiosyncrasies. The main challenge for linguistics and cognitive science is to explain how compounds are generated and how their meaning is derived. These questions have direct consequences for what we assume about language architecture and the division of labor between the different grammar components and the lexicon (Bauer, 2003; Culicover, Jackendoff, & Audring, 2017; Liberman & Sproat, 1992; Lieber & Štekauer, 2009).

While compounds are common in everyday language, our focus here is on specialized registers – search queries, print and TV ads, and navy messages – where compounding emerged as a dominant strategy for information packaging. Computational linguists in the 80s analyzed navy messages and technical manuals and observed that these registers are characterized by high number of complex, multi-word compounds, such as *forward kingpost sliding padeye unit* (Grishman & Sterling, 1989; Marsh, 1984). Leech (1963) and Rush (1998) demonstrated that TV and print ads also display a greater number of compounds compared to what is attested in the standard language, such as *ophthalmologist-tested formula* or *fashion fresh color*. Compounds also feature prominently in journalistic prose (Biber, 2003) and headlines, such as *Alibi witness hunter strike* and *Simpson trial welfare moms*, cited in Sadock (1998). More recently, compounding emerged as a productive strategy for formulating search queries – information requests submitted by users to a search engine (Barr, Jones, & Regelson, 2008; Bergsma & Wang, 2007; Pinter, Reichart, & Szpektor, 2016).

What these specialized registers have in common is the tendency to omit otherwise obligatory grammatical material, function words, as well as subjects and objects. Because of their elliptical styles, such registers are often referred to as 'telegraphic', 'reduced' or 'abbreviated' (Stowell & Massam, 2017). Compounding, however, goes beyond the omission of linguistic material, and represents an entirely different strategy for information packaging. It is notable that while in the standard language recursive compounding – compounds embedded within other compounds, such as *water supply network repairs* – are "mercifully rare" (Liberman & Sproat, 1992), they are quite common in the specialized registers, as documented by Marsh (1984), Rush (1998), Biber (2003) and others. It is a puzzle why compounding becomes so prevalent in these contexts.

Compounds in reduced registers present a particular interest because there are fewer restrictions on compound formation (Marsh, 1984; Rush, 1998). For example, bare verbs can rarely function as modifiers of nouns in English compounds, but in navy messages these types of compounds are attested: *operate mode*, *transmit antennae*, *receive sensitivity* (Marsh, 1984). Similarly, nominal compounds that consist of a verb followed by an object noun, such as *cut-throat* or *pickpocket*, are considered to be non-productive (Liberman & Sproat, 1992; Lieber, 1983) and stylistically marked in English (Adams, 1973), but they are attested in search queries. Thus, Pinter et al. (2016) cite *invent toy school project* (= 'school project for inventing toys') from the corpus of search queries, where [invent toy] is a compound, consisting of the verb and the direct object. Moreover, it is often assumed that verbs cannot be heads of compounds, with **to quick act* and **to truck drive* considered to be ungrammatical (Harley, 2009). Contrary to these claims, compounds with verbal heads are attested in search queries, such as *buick lacrosse remote start* (= 'how to remote start buick lacrosse'), discussed in (Pasca, 2013).

Previous literature on compounds in specialized registers documents the patterns but offers little in the way of explanation. Some researchers argue that compounding in naval messages and technical manuals is motivated by consideration of economy. Rush (1998) advocates the position that novel compounds in advertisement serve to attract attention and distinguish the product from other alternatives. Under this view, the usage of compounding is driven by stylistic and marketing considerations.

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

We argue instead that extensive compounding in reduced registers is non-accidental and that productivity of compounds and their unusual combinatorial patterns are a direct consequence of the grammar that generates them.

Search Queries as a Reduced Register

Search queries are produced in the context of human-computer interaction, whereby the user formulates a request for information and expects to get back the list of relevant resources. The most salient feature of search queries is their length: ~2.6 words on average (Aula, Khan, & Guan, 2010). Such brevity is achieved at the expense of functional and grammatical material. Queries freely omit definite and indefinite articles, prepositions, and auxiliary verbs. They also often drop subjects and objects. Modifications observed in queries are so extensive that they are sometimes referred to as a 'word salad'. More recent work, however, shows that while queries are extremely short, they do have specific grammatical properties, albeit different from the standard language. Barr et al. (2008) report that 70% of queries are nominal constructions, consisting of nouns and their dependents. These nominal constructions often have forms of noun phrases and nominal compounds.

The research on compounds in search queries is primarily concerned with extracting meaning and identifying user intent. The challenge is that compounds obscure grammatical relations between elements of the construction. In general, dependency relations and grammatical functions are expressed syntactically via word order. For example, a transitive verb in English has two arguments, a subject and a direct object. In active sentences the subject precedes the verb and the object follows the verb. In compounds, both subjects and objects are placed before the head noun, so there is no positional distinction between them, as in *warrior dance* (warrior is understood as the subject) and *weight loss* (weight is understood as the object). Thus, meaning relations in compounds are derived not from word order but from world knowledge and context.

Multi-word compounds present a particular challenge, since they are structurally ambiguous. Bergsma and Wang (2007) observe that the search query *two man power saw* can be parsed differently, and that depending on the structural analysis, users will see different results. Thus, [two man] [power saw] returns results about the saws operated by two people and used to fell big trees, while [two] [man power] [saw] returns non-sensical results.

Research on compounds in search queries does not say much about their structural and semantic properties, nor does it explain how such compounds are generated in the first place. To answer this question, we report results of a production experiment with the goal to establish the scope of compound productivity and identify compound types generated in the context of search. Based on the previous literature, our hypothesis is that compounding will be a productive strategy in the search query register.

Experimental Study

In this section we report results of a production study that allows us to directly contrast the information packaging strategies in the context of search to those that arise naturally in communication between humans, or the baseline. The data analyzed here were collected as part of the study on syntactic subordination in search queries. Here we reanalyze the data with the special attention to compounds. The analyses and results presented here have not been previously published.

Participants

We recruited eighty participants from the Amazon Mechanical Turk (AMT) platform, who were directed to the study hosted on Qualtrics. All participants were native speakers of English. The average age was 40 years old. 62% were male, and 38% female.

Design and Procedure

In the beginning of the study participants were introduced to the main protagonist, Maria. They were instructed to help Maria formulate information requests by typing them into a text box. The participants were first presented with a practice example, after which they saw the actual scenarios. There were two conditions. In one condition, participants learned that their information requests will be read and answered by a knowledgeable person (human condition). In another condition, the participants learned that their information requests will be answered by Google (computer condition). We hypothesized that the computer condition will generate a greater ratio of compounds compared to that in the standard language (human condition).

Stimuli

The stimuli consisted of 16 unique information search scenarios. The scenarios were formulated to evoke participants' curiosity.

Formally, each scenario contained a relative clause. Based on the association between compounds and relative clauses established in the previous literature, we hypothesized that constructions with relative clauses can be paraphrased with compounds.[1] We used three types of relative clauses: subject relative clauses, *when*-relative clauses, and *where*-relative clauses. All scenarios were amenable to a compound paraphrase. If compounding is a salient feature of reduced registers, we expect that users in the computer condition will tend to spontaneously paraphrase relative clauses with compounds, despite being primed with relative clauses. Our

[1] In the early days of transformational grammar Lees (1960) argued that nominal compounds are derived from sentences and that constructions with relative clauses are one of the steps in the compound derivation process: {The sheep has a horn. The horn is like a prong.} → The sheep has a horn which is like a prong → …→ The sheep with a prong horn → a pronghorn (Makino, 1976). Even though the transformational analyses have long been abandoned (see Downing (1977) for early criticism), they exemplify an implicit association between compounds and relative clauses and the possibility to paraphrase relative clauses with compounds.

hypothesis is that the information requests in the human condition will not show the tendency to use compounds.

Results

Participants' responses were manually coded. Information requests that were compounds were coded as 1. Everything else, including fully grammatical sentences or phrases was coded as 0. For example, *indoor pet safe carnivorous plant nursery* was coded as 1, because it is a compound. On the other hand, *what plants are safe for pets* was coded as 0 because it is a sentence; *carnivorous plants and pets* was coded as 0, because it is a Noun Phrase (NP). Queries that contained compounds but were not compounds themselves were also coded as 0. For example, *nursery with pet-friendly carnivorous plants* contains a compound (underlined) but it was coded as 0, because syntactically it is a NP rather than a compound. Note that this is a conservative strategy, since such queries were predominantly generated in the computer condition. If we include compounds that are contained within phrases that would make the differences between the human and the computer conditions even more robust in the direction predicted by the study.

Collapsing across the scenarios, we found that there was a strong preference for using compounds in the computer condition (M = 0.34, SD = 0.32) compared to the person condition (M = 0.03, SD = 0.09). The differences between the two conditions were statistically significant on a two-sample two-sided t-test ($t(78) = 5.5$, $p < 0.001$). The results are shown in Figure 1.

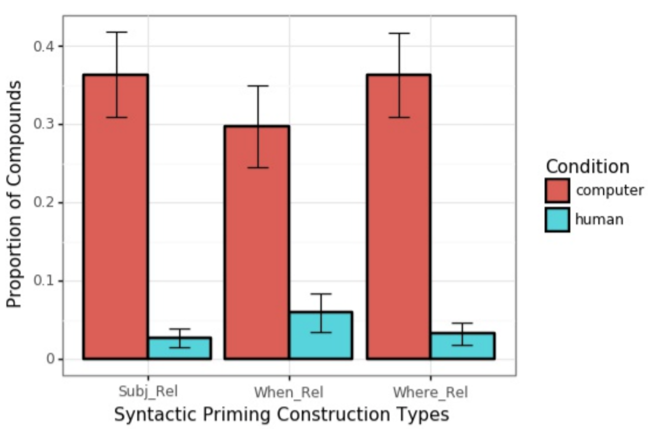

Figure 1: Proportion of compounds in information requests directed to computer (computer condition) and people (person condition). Error bars indicate +/-1 SE.

Discussion

The analysis of the data revealed that participants in the computer condition used compounding to a much greater extent than the participants in the human condition. It is important to point out that occasionally we do encounter fully-formed sentences in the computer condition, but their number is rather small. This is not unexpected. White, Richardson, and Yih (2015) report that 3.8% of queries are grammatical sentences. In the human condition, the main tendency was to use fully formed grammatical sentences. Despite being primed with constructions that are amenable to the compound paraphrases, participants used few compounds.

Turning now to the data generated by participants in the computer condition, several types of compounds can be distinguished. The classification adopted here is based on the types of semantic relations that hold between the elements of compounds, as proposed in Jackendoff and Wittenberg (2014). First, there are compounds such as *carnivorous plants*, *deaf employee*, and *vegan meals*, which exemplify a modification relation between the head of the compound and its dependent. In their basic form, such compounds can be paraphrased as 'X that are Y', i.e. 'plants that are carnivorous'. Second, there are compounds such as *soil composition*, in which the head is a functor that takes the dependent, *soil*, is an argument. They can be paraphrased as 'X of/by/…Y' (cf. Jackendoff, 2009). These compounds can be characterized as function-argument compounds (cf. argument-predicate compounds in Liberman & Sproat, 1992). Finally, in some compounds, such as *margaret mead culture* (= 'culture that Margaret Mead researched'), the head and the dependent are arguments of an unexpressed functor. In this example, the unexpressed functor is predicate CONDUCT-RESEARCH; it takes MARGARET MEAD and CULTURE as arguments. We refer to this type as co-argument compounds (cf. argument-argument compounds in Liberman & Sproat, 1992). The specific semantic relation between the elements in these compounds is derived from world knowledge. Table 1 summarizes major compound types in search queries. [2]

Table 1: Types of compounds in search queries

Relations between elements in compounds	Examples
modification	vegan meals
function-argument	soil composition
co-arguments	margaret mead culture

Notably, compounding is recursive, and the majority of data in our study were compounds built out of other compounds. Thus *chicago stadium vegan food* (='chicago stadium that serves vegan food') encodes different types of semantic relations, including modification (*vegan food*), as well as the co-argument relation, where the contextually supplied function SERVE takes CHICAGO STADIUM and VEGAN FOOD as its arguments.

[2] A reviewer suggests that information about the frequency of each semantic relation (modification, function-argument, and co-argument) might be useful. However, in our data the majority of compounds are not uniquely associated with a single type; they encode multiple semantic relations simultaneously, which makes unambiguous classification unattainable.

This discussion leads us to the following questions:

1. How are the observed compound types generated and interpreted?

2. What is the principled explanation for the predominance of compounds in the search query register?

Linear Grammar

When properties of compounds are evaluated together, the following picture emerges. There is evidence for a productive rule that concatenates elements to produce a novel compound. There is no evidence that lexical categories play a role. According to Jackendoff (2009), the existence of near-minimal pairs such as $[\text{marine}_A\ \text{life}_N]_N$ and $[\text{sea}_N\ \text{life}_N]_N$ suggests that compounding is not sensitive to word classes. The meaning, rather than being derived from the meaning of the parts and the way they combine together, is idiosyncratic and heavily dependent on context. This is very different from how the rest of the language works. Based on these observations, Jackendoff (2009) proposed that compounds are relics of an older system, a protolanguage (Bickerton, 1990; Givón, 2008). Protolanguage can be viewed as an evolutionary predecessor of contemporary languages, a more rudimentary system that is characterized with the absence of a fully-fledged syntactic component present in modern languages. Nevertheless, such linguistic system can be quite effective. Speakers can have access to words and their meanings, and there is a concatenation rule that combines words together to produce multi-word expressions. Following Jackendoff and Wittenberg (2017), we refer to this type of system as linear grammar.

Jackendoff and Wittenberg (2017) propose that as the language evolved, linear grammar did not disappear, but continued to co-exist along with the more complex grammatical system of the modern language. The authors argue that linear grammar operates in a range of contexts, including child language, pidgins, emergent sign languages, agrammatic aphasia, and other. All these language varieties are characterized by the lack of function words, relatively free word order, and heavy reliance on context. In the case of child language, linear grammar becomes a stage in language development. In the case of language impairment, speakers fall back on linear grammar when access to fully-developed grammar is blocked. General adult population accesses linear grammar in a range of contexts. In fact, Jackendoff and Wittenberg (2017) argue that compounds are a product of linear grammar. We propose to extend this analysis to search queries. We argue that compounds are so dominant in the context of search, because search queries themselves are generated by linear grammar. Indeed, similar to other language varieties, search queries demonstrate omission of functional material, optionality of the core arguments, and flexible word order that is governed by pragmatic considerations. In such a system, context rather than syntax plays a major role.

Consider how the major types of compounds can be derived in such a system. The three types of semantic relations in compounding are modification, function-argument, and co-argument relations (cf. also Liberman & Sproat, 1992). Rather than being derived by a procedural syntactic rule of the form N→ N N, the compounds are licensed by declarative schemas or interface rules that determine the mapping between phonology and semantics. We assume that these constructions can be derived without reference to the syntactic component or lexical classes, such as nouns or verbs.

The schema in (1) establishes modification relations, where Y functions as a modifier of X (from Jackendoff & Wittenberg, 2014). Such a schema can license *vegan meals*, as shown in (2).

(1) Modification schema
Phonology: $[_{\text{Utterance}}\ \text{Word}_1, \text{Word}_2]_3$
Semantics: $[X_2; Y_1]_3$

(2) Derivation of *vegan meals*
Phonology: $[_{\text{Utterance}}\ \text{vegan}_1, \text{meals}_2]_3$
Semantics: $[\text{MEALS}_2; \text{VEGAN}_1]_3$

The rule in (3) is a function-argument schema (Jackendoff & Wittenberg, 2014), which specifies that Word_2 is semantically a function that takes the meaning of Word_1 as an argument. This rule derives the meaning of right-headed compounds, such as *soil composition* (= 'composition of soil'), shown in (4).

(3) Function-argument schema
Phonology: $[_{\text{Word}}\ \text{Word}_1, \text{Word}_2]_3$
Semantics: $[F_2\ (X_1)]_3$

(4) Derivation of *soil composition*
Phonology: $[_{\text{Utterance}}\ \text{soil}_1, \text{composition}_2]_3$
Semantics: $[\text{COMPOSITION}_2\ (\text{SOIL}_1)]_3$

Finally, the co-argument schema in (5), from Jackendoff and Wittenberg (2014), supplies a contextually-relevant function that takes the denotations of compound elements as arguments. (5) can account for *margaret mead culture*, as shown in (6).

(5) Co-argument schema
Phonology: $[_{\text{Utterance}}\ \text{Word}_1, \text{Word}_2]_3$
Semantics: $[F\ (X_1, Y_2)]_3$

(6) Derivation of *margaret mead culture*
Phonology: $[_{\text{Utterance}}\ (\text{margaret mead})_1, \text{culture}_2]_3$
Semantics: $[\text{CONDUCT-RESEARCH}\ (\text{MARGARET-MEAD}_1, \text{CULTURE}_2)]_3$

These schemas can be extended to multiword compounds, such as *chicago stadium vegan food* (= 'chicago stadium that serves vegan food'). Structurally, such a compound can be derived by the application of the co-argument schema to [chicago stadium] and [vegan food]. The subcomponents of this compound, [chicago stadium] and [vegan food] are themselves compounds. *Vegan foods* is derived by the modification schema in (1). *Chicago stadium* is a bit more complex, as it involves both the modification schema and the co-argument schema. The latter is needed to derive the spatial location of the stadium: IN (CHICAGO, STADIUM). Combining the modification and the co-argument schema we get [STADIUM$^{\alpha}$; IN(CHICAGO, α)] for *chicago stadium* (cf. Jackendoff, 2009). (7) demonstrates the final step in the derivation, and the application of the function-argument schema.

(7) Derivation of [*chicago stadium*]$_1$ [*vegan food*]$_2$
Phonology: [$_{\text{Utterance}}$ Word$_1$, Word$_2$]$_3$
Semantics:
[SERVE ([STADIUM$^{\alpha}$; IN(CHICAGO, α)]$_1$,[FOOD;VEGAN]$_2$)]$_3$

Consider now how the proposed analysis extends to the cases that are predicted to be non-productive in the standard language, including *invent toy* (Pinter et al., 2016) and *remote start* (Pasca, 2015). Harley (2009) proposes to explain unavailability of compounds such as **to quick act* and **to truck drive* with the assumption that verbs cannot host compounds in English, a parametric feature. The linear grammar approach presented here does not assume sensitivity to lexical classes or grammatical relations, and derives such compounds by the application of the schemas discussed above. For example, in the linear grammar model, *invent toy* is derived by the function-argument schema, as shown in (8). [3]

(8) Derivation of *invent toy*
Phonology: [$_{\text{Word}}$ invent$_1$, toy$_2$]$_3$
Semantics: [INVENT$_1$ (TOY$_2$)]$_3$

Finally, one interesting tendency that we observed in our production experiment is to follow well-formed compounds by modifiers that provide additional information about location or time, as in [[deaf employee] restaurant] [LA]. While traditionally such constructions would not be considered compounds, they nevertheless present an example of creativity. More importantly, such constructions can be derived by the linear grammar model adopted here. Specifically, *deaf employee restaurant* will be derived by the function-argument schema, where the contextually-supplied function EMPLOY takes RESTAURANT and DEAF EMPLOYEE as its argument. The final element, LA, can be integrated via the modification and the co-argument schema.

General Discussion

In this study we investigated compound production in the context of a specialized reduced register – search queries. We demonstrated that in the context of search compounding becomes the dominant strategy for presenting information. We proposed that the prevalence of compounds in search queries is explained by the fact that queries are generated by a qualitatively different type of grammar, linear grammar (Jackendoff & Wittenberg, 2017). These findings have several implications for the debate on the processes of compound generation and the grammar architecture more generally.

First, the linear grammar model adopted here is rooted in a particular theoretical approach to grammar – Parallel Architecture (Jackendoff, 1997; Jackendoff & Audring, 2016). The central feature of the Parallel Architecture is the assumption of relative independence of grammar components and its rejection of syntactocentrism. A direct consequence from this assumption is that certain models of the grammar can lack a fully-fledged syntactic component altogether. This is the case with the linear grammar. In such a grammar, the explanation of the otherwise puzzling features, such as lack of function words, optionality of arguments, flexible word order and heavy reliance on context for interpretation comes for free. Such a model readily explains the predominance of compounds in reduced languages, including the structures that are predicted to be impossible by the grammar of the standard language.

Second, our analysis of compounds as a product of linear grammar assumes that such constructions are licensed by declarative schemas ([$_N$ N N]) rather than by procedural rules in the syntactic or morphological component (N → N N), which are postulated within the mainstream generative grammar framework. As long as the schemas' constraints are satisfied, a novel expression is generated. More importantly, because in the Parallel Architecture framework the relationship between schemas and their lexical instantiations is captured in terms of inheritance, this allows us to still capture parallels between compounds with idiosyncratic meaning stored in the lexicon (*cut-throat* or *pickpocket*) and novel constructions licensed by the declarative schemas (*invent toy*). Thus, this approach provides an elegant explanation for compound productivity and their idiosyncrasies.

Finally, our analysis contributes to a better understanding of the process of reduction observed in specialized registers. Previous studies of language in reduced registers have shown that brevity is achieved at the expense of function words – prepositions, articles, and auxiliary verbs are usually omitted (cf. Boot et al., 2019). However, the analysis of search queries presented here revealed that in addition to omitting grammatical elements, users actively employ compounding, a strategy that restructures the message and eliminates function words in the process. Consider, for example, the distribution of prepositions. Prepositions are often used to express dependency relations between grammatical elements, as in [collector [of the works [of an artist]$_{PP}$]$_{PP}$]$_{NP}$ (Costello & Keane, 2000). When the same information is presented as a compound, [artist collector]$_N$, the prepositions are eliminated altogether. A similar observation pertains to determiners. For example, when *The Liberty Bell Award by Chicago Tribune*

[3] The only difference with the previous examples is that *invent toy* is a left-headed compound, so the mapping of the functor and argument is different than in right-headed compounds such as *soil composition*.

is expressed as a compound, it becomes *Chicago Tribune (*The) Liberty Bell Award* (Sadock, 1998). From this perspective, deletion of functional material can be viewed as a byproduct of restructuring, and compounding, specifically.

References

Adams, V. (1973). *An introduction to modern English word-formation* (No. 7). London: Longman.

Aula, A., Khan, R. M., & Guan, Z. (2010). How does search behavior change as search becomes more difficult? In *Proceedings of the 10th SIGCHI conference on human factors in computing systems*, CHI 10 (pp. 35-44). New York, NY: Association for Computing Machinery.

Barr, C., Jones, R., & Regelson, M. (2008). The linguistic structure of English web-search queries. In *Proceedings of the 2008 conference on empirical methods in Natural Language Processing* (pp. 1021-1030). Honolulu, HI: Association for Computational Linguistics.

Bauer, L. (2003). *Introducing linguistic morphology*. Edinburgh: Edinburgh University Press.

Bergsma, S., & Wang, Q. I. (2007). Learning noun phrase query segmentation. In *Proceedings of the 2007 joint conference on empirical methods in Natural Language Processing and computational Natural Language Learning* (pp. 819-826). Prague: Association for Computational Linguistics.

Biber, D. (2003). Compressed noun-phrase structures in newspaper discourse: The competing demands of popularization vs. economy. In J. Aitchison & D. Lewis (Eds.), *New media language*. London: Routledge.

Bickerton, D. (1990). *Language and species*. Chicago: University of Chicago Press.

Boot, A. B., Tjong Kim Sang, E., Dijkstra, K., & Zwaan, R. A. (2019). How character limit affects language usage in tweets. *Palgrave Communications*, *5*(1).

Costello, F. J., & Keane, M. T. (2000). Efficient creativity: Constraint-guided conceptual combination. *Cognitive Science*, *24*(2), 299-349.

Culicover, P. W., Jackendoff, R., & Audring, J. (2017). Multiword constructions in the grammar. *Topics in Cognitive Science*, *9*(3), 552-568.

Downing, P. (1977). On the creation and use of English compound nouns. *Language*, *53*(4), 810-842.

Givón, T. (2008). *The genesis of syntactic complexity*. Amsterdam: John Benjamins.

Grishman, R., & Sterling, J. (1989). Analyzing telegraphic messages. In *Speech and natural language: Proceedings of DARPA workshop* (pp. 204-208). Philadelphia, PA.

Jackendoff, R. (1997). *The Architecture of the language faculty*. Cambridge, MA: MIT Press.

Jackendoff, R. (2009). Compounding in the parallel architecture and conceptual semantics. In R. Lieber & P. Štekauer (Eds.), *The Oxford handbook of compounding*. Oxford: Oxford University Press.

Jackendoff, R., & Audring, J. (2016). Morphology in the parallel architecture. In J. Audring & F. Masini (Eds.), *The Oxford handbook of morphological theory*. Oxford, UK: Oxford University Press.

Jackendoff, R., & Wittenberg, E. (2014). What you can say without syntax: A hierarchy of grammatical complexity. In F. J. Newmeyer & L. B. Preston (Eds.), *Measuring grammatical complexity*. Oxford: Oxford University Press.

Jackendoff, R., & Wittenberg, E. (2017). Linear grammar as a possible stepping-stone in the evolution of language. *Psychonomic Bulletin & Review*, *24*, 219-224.

Harley, H. (2009). Compounding in distributed morphology. In R. Lieber & P. Štekauer, R. (Eds.), *The Oxford handbook of compounding*. Oxford: Oxford University Press.

Leech, G. N. (1963). Disjunctive grammar in British television advertising. *Studia Neophilologica*, *35*(2), 256-264.

Lees, R. B. (1960). *The grammar of English nominalizations*. The Hague: Mouton & Co.

Liberman, M., & Sproat, R. (1992). The stress and structure of modified noun phrases in English. In I. Sag & A. Szabolcsi (Eds.), *Lexical matters*. Stanford: CSLI Publications.

Lieber, R. (1983). Argument linking and compounds in English. *Linguistic inquiry*, *14*(2), 251-285.

Lieber, R & Štekauer, R. 2009. (Eds.), *The Oxford handbook of compounding*. Oxford: Oxford University Press.

Makino, S. (1976). Nominal compounds. In M. Shibatani (Ed.), *Japanese generative grammar*. Brill.

Marsh, E. (1984). A computational analysis of complex noun phrases in Navy messages. In *Proceedings of COLING'84* (pp. 505-508). Stanford, CA: Association for Computational Linguistics.

Pasca, M. (2013). Open-domain fine-grained class extraction from web search queries. In *Proceedings of the 2013 conference on empirical methods in natural language processing* (pp. 403-414). Seattle, WA: Association for Computational Linguistics.

Pinter, Y., Reichart, R., & Szpektor, I. (2016). Syntactic parsing of web queries with question intent. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies* (pp. 670-680). San Diego, CA: Association for Computational Linguistics.

Rush, S. (1998). The noun phrase in advertising English. *Journal of Pragmatics*, *29*(2), 155-171.

Sadock, J. M. (1998). On the autonomy of compounding morphology. In S. G. Lapointe, D. K. Brentari, & P. M. Farrell (Eds.), *Morphology and its relation to phonology and syntax*. Stanford: CSLI Publications.

Stowell, T., & Massam, D. (2017). Introducing register variation and syntactic theory. *Linguistic Variation*, *17*(2), 149-156.

White, R. W., Richardson, M., & Yih, W. T. (2015). Questions vs. queries in informational search tasks. In *Proceedings of the 24th international conference on World Wide Web* (pp. 135-136).

Appendix: Experimental Stimuli

Stimuli priming subject relative clauses

1. [**Bitcoin bus**] Maria wants to know if there is a bus that only accepts bitcoins. What should she type into the text box? 2. [**LA restaurant**] Maria knows that there is a restaurant in LA that only hires deaf people, and wants to know the location of the restaurant. What should she type into the text box? 3. [**Baseball**] Maria wants to know how old the player who set the record for the most home runs in 1975 was at the time. What should she type into the text box? 4. [**Philanthropy**] Maria wants to know the names of super rich people in the Bay Area who pledged to give away most of their wealth. What should she type into the text box? 5. [**Carnivorous plants**] Maria wants to get carnivorous indoor plants that are safe for pets, and wants to know which nursery carries such plants. What should she type into the text box? 6. [**Chicago stadium**] Maria knows that there is a sport stadium in the Greater Chicago Area which serves vegan meals, and wants to know who is the owner of the stadium. What should she type into the text box?

Stimuli priming *when* relative clauses

1. [**Bees**] Maria wants to know what bees do when animals try to take their honey. What should she type into the text box? 2. [**Coastal cities**] Maria wants to know what will happen to coastal cities when sea-level rise reaches 1 inch. What should she type into the text box? 3. [**Einstein's brain**] Maria is trying to remember a popular quote attributed to Thomas Stoltz Harvey when it was discovered that he stole Albert Einstein's brain. What should she type into the text box? 4. [**Kasparov**] Maria wants to know what Garry Kasparov said when he was defeated by Deep Blue, an IBM supercomputer. What should she type into the text box? 5. [**Tesla**] Maria wants to know how Nikola Tesla reacted when he learned that Guglielmo Marconi won the Nobel Prize for the development of radio technology. What should she type into the text box?

Stimuli priming *where* relative clauses

1.[**Tectonic plates**] Maria wants to know about the geological properties of places where two tectonic plates meet. What should she type into the text box? 2. [**Trees**] Maria wants to know the composition of the soil in forests where trees grow above 300 feet. What should she type into the text box? 3. [**Margaret Mead**] Maria wants to know what the culture where Margaret Mead did her research is most famous for. What should she type into the text box? 4. [**Indian tanneries**] Maria is interested in the economic status of the states in India where tanneries are the main source of water pollution. What should she type into the text box? 5.[**Vaping**] Maria is interested in whether people who live in areas where vaping is illegal are healthier compared to the general population. What should she type into the text box?