Summary

In my dissertation, I introduce a set of measures designed to estimate the syntactic information carried by nouns. I then evaluate the relationships between these measures and processing time in a number of domains and experimental paradigms. This research extends a line of work that originated in morphological processing but which has recently been extended to syntactic analogues of morphological inflection (i.e., 'prepositional cases'; Baayen et al., 2011; Hendrix, Bolger, & Baayen, 2016; Lester & Moscoso del Prado Martín, 2015). For languages that change the form of a word to reflect its syntactic properties (*inflecting languages*), the distribution of the tokens of words across their inflectional variants has been shown to influence lexical comprehension (Moscoso del Prado Martín, Kostić, & Baayen, 2004), production (Baayen, Levelt, Schreduer, & Ernestus, 2008), and acquisition (Stoll et al., 2012). The converging evidence suggests that words with more uncertain (i.e., uniform) distributions across their various forms are easier to comprehend but more difficult to produce. Scaling up, I ask whether supra-lexical – that is, syntactic – distributions influence the processing of individual words.

In the following sections, I define the measures of syntactic variability (henceforth, *syntactic diversity*). Next, I correlate these measures with response times from a previously published database of bare-noun picture-naming data (Bates et al., 2003). Based on the results, I further distill the notion of syntactic diversity to yield six new measures (nine total). I report a replication of the Bates et al. study, again predicting response times as a function of syntactic diversity. I discuss implications of the findings for the representational and functional architecture linking words to syntax. Finally, I present a number of future refinements and extensions, including alternative measures, experiments, processing modalities (comprehension vs. production), and potential implications for cross-linguistic variation.

Syntactic aspects of lexical access

Current psycholinguistic models of lexical access typically link lexical forms to a set of syntactic categories (Carmazza, 1997; Levelt, Roelofs, & Meyer, 1999; Oppenheim, 2010). At the most basic level, words must be specified for their part-of-speech (noun, verb, etc.). For nouns, further subdivisions have been proposed, including gender (Schriefers, 1993; La Heij et al., 1998; Schriefers & Jescheniak, 1999; cf. Miozzo, Costa, & Caramazza, 2002; Schiller & Caramazza, 2003) and countability (Gregory, Varley, & Herbert, 2012). These models either spread activation from word nodes to syntactic category nodes (but not the reverse) or else treat syntactic categories as 'clamps' for the search space for words which have been called up by the syntactic processor. In either case, syntactic information is not thought to be borne by the noun per se, but to mediate structure building and lexical selection. Strong support for this notion came from La Heij et al. (1998), who found that words of the same syntactic gender could facilitate each other (the *gender congruence effect*), but only if the target word was produced in a syntactic context (e.g., DET + NOUN or DET + ADJ + NOUN, for which the gender information was relevant). However, since then, a number of studies have turned up syntactic category effects in non-syntactic production tasks (i.e., bare-noun production; e.g., Cubelli et al., 2005).

Beyond categories, some “empirical modeling” techniques (Chang & Fitz, 2014) have linked syntactic affinities of words to different patterns of production. For example, nouns that are more likely to be modified by a relative clause in general are more likely to occur with zero-marked relative clauses in any given (non-subject-extracted) token (Wasow, Jaeger, & Orr, 2011). Activating the word sets up a syntactic expectation well beyond the simple category information usually studied in experimental tasks. Similarly, verbs are known to associate with particular sets of argument-structure constructions. The presence of these options is known to impact sentence-level planning and articulation (Ferreira, 1996; Myachykov et al., 2013; Hwang & Kaiser, 2014). Such studies hint that syntax and lexicon are more closely related than some (psycho)linguisic models typically allow, especially in the case of nouns. This notion is codified in the constructionist approaches to language, which treat syntax and lexicon as a continuum (Langacker, 1987; Goldberg, 1995). While suggestive, these studies do not directly address the role of that syntactic information in processing the noun itself (as the category-driven experiments do). My dissertation aims to flesh out our understanding of noun processing and its relationship to syntax.

Operationalization of syntactic diversity

I adopt a low-level, aggregate perspective on syntactic relations. In so doing, I necessarily oversimplify the scope of variability and hierarchical organization that has been worked out in astonishing detail in theories of language structure. However, such simplifications are often necessary at the earliest stages of model-building (or of learning more generally; e.g., Elman, 1993). This is particularly true when demonstrating a basic proof of concept.

Selecting a syntactic formalism

How we understand the syntactic distribution of nouns depends absolutely on the formalism we use to represent syntactic relations. Many such formalism exist, so a decision must be made with respect to which one is the most appropriate for the task at hand. For my purposes, four criteria are particularly relevant, all of which can be seen as applying to a syntactic domain of whatever size (morphological, phrasal, clausal, and so on). First, the formalism must provide a means for unambiguously determining the relative sequential position of a target word relative to syntactically related words. Second, it must capture which other words in a given syntactic domain depend on the target noun for their realization and/or interpretation. Third, it should provide ready labels for the *syntactic function*(*s*) served by a given noun. By function, I mean that the formalism should discriminate at some degree of granularity the various types of syntactic relationships that nouns may enter into with respect to other words. Fourth, it should allow for straightforward computer-automated processing.

The first criterion, related to encoding of word order, is satisfied by almost all contemporary formalisms (though some systems designed to handle variable discontinuities in syntactic dependency – *scrambling* – avoid explicitly specifying linear order outside of the linguistic token itself; see Mel'čuk, 1988). Regarding the second and third criteria (which words link up syntactically and by what sort of relation, respectively), formalisms vary with respect to how transparently they reflect this information. Consider perhaps the most widely used syntactic formalism: the *phrase-structure* (PS) tree. PS trees consist of typed nodes and (non-typed) arcs. Nodes represent lexical items (*terminal nodes*) and groups of words (*non-terminal* or *phrasal nodes*). These nodes are connected via vertical arcs that indicate which lower-level nodes are bound to which higher-order nodes (*immediate consituency*; Bloomfield, 1933; Chomsky, 1957). To determine which words are related to a target, one must traverse a potentially complex path via the set of intervening arcs and nodes. While not computationally intractable, the complexity of these paths makes the phrase-structure tree formalism a somewhat cumbersome choice, if only for purposes of exposition.

Furthermore, because the connecting arcs are untyped, information about functional relationships between words must be distributed throughout the tree. Identifying a given relationship requires one to consider at least the types of nodes intervening between the words (if one were to trace a path along the arcs that connect them), as well as the relative positioning of the words with respect to those nodes. For example, the word *stealthy* in the noun phrase *stealthy owl* can only be identified as an adjectival modifier of *owl* in a typical PS tree by (1) its subordination to an AdjP (adjective phrase) node (2) that is left-sister to the word-class non-terminal N node (3) that dominates *owl* and (4), where both AdjP and N are eventually dominated by a single higher-order NP (noun phrase) node. The distributed nature of PS grammars thus presents a somewhat of a challenge for my fourth criterion.

An increasingly popular alternative to the PS notation is the *dependency graph* (DG; Tesnière, 1959; Mel'čuk, 1988). DGs are trees whose nodes represent lexical items and whose arcs represent typed syntactic relations. By convention, DG formalisms tend to include only binary relationships between words, with one privileged relationship linking a word (usually the verb in a finite clause) to the "utterance generator" *root* node[1]. These relationships are known as *dependencies*. Dependencies are asymmetrical, in that one word – the *head* – licenses the presence of the other word – the *modifier* (*governor* and *subordinate* in language of Tesnière, 1959). Heads, their modifiers, and the dependencies that bind them are together known as *constructions*. To avoid confusion between this and other uses of the term construction (e.g., the "construction" of Construction Grammar; Goldberg, 1995), I will refer to the head-modifier-dependency trio as a *bundle*. Much more could be said about the nature of dependencies, including how headship is decided, what types of dependencies are allowed, whether dependencies can only hold between words, or whether supra-lexical (i.e., 'phrasal') nodes should be allowed, and so on. Such questions are addressed from theoretical (e.g., Mel'čuk, 1988; Hudson, 2007) and practical (e.g., Nivre, 2005) perspectives elsewhere. Here, I rely on the standards described by de Marneffe and Manning (2008) for English.

The Stanford DG formalism (SDG; de Marneffe & Manning, 2008) readily meets the criteria I laid out above. While other DG formalisms ignore word order (Criterion 1), SDG supplies for each bundle the sequential position of the head and modifier within the overall string. The question of which words are related (Criterion 2) is replaced by a simpler question – which words are *directly* functionally related. For any target, the set of related words consists of those that are bundled with the target as head or modifier. The nature of these relationships (Criterion 3) is plainly indicated by a functional tag (e.g., *nsubj* for the clausal subject relation). The direct representation of each of these pieces of information within SDG means that it allows for straightforward computer-automated processing (Criterion 4).

SDG is not without its limitations. Perhaps chief among them is that it cannot directly associate meaning to complex structures (i.e., constructions involving more than one dependency). Consider the so-called *caused-motion construction,* which in English takes the form X[*agent*] VERB[*cause* + *move*] Y[*theme*] PREP[*path*] Z[*ground*], as in *Claude flicked the letter through the mail slot.* Fully generalized frames of this sort can be associated directly with various types of meaning. These meanings are revealed through phenomena such as semantic coercion (e.g., intransitive *laugh* takes a force-dynamic interpretation in *The audience laughed the speaker out of the session*), selectional restrictions (e.g., what types of words may surface in each syntactic "slot"), and so on. Importantly, the interpretations derived from instances of the caused-motion construction are not reducible to the content of the component phrases or words. Therefore, any realistic grammatical formalism must be able to account for the non-decompositional meanings that attach to syntactic templates at the phrasal, clausal, and supra-clausal levels (for a compelling discussion of such meanings, see Goldberg, 1995). While I acknowledge the importance of such constructions for our understanding of the *true* syntactic space of any language, I set these concerns aside for later research. I do so for two main reasons. First, no grammar of any language purports to account for every construction in that language. This is no more true of words than it is of higher-order constructions. In fact, the constructionist approach has spurred generations of construction-hunters to identify and catalog their quarry with greater and greater levels of precision. The more constructions are uncovered, the further removed seems the goal of an exhaustive taxonomy. Therefore, even if I wanted to derive syntactic measures from parses reflecting the true syntactic space, I should always face the possibility – in truth, inevitability – of incompleteness. I tackle this necessary incompleteness by assuming only the reality of dependency bundles and the set of typed relations specified by SDG. This assumption comes with the caveat that

1 Root nodes are common to dependency grammars and phrase-structure grammars. However, the notion of root differs across the two formalisms. In DGs, the root connects directly to a lexical item, in agreement with the *lexicalist hypothesis* (Levelt, 1989; Bresnan, 2001). The lexicalist hypothesis states that words project their own syntactic structures to be unified with other co-active words. Syntactic projections are often referred to as *subcategorization frames* (Chomsky, 1965).

we are dealing with only a 'toy' representation of the full grammar. Future improvements may replace or elaborate the relations considered here. Second, while the set of dependencies in SDG is rather small, even a conservative estimate of the total number of syntactic constructions in English is much larger. Moreover, the frequency distributions of these structures should follow a Zipf-Mandelbrot distribution. This means that the expected frequencies for the vast majority of structures are extremely low. Therefore, we cannot expect finite samples of the size typical for syntactically annotated English corpora to produce reliable frequency estimates for the bulk of these constructions. By paring the space of alternatives down, I increase the likelihood of observing a sufficient number of tokens for each type to support – at the defined granularity – reliable estimates of frequency, hence diversity. For these reasons, I consider SDG a desirable and appropriate formalism for beginning my investigation of syntactic diversity within the lexicon (though see my proposals for scaling up in the Future Directions below).

Syntactic diversity as Shannon entropy

The syntactic diversity of a noun *w* has two key components: (1) the set of possible syntactic constructions *C* and (2) the probability that *w* occurs in each construction c in *C.* Nouns should be considered more diverse to the extent that *C* increases. This relationship captures the common sense intuition that nouns that occur in a greater variety of constructions are more syntactically diverse. But this is only half the story. To see this more clearly, imagine that we extract 1000 instances of some noun from a corpus. We find tokens embedded in each of the possible syntactic structures defined in *C*. According to the metric just introduced, the noun exhibits maximal diversity. However, looking more closely, we notice that 900 tokens occurred in a single construction c_1, while the remaining 100 tokens are distributed relatively evenly across the remaining constructions. Now consider a different noun that also occurs in every available construction, but which occurs equiprobably in each construction. According to our first metric, the two nouns are equally diverse. And yet, our intuition suggests that the second noun is much more diverse than the first. The optimal measure of diversity should take both sources of information into account: *instance* (did it occur?) and *rate* (how frequently?). The measure should also account for the full distribution simultaneously (i.e., by taking some central tendency of the noun's distribution across all constructions). One measure that satisfies all of these criteria is the Shannon entropy *H* (Shannon, 1948). *H* measures the average amount of uncertainty within a distribution. Applied to syntactic distributions as defined above, it represents how uncertain we are about assigning a given noun to any of the available constructions. In processing terms, it measures the richness of the spreading activation between lemmas and syntactic structures, with high uncertainty translating into richer patterns of activation. More specifically, *H* increases as the number of possible constructions increases *and* as the frequency distribution across these constructions approaches maximum uncertainty (equiprobability, or equally strong sets of connections). Therefore, holding the dimensionality of the syntactic space constant, the most diverse noun is the one that is least biased towards a particular subset of the possible constructions. *H* has proved useful for estimating the syntactic diversity of full grammars (probabilistic context-free grammars) induced from treebanks (Moscoso del Prado Martín, 2014), as well as for estimating the morphological diversity of words (Moscoso del Prado Martín et al., 2004).

To apply the Shannon entropy, we need to define a syntactic space (i.e., sets of possible constructions) compatible with the SDG formalism outlined above. As a first step, I distinguish between three such spaces. First, nouns may register distributional information about all dependencies to which they belong, irrespective of whether they serve as heads or modifiers. Therefore, I define a total syntactic distribution for which each cell contains the joint probability $p(w, d)$ of target noun *w* in dependency *d.* I refer to the (joint) entropy of this distribution as H_t for *total entropy*. But this measure may be decomposed further. By taking hierarchical status into account, we can dissociate the diversity of relations for the which the noun is a head or a modifier. This decomposition may be necessary given evidence that heads and modifiers are treated differently by the syntactic machinery (e.g., Bürki et al., 2016), which may have consequences for lexical access more generally. Therefore, I define two new distributions consisting of the joint probabilities $p(w, d, r\text{=head})$ and $p(w, d, r\text{=modifier})$, respectively, for target noun w occurring in dependency d in head or modifier role r. I refer to the (joint) entropies of these distributions as H_h and H_m, for *head diversity* and *modifier diversity*. These three distributions are schematized in Figure 1.

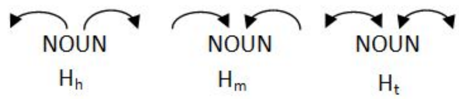

Figure 1: Distribution types for diversity measures. NOUN represents the target name. The arrows point from heads to modifiers (only targets shown). Double-sided arrows indicate that the head/modifier contrast is not taken into account. The position of the arrows reflect linear order of target and its bundle-mate.

Pilot studies: Syntactic diversity in bare-noun object naming

The primary assumption underlying this research is that the syntactic potential of a noun should have implications for lexical processing in the same way that inflectional potential influences recognition latencies for words in case-marking languages. Moscoso del Prado Martín et al. (2004) show for Serbian that the more uncertain the inflection of a given stem, the *faster* it is recognized. Such a finding supports a connectionist interpretation, whereby activation spreads between a central lemma node and its set of inflectional variants. Where this connectivity is evenly distributed across the available forms, the overall lemma receives more efficient support from its inflected variants. This support rapidly boosts lemma activation, leading to faster recognition in visual lexical decision. Morphological inflection presents an interesting case because it translates syntactic relations into formal variants. That is, inflected word forms wear their syntactic functions on their sleeves. For example, Russian собак-а [sobak-a] 'dog-NOM' applies primarily to sentential subjects. Simply knowing the surface form allows us to infer something about its *in situ* syntactic role. By extension, nouns that surface regularly in each of their possible inflectional variants are also nouns that occur relatively evenly across a number of syntactic dependencies. Thus, for inflectionally rich languages, inflectional diversity might partially reflect syntactic diversity. However, for languages with limited morphology such as English, 'inflectional' functions are achieved almost exclusively through syntax. For example, the morphological *cases* of highly inflecting languages correspond largely to phrasal structures in English. Despite the difference in locus, the 'syntactic inflections' of English show an effect similar to that found for case-inflecting languages: nouns that have more diverse distributions across preposition types are recognized (Baayen et al., 2011; Lester & Moscoso del Prado Martín, 2015) and produced (Hendrix, Bolger, & Baayen, 2016) faster. In the terminology introduced above, these studies measure the lexicalized modifier diversity for nouns in a single dependency relation (*pobj*, for 'object of a preposition'). The diversity is lexicalized in that the frequency distribution is defined over prepositions as opposed to abstract syntactic categories. Therefore, these studies address lexical diversity within syntactic constructions.

Pilot Study 1a: Syntactic Diversity of Heads and Dependents

In Pilot Study 1a, I extend the prior research in several ways. First, I consider fully abstract (i.e., non-lexicalized) syntactic relations. Second, I measure the cross-constructional diversity of nouns as opposed to the within-construction diversity studied elsewhere. I follow Hendrix, Bolger, & Baayen (2016) in applying these measures to production. However, the picture-naming task employed there involved comprehension-to-production priming via the presentation of a syntactic context (preposition

+ determiner, e.g., *in the*) prior to the picture. For this reason, their results are ambiguous with respect to the locus of facilitation: should they be attributed to lexical features of the picture name or to aspects of the syntactic processes linking the name to the primed context? We have some hint to the answer: the ERP oscillations for their diversity measure (actually a measure of the point-wise distance between the target distribution and the distribution of the typical noun) were similar in phase to those observed for word frequency. Word frequency is uncontroversially interpreted as a lexical effect. Moreover, the presence of the prime could only interfere with this effect by conditioning the likelihood of the noun. That the effect remains despite the prime, and that it shares its electrophysiological profile with the frequency measure suggests that the latter is most likely also lexical in nature.

In an attempt to clarify this point, I begin with a purportedly non-syntactic task: bare-noun object naming. In this paradigm, participants are presented with images of objects and asked to say their names aloud. The measures of interest are what name is produced and how long it took to produce. Ostensibly, the task is non-syntactic in that the participants are not required to produce any syntactic structure. Empirically, La Heij et al. (1998) report syntactic category effects (the gender congruence effect) for noun-phrase production (unambiguously a syntactic task) but not for bare-noun production. However, the noun-phrase result has also been interpreted as reflecting selection of the appropriate determiner – that is, as a lexical effect (Schiller & Caramazza, 2003). This possibility challenges the syntactic interpretation of the contrast documented by La Heij and colleagues. In addition, more recent research has reported effects from a number of syntactic categories on production latencies in bare-noun production. For example, Duràn and Pillon (2011), using a blocked priming task, found faster response times for trial blocks containing only nouns or only verbs than for blocks containing nouns and verbs (a *word-class-congruence* effect; WCC). Unlike gender congruence, WCC cannot be attributed to additional lexical search functions. Thus, it appears (1) that lexical access may involve obligatory activation of syntactic information, even in the absence of syntactic encoding, and (2) bare-noun naming can tap into this relationship (see also Gregory et al., 2012; de Simone & Collina, 2015).

A common feature of the prior research is the focus on categorial information. In connectionist terms, these studies attempt to preactivate an abstract syntactic category node which shares links among words belonging to that class. Preactivation can facilitate (priming; Gregory et al., 2012) or inhibit (picture-word interference; de Simone & Collina, 2015) access to the target, depending on the task and design. Categorial constraints of this kind figure prominently in models of sentence production (Dell et al., 2008). However, these category labels are actually quite complex generalizations over both morpho-syntactic and morpho-phonological distributions. As such, it is not clear that they actually hold as independent symbols in the lexical network (Schiller & Caramazza, 2003). For example, noun lemmas could connect directly to the set of inflectional and syntactic structures with which they combine, with similar connectivity producing similar effects across lemmas (a 'ghost-category' effect). I do not seek to resolve this rather thorny (and potentially low-payoff) debate. Instead, I side-step the issue altogether by eliciting only one word class – nouns – and measuring the finer-grained distributional information directly. Doing so ensures that any positive effects must be interpreted as stemming from combinatoric potential, regardless of whether a NOUN node is supposed to intercede between (or be co-active with) that information and the lemma.

Data. To establish proof of concept, I re-analyze the English object-naming data that were originally published as part of the International Picture Naming Project (IPNP; Bates et al., 2003). In addition to response times, these data contain a wealth of control statistics, providing a rigorous crucible for my syntactic diversity measures. The original data were collected using a bare-noun naming paradigm. Participants were presented with 520 black-and-white line drawings. They were instructed to name the images with a bare noun (e.g., “Banjo!”) as quickly and accurately as they could, with no hesitations, fillers (*uh/um*), or other words. Participants were given up to 3 s to respond before the experimental software automatically cycled to the next trial. The procedure, design, and stimuli are discussed in more detail in Bates et al. (2003).

From the full dataset, I extracted the following data per image (with the exception of animacy, which I coded manually):

- *dominant response*: the most frequently offered name for each image
- *response time:* mean response time (ms) for dominant responses only
- *dominant response frequency*: log lemma frequency of the dominant responses
- *initial fricative*: whether the word begins with a fricative
- *phonological complexity*: number of syllables (CELEX)
- *name diversity*: Shannon entropy of the set of different names given to the image
- *shared name*: whether an image shared its name with one or more other images (*yes/no*)
- *objective image complexity:* size of image file (kb)
- *subjective image complexity*: perceived number of components within image
- *animacy:* whether or not the image depicts an animate entity

Next, I calculated the three syntactic diversity measures for each dominant response. For this I needed a treebank of English sentences. I selected the freely available Open American National Corpus treebank (a subpart of the American National Corpus; Reppen, Ide, & Suderman, 2005; parsing by Rasul Kalajahi). The OANC is well suited to the present investigation in that it is relatively large and has the advantage of containing predominantly American English (a superlect of the variety spoken by participants in the original IPNP study).The OANC treebank contains 305,290 sentences (roughly 13 million words) parsed automatically using the constituent-based parser of Charniak & Johnson (2005). The parse trees are in Penn-Treebank (phrase-structure) format.

For each dominant response, I extracted every tree that contained that word. Because of the nature of the phrase-structure trees, I limited the search to include only those responses that appear as a single uninterrupted string of word characters (n = 416) and which were tagged as a noun (singular or plural). This search returned 55,069 trees (median = 40 trees per word; mean = 132.38). I then converted these trees into dependency graphs using the Stanford conversion software (de Marneffe & Manning, 2008). I created a frequency vector for each target word by tabulating the number of times that it occurred in each dependency type. The number of cells in the vector was equal to the size of the set union of dependency types occurring with any noun in the sample. I take the Shannon entropy of this distribution to be the total syntactic diversity for a given word, which I abbreviate as H_t. I also created individual vectors for just those dependencies in which the target noun served as head or modifier, respectively. I take the Shannon entropy of these distributions to be the head and modifier diversities, abbreviated H_h and H_m, respectively,. Rather than taking the entropy over the raw probabilities (the maximum-likelihood approach), I smooth the estimates according to the procedure described in Chao, Wang, & Jost (2013). This technique helps to curb the underestimation bias associated with smaller sample sizes. A consequence of this correction is that diversity is decoupled from pure frequency (for a thorough discussion of the suitability of this approach for corpus-linguistic data, see Moscoso del Prado Martín, 2016b). After computing the entropies for each word, I merged the resulting diversity values with the IPNP database.

Because H_h and H_m are calculated over distributions that overlap with that of H_t, the measures necessarily capture some of the same information. If substantial, this overlap can create problems for models that (hopefully) link explained variance to the appropriate source of information. This situation is known as *(multi-)collinearity*. If two or more predictors compete to explain the same variance, then the model will have difficulty estimating the correct strength of the relationship between each variable and the observed data. Coefficients can vary wildly, including changing sign, for a single predictor as increasingly collinear co-predictors are injected into the model (Baayen, 2008). It can also undermine generalizeability to new data (e.g., Dormann et al., 2012). I test for multicollinearity across my three diversity measures using the function *collin.fnc* from Harald Baayen's *languageR* package (Baayen, 2013) for *R*. The analysis revealed a moderate amount of collinearity (k = 24.52). This value falls within the acceptable range by some standards (e.g., $k <= 30$ from Baayen, 2008). Nevertheless, I attempt to clean what collinearity there is from the data. To accomplish this, I apply an Independent Component Analysis (ICA) using the fastICA algorithm as implemented in the *fastICA* library (Marchini, Heaton, & Ripley, 2013) for *R*. The fastICA algorithm reconstructs a set of maximally independent “source” signals (the number is determined *a priori*) whose linear transformation produces the observed n-dimensional signal. The resulting signals, or *components*, can be interpreted relative to the original predictors by examining the so-called *mixing matrix.* The mixing matrix contains the n coefficients that must be applied to the points in (a normalized and decorrelated version of) the original n-dimensional space to project them into the 'rotated' component space. The coefficients are known as *loadings* and reflect how strongly each predictor influences the position of points in the transformed component space. For the current analysis, I generated three components (equal to the number of predictors). The loadings for the three diversity measures within each component are plotted in Figure 2 below. As Figure 2 shows, the two hierarchically sensitive measures H_h and H_m influence the data in relatively strong and distinctive ways (the contrast between components 1 and 2), though, as expected, both overlap with total diversity (component 3). For modeling purposes, I substitute the components (i.e., the estimated source matrices) for the original variables.

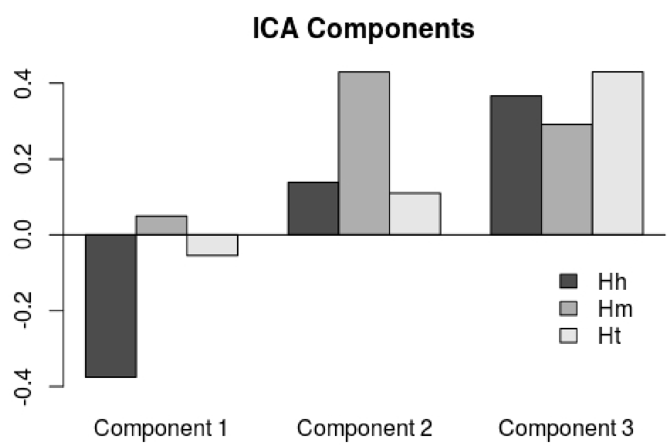

Figure 2: Loadings of diversity measures across independent components. Component 1 is most strongly associated with H_h and Component 2 with H_m. Component 3 reflects information common to all three diversity measures.

Results. I model the data with a *general additive regression* predicting (untransformed) mean response times from the Bates et al. (2003) data. I include each of the variables listed above as predictors, minus the diversity measures (which were substituted by the ICA components). I also include smoothing via thin-plate regression splines for a subset of the numeric predictors: *dominant response frequency*, *name diversity*, *objective image complexity*, and the three ICA components. Spline-based smooth terms balance tightness of fit against model complexity to avoid overfitting while accounting for possible non-linearities. The other numeric predictors had too few distinct observations to accommodate the smooth. Beginning with the maximal model (i.e., the model consisting of all predictors), I performed a step-wise backward elimination of non-significant factors until all remaining predictors were significant at $\alpha = .05$. Table 1 summarizes the resulting *minimal adequate model* (Zuur et al., 2009). Coefficients, error, and significance level are reported for the unsmoothed terms; only significance levels are reported for the smoothed terms (parameter estimates cannot be calculated for smoothed terms).

Table 1: Results of generalized additive regression (RT data from Bates et al., 2003)

Unsmoothed parameters

Term	***b***	**SE**	***p***
initial fricative (fricative → no fricative)	-33.38	14.57	<.05
shared name	-62.48	28.61	<.05

Smoothed parameters

Term	***p***
name diversity	<.001
dominant response frequency	<.001
ICA component 1	<.001

Overall model stats: adj. $R^2 = .58$

First, the significant unsmoothed control variables. As expected, initial frication delayed naming. The model revealed an approximately 33 ms processing advantage for words without initial frication. This effect presumably arises at later stages of lexical processing (i.e., during articulation). A much stronger advantage was revealed for words that shared their dominant name with one or more other images, ~ 62 ms. This advantage likely arises from a form of satisficing (e.g., Simon, 1956) during lexical search. Names that are shared across multiple images could be semantically more flexible, perhaps due to greater generality (in the extreme, any of the images could have been referred to as *thing*). If one can classify an image at a higher level of generality, fewer of the distinctive semantic features associated with the image would need to be accounted for during name selection. Hence, facilitation.

The results for the smoothed terms are plotted in Figures 3-5 below. In these graphs, the x-axis represents the variable in question, the y-axis represents the estimated effect on reaction times, the dark line represents the smoothed regression curve, and the shading indicates the 95% confidence bands for the predicted effect at each value of the predictor. The dotted horizontal line marks 0 ms (absence of effect). Note the y-axis scales: these predictors have very different effect sizes. In particular, the effects of the two smoothed control variables – *name diversity* (Figure 3) and *dominant response frequency* (Figure 4) – are much larger than the syntactic diversity effect.

Figure 3 shows that *naming diversity* had a strong inhibitory effect on response times: when participants tend not to select the same name for an image, response times for that image increase (even though the modeled response times come exclusively from dominant responses). The rate of increase is ~ 226 (± 11) ms/nat of naming uncertainty. Figure 4 shows that *dominant response frequency* facilitated response times at a rate of roughly -28 (± 5) ms/nat. These two effects were also the strongest predictors of response times in the original analysis (Bates et al., 2003), even though that model differs substantially from the one reported here.

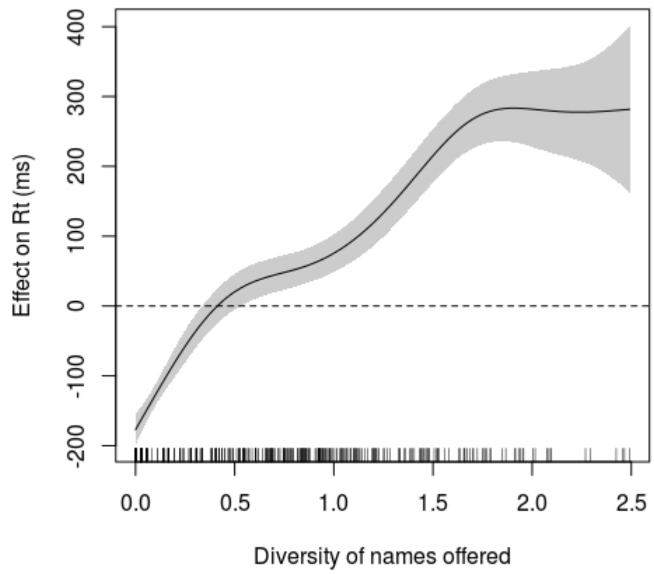

Figure 3: Effect of *name diversity*. When participants tended to agree on the name of an images, the names were produced faster. To the extent that participants disagreed, naming was slowed at a rate of ~226 ms/nat.

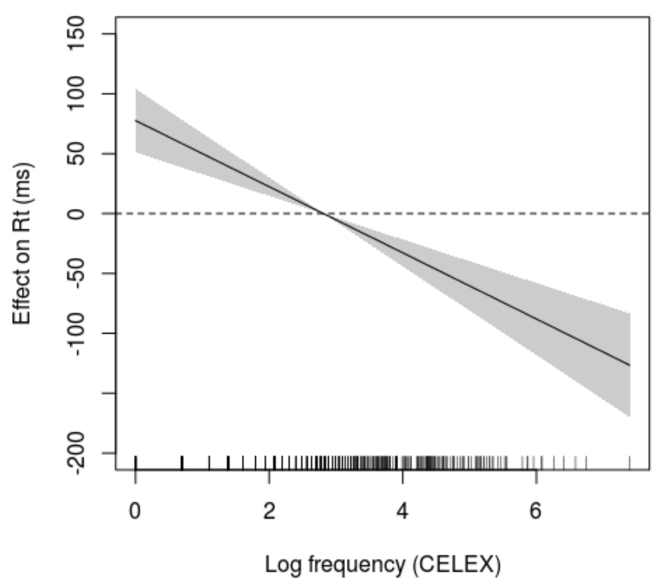

Figure 4: Effect of frequency. The more frequent the name, the faster it is produced (at a rate of ~28 ms/nat). As usual, the effect is (roughly) linear in the logarithmic scale.

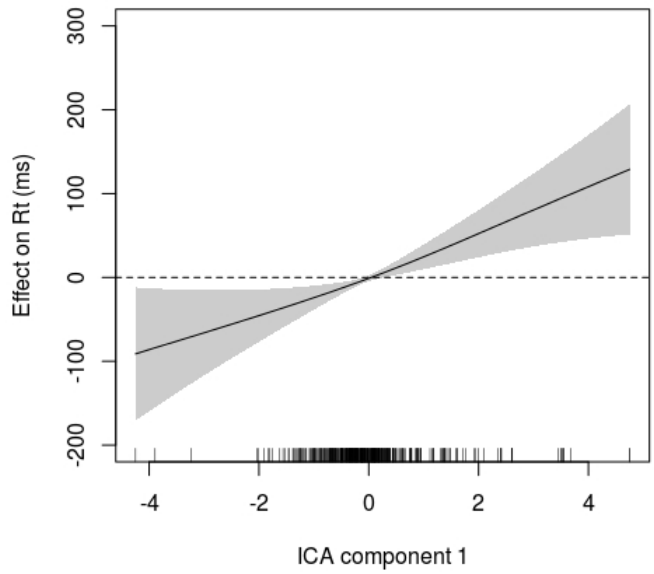

Figure 5: Effect of ICA component 1. As shown in Figure 2, ICA component 1 loads negatively for H_h, with little connection to H_m or H_t. This plot shows that words with overall higher headship diversity (moving from 0 to -4) are produced faster than words with low headship diversity (moving from 0 to +4). The linear effect amounts to an increase of ~27 ms/unit.

Finally, only one of the ICA components proved significant. Component 1 (Figure 4) loads negatively on H_h and correlates positively with response times at a rate of ~27 (± 7) ms/unit. This pattern suggests that nouns that serve as head for a diverse array of syntactic dependencies are accessed faster than words that head only a limited subset of the possible dependencies. The loadings in Figure 1 suggest further that this effect is largely independent of modifier diversity or overall diversity.

Discussion

Syntactic diversity was predictive of response times, suggesting that when we prepare single words for production, we obligatorily access their syntactic information (Cubelli et al., 2005; de Simone & Collina, 2015; Duràn & Pillon, 2011; Gregory et al., 2012). This finding is in-line with a spreading activation account, whereby words are linked bidirectionally to nodes representing syntactic relations. The bidirectionality is implied by the fact that the syntactic information was not actively deployed (no other words were produced in relation to the target) but nevertheless affected production speed. Activation spreads from the word representation (sometimes referred to as a *lemma*), to the syntactic network, and *back again*. This finding is not compatible with theories of uni-directional spreading from words to syntax (Caramazza, 1997; Levelt et al., 1999), but could fit the selection-through-boosting approach of Oppenheim (2010). In this case, however, the boost would not be an arbitrary flood of activation (as it is in Oppenheim's model), but principled, lexeme-specific feedback from the peripheral networks, including the syntactic network. The results also hint that the relationship between nouns and syntax is not blind to function: headship diversity had an independent effect on response times.

The results suggest that hierarchically informed distributions better discriminate differences in processing speed than general syntactic diversity. However, the current formulations overlook a major aspect of English syntax, namely, word order. Word order is important for English, not only structurally (English is a highly *configurational* language; Hale, 1982), but in terms of processing. For example, the possibility of variable word orders has been shown to *decrease* sentence onset latencies in English (Ferreira, 1996). These results have been taken as evidence for a radically incremental parser that operates most efficiently when maximizing the space of possible continuations. Similar effects might play out during the planning and production of noun phrases. If this is true, then perhaps only rightward-facing dependencies are of any consequence for determining speed of access. In the next section, I introduce six new measures to test the importance of word order relative to hierarchical function and apply these measures to the Bates data.

Pilot study 1b: Syntactic diversity of left- and rightward heads and modifiers

I account for word order by splitting each of the three distributions introduced above into two vectors, resulting in six new vectors (nine total). Each new vector reflects the probability distribution over the joint probabilities of a word *w* with each possible dependency *d* in a particular functional relation *r* in a particular sequential direction *s*, or *p*(*w, d, r, s*). *r* can take the value *head* or *modifier.* The critical new factor *s* can take the value *rightward* or *leftward.* These new measures are presented below in Figure 6.

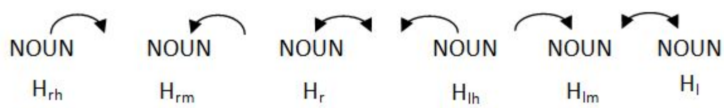

Figure 6: Refined distribution types for measures, now reflecting word order (right or left). Placement of arrow relative to the target NOUN indicates sequence; arrows point from heads to modifiers; double-headed arrows indicate that the headship contrast is not considered.

The contrast between rightward and leftward dependencies can be reframed in typological terms as the difference between head-final and head-initial structures (Liu, 2010). Leftward headship and rightward modifiership constitute head-final structures. Rightward headship and leftward modifiership constitute head-initial structures. English strongly prefers head-initial structures. Therefore, we expect more efficient processing of head-initial as opposed to head-final structures. That is, we expect facilitation for nouns with high head-initial diversity and inhibition for nouns with high head-final diversity.

Data and Methods. The data were the same as those in Pilot Study 1a, with the addition of the six new diversity measures (Figure 6). To compare the relative effects of hierarchical and word-order information, I include all nine diversity measures in the analysis. As in the previous analysis, the threat of multicollinearity is high for these variables, given that they are drawn from nested distributions. Indeed, collinearity between the variables is unacceptable ($k > 65$). I again clean the collinearity via an independent component analysis. Simply creating the same number of components as predictors (= nine components) as I did before may force the algorithm to draw difficult-to-interpret contrasts). Instead, I first run a principal component analysis (PCA) on the data. I then look at the cumulative explained variance of the resulting components and take the number of components that account for 95% of overall variance. While the algorithms underlying PCA and ICA are different, PCA serves as a rough-and-ready heuristic for an otherwise arbitrary decision in ICA. Based on the PCA, I generate five new independent components using the fastICA algorithm. The mixing matrices for each of the components are plotted in Figure 7.

Results. I performed another generalized additive regression to predict mean response times. The model was the same as that reported in Pilot Study 1a, except that the five new independent components replaced the three components analyzed there. As before, the components were smoothed using thin-plate regression splines. All other predictors were treated in the same manner as reported above. I began with the maximal model and performed a step-wise backward elimination of factors until reaching the minimal adequate model. Partial effects and significance levels are reported in Table 2 below.

Table 2: Results of generalized additive regression (RT data from Bates et al., 2003)

Unsmoothed parameters

Term	***b***	**SE**	***p***
initial fricative (fricative → no fricative)	-35.97	14.44	<.05
shared name	-57.76	28.47	<.05
subjective visual complexity	22.37	11.20	<.05

Smoothed parameters

Term	***p***
name diversity	<.001
dominant response frequency	<.001
ICA component 3	<.05
ICA component 5	<.001

Overall model stats: adj. R^2 = .58

The controls had effects nearly identical to those reported above, so I do not discuss them further. The novel findings concern ICA components 3 and 5. Figure 8 shows that or every unit increase in ICA component 3, response times increase by ~ 19 ms (± 7 ms), somewhat smaller than the effect size of component 1 from Pilot Study 1a. The data also suggest a quadratic shape to the effect, with a null correlation gradually bending into a positive correlation. To interpret this effect, we turn to the loadings of the predictors, which are shown in Figure 9. The negative space of component 3 reflects high rightward diversity across heads and modifiers; the positive space loads weakly for $H_{(l)m}$. However, notice that the confidence band overlaps with 0 in the negative space, while deflecting upwards for the positive space of component 3. This pattern suggests that words with strong rightward diversity are not necessarily facilitated; rather, words incur a cost by *not* providing for enough rightward flexibility.

The positive space of component 5 reflects high headship diversity (Figure 11). The negative slope of the effect (Figure 10) shows that nouns are produced faster when they imply a more diverse set of dependencies as head *but not as rightward modifier*. Conversely, nouns are produced more slowly when they imply a more diverse set of (rightward) dependencies as modifier *but not as head*. The strength of this effect is comparable to that uncovered in the previous analysis, falling in the range of ~28 (± 7 ms) ms/unit.

Discussion. By introducing word-order information, I revealed a more fine-grained pattern of effects, tuned primarily to rightward direction. As before, headship diversity is generally facilitatory. This effect stems primarily from the rightward direction, but appears to receive support from other domains as well. The positive loading of H_{lm} suggests that words that can more easily *complete* a projected structure are also preferentially accessed. Taken together, the effects of components 3 and 5 suggest that facilitation occurs in the absence of H_{rm},, and is particularly associated with headship diversity.

This pattern of effects raises two important questions: why do English production latencies show a special sensitivity to rightward syntactic diversity, and why should contrasts in hierarchical function push latencies in different directions? The first question receives a partial answer from research on sentence production, which finds that English speakers begin sentences earlier when they have more structural options for completing the sentence (Ferreira, 1996). This effect was interpreted as evidence for a radically incremental production system that prefers to maximize fluency by eliminating in as much as possible the need for long-distance advance planning. Presumably, local features at the choice point will resolve uncertainty about how to proceed, allowing the speaker to flexibly encounter a broader array of potential biases. As Swets & Ferreira (2002) point out, this may be just one of many production modes, with the choice being based on the current task demands. Ferreira (1996) asked participants to produce the utterances as quickly as possible, which has been shown by Swets & Ferreira to encourage incremental as opposed to advanced processing. Returning to the picture naming data, Bates et al. (2003) also asked participants to produce the names as quickly as possible. Thus, the symmetry between the findings may be a reflection of the task. However, the facilitation we observed was reserved for noun-as-head. Why should diverse nouns-as-modifiers interfere with facilitation?

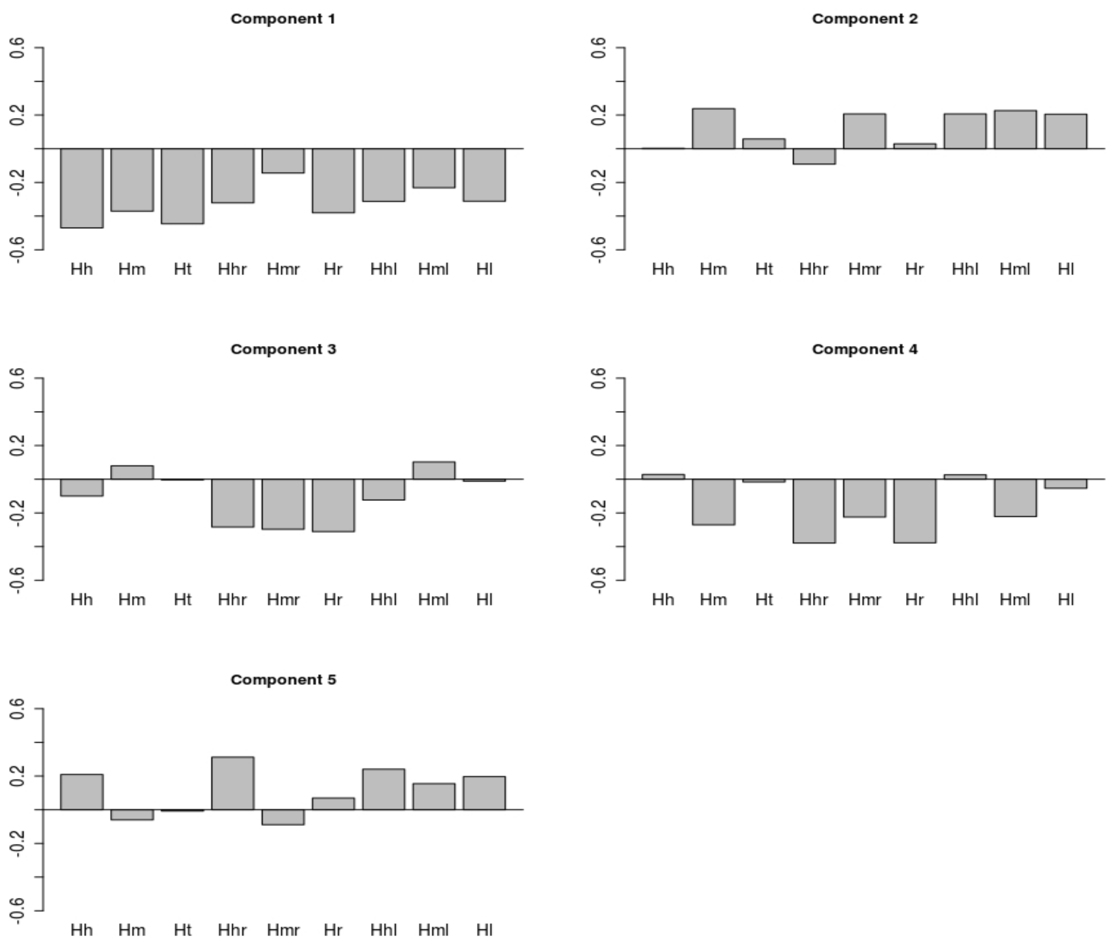

Figure 7: Loadings of the diversity measures across components.

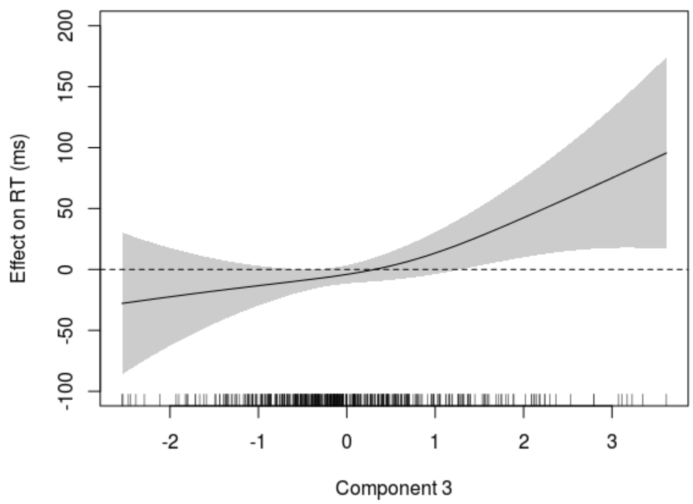

Figure 8: Effect of ICA component 3 on mean response times. Moving from 0 on the x-axis to the left, we see a marginal facilitative effect of H_r, or rightward-facing diversity. Moving from 0 to the right, we see an inhibitory effect of *lack* of rightward diversity (with small contributions from $H_{\{l)m}$).

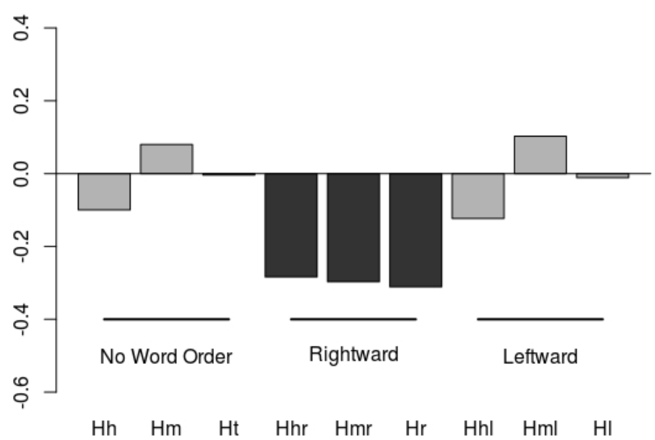

Figure 9: Component loadings for ICA component 3. Negative loadings for rightward diversity in general. Slight positive loadings for modifier-ship diversity, mostly from leftward modifier-ship diversity.

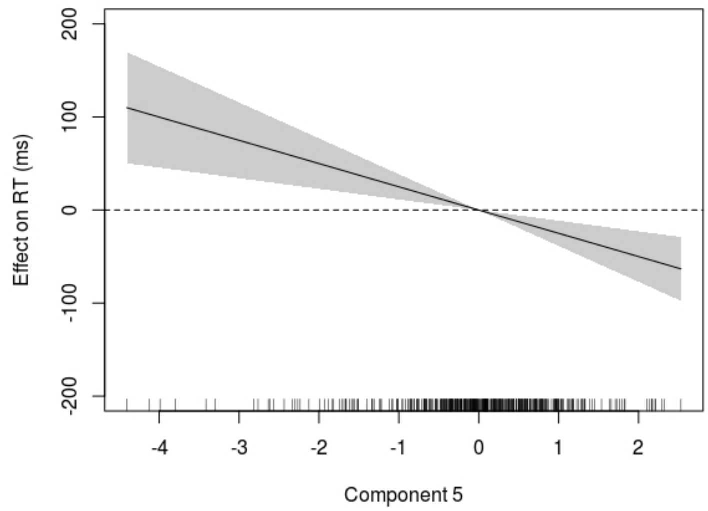

Figure 10: Effect of ICA component 5 on mean response times. Moving from 0 on the x-axis to the left, we see a facilitatory effect of H_{rh}, or rightward-facing headship diversity (with some support from H_h and H_{lh}). Moving from 0 to the right, we see an inhibitory effect of H_m and H_{rm}, or rightward-facing modifier-ship diversity.

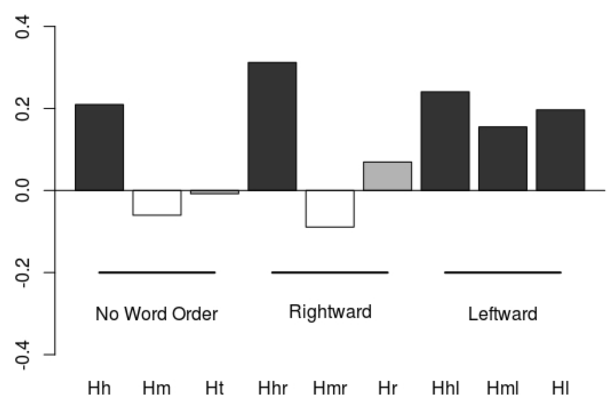

Figure 11: Component loadings for ICA component 5. Positive loadings for headship diversity in general, and in particular, H_{rh}, or rightward-facing headship diversity, with support from leftward diversity in general. Negative loadings for modifier-ship diversity, mostly from rightward modifier-ship diversity.

The present data do not allow us to pinpoint the source of the modifier inhibition, but several possibilities present themselves. First, it may arise from the combination of the dependency formalism and task. In particular, the task involves selection of an utterance, for which the noun must serve as modifier. SDG does not allow 'non-rooted' graphs, or structures with a single head. Instead, every utterance is presumed to attach to a common ROOT node, which serves as an initial head. ROOT projects a modifier (usually a verb in clause-level parses) which serves as the entry point to the remainder of the graph. Thus, in the single-word frames elicited here, the speakers would be forced to anchor the noun to ROOT as modifier. If there is a representational distinction between head and modifier relations, then the inhibitory effect could reflect competitive interference where a decision among the possible modifier relations must be made. This explanation also captures why headship diversity should *not* exhibit interference in this task: no (rightward) modifier relations are actually selected.

Another explanation could be that modifiers and heads are processed asynchronously, and that over time these temporal offsets become codified as general activation potentials for the lexical nodes. Recent electrophysiological research suggests that determiners (the modifiers) are processed *after* nouns (heads), even though the words are produced in the opposite order (Bürki et al., 2016). If this asynchrony generalizes to all types of heads and modifiers, then nouns that habitually serve a wide array of modifier relationships might accumulate inhibitory connections leading to a slow-down in overall speed of access. The opposite would hold of nouns that habitually serve as heads, in that heads are given sequential priority.

Pilot Study 2: Replication of Bates et al. (2003)

The effects have so far been consistent; however, we have only explored one dataset. Furthermore, the mean response times we have considered so far are not ideal for such fine-grained analyses. For example, they limit the power of the analysis by greatly reducing the number of data points. They also mask individual differences in the way that participants react to the different stimuli. Therefore, to test the generalizeability of the effects, and to clarify their shapes and magnitudes, I collected new data using the same methods reported in Bates et al. (2003).

Stimuli and design. Image stimuli were taken from the set of 520 black-and-white line drawings of common objects that were used in the original IPNP research. Each participant saw 200 of the original 520 images. These 200 images were randomly selected at the onset of each experimental session, meaning that each participant saw a unique set of images. Paring down the stimulus set in this way made for a more expedient replication. Randomizing the subsets guards against unintentional sampling biases in pursuit of this expediency. The 200 randomly selected images were further grouped into 4 sets of 50 (groups assigned randomly). Order of presentation within these groups was of course also random. All images and text were presented in black on a white background.

Participants. 48 undergraduates from a public university on the west coast of the United States were recruited to participate (N(female) = 37; mean age = 20.85), all of whom were native speakers of English with normal or corrected-to-normal vision. All participants were treated in accordance with the American Psychological Association guidelines for ethical human research.

Procedure. The experiment was carried out in a dimly lit, sound-attenuated room. All experimental materials were presented via the experimental software OpenSesame v. 3.1.2 (Mathôt, Schreij, & Theeuwes, 2012). At the outset, participants were told that they would be shown a series of images, and that their task was to say the name of each image aloud. They were instructed to say the name in isolation (“Banjo!”) as quickly and accurately as possible, and to avoid producing hesitations or fillers prior to saying the word. Finally, they were told that they had a maximum of three seconds to name the image before the next trial would begin, and that they should remain silent if they could not find an appropriate name for the image before timeout. Before beginning the proper trials, participants were trained on a set of three images taken from an alternative set of images (Bonin et al., 2003). These images were the same for all participants and were selected so as not to overlap with images from the IPNP set. During the training and critical trials, participants saw a fixation cross for 250 ms, followed by a white 'blankout' screen for 500 ms, then the image for a maximum of 3 s. Images were presented in four 50-image blocks with opportunities to rest after each of the first three blocks (to minimize fatigue effects). Responses were recorded with a Sony ECM-909 stereo microphone set to 90-degree spread. Responses were transcribed, and response times were coded by hand using the audio-editing software Audacity[2].

Response processing. Following Bates et al. (2003), I only analyze the dominant responses. This is because participants who offer unexpected names for images may have been experiencing other difficulties at the time of processing. However, the random-subsampling approach applied here precludes my calculating new dominant responses. This proves not to be too much of an issue. The proportion of matches between the newly elicited responses and the dominant responses listed in IPNP was 60% or greater for ~80% of the original 520 images (images with majority other-names consisted of < 9% of the total pool). Hence, the participants in the present sample tended to produce names in accord with those produced in the original IPNP study. Because I analyze the same names listed in IPNP, I can also carry over the full set of norming statistics from that database.

De-correlating predictors. I again de-correlated the syntactic diversity measures ($k > 60$) using fastICA. I recompute the ICA to account for the random element of the fastICA algorithm. This random element may lead to differences in the directions of loadings and slight variability in the strength of the loadings. An initial PCA suggested that five orthogonal components can account for ~95% of the original variance. Therefore, I computed five independent components (these components were extracted simultaneously as opposed to sequentially). Component loadings are plotted in Figure 8. Notice the similarity between Component 3 from Pilot Study 1b (Figure 7) and Component 5 in the present analysis: both contrast $H_{(r)h}$ and $H_{(r)m}$.

Data processing. To better approximate the assumptions of the model, I removed all outlier response times, defined as those which fell two standard deviations above or below the overall mean (5% of tokens). I then applied an inverse-square transform to the remaining values to correct for heavy rightward skew (suggested by a Box-Cox power analysis; $\lambda = -2$). Model comparison validated these transformations of the data: residuals from untransformed-RT models were non-normally distributed, with a strong tendency to overestimate RTs in the middling range; residuals from the transformed-RT models were well-behaved.

Results. I again performed a generalized additive regression predicting (transformed) response times. To account for multiple observations per stimulus, I include random intercepts for image. To reduce autocorrelational artifacts (whereby participants are heavily influenced by their performance on prior trials – beyond the scope of the variables of interest), I included a variable *sequence,* which reflects the order of presentation per subject. I added by-participant penalized factor smooths for *sequence* to account for differences in the autocorrelative behavior between participants. Finally, I added a first-order (t-1) autoregressive process relating immediately prior error to target error by a proportion of $\rho = .3$ (an arbitrary decision; however, its success in cleaning out correlated error was validated by visual inspection of the resulting autocorrelation curves per participant; see Baayen, van Rij, de Cat, & Wood, *to appear*).

As controls, I included all of the measures from the prior analyses. Spline-based smooths were applied to *dominant response frequency, name diversity,* and *objective visual complexity.* I added participant-based variables for *age* and *sex.* Critical predictors were the five ICA components. Each was fit with a spline-based smooths, as well. Beginning with the maximal model, I performed a

2 http://audacityteam.org

backward selection of factors until obtaining the minimal adequate model. Model stats are shown in Table 3[3].

Table 3: Results of generalized additive regression (replication)

Unsmoothed parameters

Term	***b***	**SE**	***p***
animacy (animate → inanimate)	2.6 X 10^8 (31.3)	9.3X10^9 (10.9)	<.01

Smoothed parameters

Term	***p***
name diversity	<.001
dominant response frequency	<.001
objective visual complexity	<.05
ICA component 1	<.05

By-trial parameters

Term	***p***
sequence-by-participant	<.001
image (intercept adjustments)	<.001

Overall model stats: adj. R^2 = .32

The model revealed significant effects for the autocorrelation and image terms, confirming the presence of structured subject- and item-bound variance beyond that which can be attributed to the other predictors. Unlike the prior analyses, we observe a facilitatory effect of *animacy,* with a ~20 ms advantage for images depicting animate entities over inanimate entities. This model also differs from the previous ones in that *objective* rather than *subjective visual complexity* was a better predictor of response times. For every 10,000 bytes of additional visual information, response times increased by ~6 ms (± 4 ms). Another notable difference between this and the prior models is the lack of an *initial fricative* effect. While this variable was removed as a non-significant contributor at an alpha of .05, it did exhibit a marginal trend (p = .06) in the expected direction.

The effects that were shared by this and the previous models tended to be highly attenuated. *Dominant name frequency* showed facilitation at a rate of -13.8 ms (± 2.8 ms) per nat increase – about half the previously observed effect. Likewise diminished was the effect of *name diversity*: a one-nat increase resulted in only a ~84.1 ms (± 7 ms) increase in naming latencies. This effect size is much smaller (by a magnitude of 3) than was reported in Pilot Studies 1a/b. One possible explanation is that this measure depends heavily on the sampling population. Participants in the original IPNP study may have *actually* had a broader array of potential names to consider than those in the present trial. Another possibility is that halving the number of trials led to a general decrease in competitive activation across

3 Coefficients are given in the transformed scale for unsmoothed parameters. Approximate transforms into the response scale (ms) are offered in parentheses.

name types (fewer images named = fewer semi-active competitors). Thus, the *functional* diversity may have remained unchanged, while the lower overall competition drained the potency of the effect. Rough support for this explanation comes from the fact that the *name diversity* effect shifts upward slightly – particularly at the upper extremes – as *sequence* increases (though the interaction does not reach significance in a linear model). That is, naming takes numerically longer for images with high name diversity when more images have previously been named. While interesting, this trend lies outside the scope of the present study. Most important is the fact that the effect of name diversity remains significant, in the same direction, and with the same smooth shape as reported by Bates and colleagues.

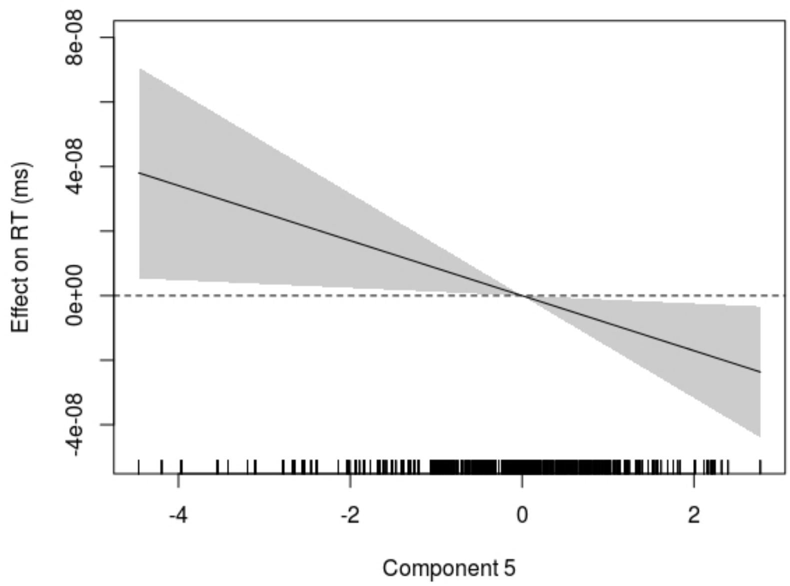

Figure 12: Effect of ICA component 5 on mean response times. Negative values reflect $H_{(r)m}$. Moving from 0 on the x-axis to the left, we find an inhibitory effect of $H_{(r)m}$. Moving from 0 to the right, we see the expected facilitatory effect of (rightward) headship diversity (with small contributions from $H_{l(h)}$).

The only ICA component to surface as significant was *component 5*, repeated as Figure 13. This component contrasts $H_{(r)h}$ and $H_{(r)m}$ similarly to component 3 from Pilot Study 1b, with a slightly heavier loading of H_{lh} in the same direction as the other headship-based measures. In this case, the headship diversity measures load positively while the modifier diversities load negatively. The effect of *component 5* is plotted in Figure 12. The negative slope replicates the previously observed effects: headship diversity, particularly rightward headship, is facilitatory; modifier diversity, particularly in the rightward direction, is inhibitory. The rate of facilitation/inhibition is ~9 ms (± 4 ms) per unit.

Discussion. A replication of the original bare-noun object-naming study of Bates et al. (2003) showed a similar pattern of effects to those reported in the previous two analyses. Crucially, the syntactic diversity measures had a similar effect on response times. It seems that accessing a noun obligatorily involves accessing the syntactic information of that noun. Moreover, this replication supports the interpretation that syntactic information is functionally differentiated. More specifically, there appears to be a close relationship between hierarchy and word order, generating interesting predictions about how processing efficiency might be managed in languages with different order/hierarchy correlations (for example, head-final languages such as Japanese; more on this below). All of this from a 'non-syntactic' task. Finally, the reliability of these effects across datasets (disregarding the differences in effect size) speaks to the generalizeability of the finding, but also the validity of the IPNP data as norming material.

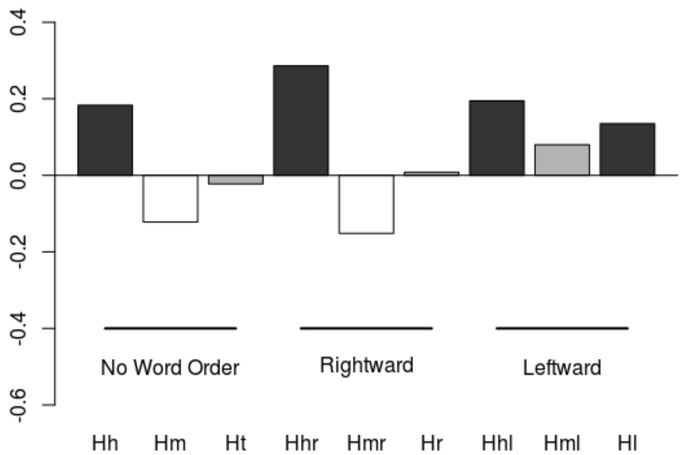

Figure 13: Component loadings for ICA component 5. Positive loadings for headship diversity in general, and in particular, H_{rh}, or rightward-facing headship diversity, with support from leftward (headship) diversity. Negative loadings for modifier-ship diversity, mostly from rightward modifier-ship diversity.

General Discussion

The primary goals of this research is to document and measure the types of syntactic information that are functionally relevant for lexical access. Most previous psycholinguistic research has focused on how membership within syntactic categories (e.g., part-of-speech, mass/count, gender, and so on) affects the processing of words. However, modern linguistic theory suggests that category membership is only half the story, and that words must also reflect their specific combinatoric potential. This is the difference between coarse-grained generalization, which defines a *possibility space* ('X belongs to the class NOUN; members of this class are licensed in constructions A, B, C, ,,,') and fine-grained distributional information, which defines the *probability space* (X occurs with probability *P* in constructions A, B, C, ...'). I argue that the former is a simply a bad approximation of the latter, that syntactic similarity (as implied by part-of-speech labels) is derived primarily from the bottom up, and that processing should primarily reflect distributional information (with the understanding that categories are lossy summaries of that information).

To reach these goals, I proposed a set of measures sensitive to the syntactic distributions of words. As a first approximation, I defined the 'basic unit' of syntax as the functionally typed *dependency,* adopted from the general class of Dependency Grammars (e.g., Mel'čuk. 1988; Hudson, 2007; Nivre, 2005; de Marneffe & Manning, 2008). I further contrasted two dimensions of syntactic information: hierarchical relationships (head/modifier) and word order (leftward/rightward relationships). I defined syntactic information as the diversity (smoothed Shannon entropy) of the probability distributions of words across the set of dependency types. In a set of exploratory analyses, I correlated the syntactic predictors with bare-noun picture naming latencies. Importantly, this paradigm should have minimal impact on syntactic processing because the participants are not required to produce syntactic structures (beyond the marginal 'single-word utterance frame'). The studies converged on two principles, which I shall provisionally refer to as *head priority* and *modifier delay.* The *head priority* principle states that nouns that imply a more diverse array of modifying relationships should be processed faster. This principle surfaces in the sentence production literature as the so-called *opportunistic hypothesis.* The opportunistic hypothesis states that we prefer to plan ahead as little as possible, and that we buttress our short-sightedness by maximizing the options for continuation at any given *choice point* (the point in speaking at which we have 'committed' to a particular syntactic structure). For example, Ferreira (1996) showed that speakers were quicker to begin speaking when the stimulus allowed for either a double-object or prepositional-object encoding (e.g., *Hardy handed Littlewood the proofs; Hardy handed the proofs to Littlewood*); they were slower to speak when the stimulus committed them to one or another. Crucially, Ferreira showed that this effect could be achieved by choice of verb: verbs that strongly selected for one or the other construction led to slower onsets, while verbs that were equally suitable in either construction led to faster onsets. While Ferreira looked at verbs, the syntactic distributions underlying his manipulation are fundamentally comparable to those described here for nouns. He is essentially estimating the entropy of (a portion of) the rightward headship distribution. According to most dependency grammar accounts, verbs are head to all of their arguments. All else being equal, verbs that occur with both constructions will provide broader coverage of the probability distribution (higher entropy) than verbs that occur with only one. Translating then, higher rightward headship diversity of the verb (higher head-initial diversity) correlates with faster sentence onsets. This effect could be construed as a scaled-up version of the lexical effects considered in the present study. Hence, *head priority* may be considered a special case of the *opportunistic hypothesis.*

Modifier delay states that words that, as modifiers, imply a more diverse array of upcoming headship relations are processed more slowly. This principle does not have a direct precedent to my knowledge; however, it does fit with some recent electrophysiological evidence which suggests a temporal asymmetry in the processing of heads and modifiers. In a picture-word interference study, Bürki et al. (2016) found that ERP deflections prompted by phonological similarity of the distractor preceded those prompted by gender congruence. The former is associated with the latest stages of lexical processing (Schriefers, Meyer, & Levelt, 1990), while the latter is interpreted as priming of the appropriate determiner (Schiller & Caramazza, 2003; cf. La Heij et al., 1998). Thus, phonological encoding (a late process) is completed for the head prior to the lexical access (an early process) of the determiner. That is, heads are processed before their modifiers, at least with respect to simple determiner phrases for a language with head-final order (I, along with the dependency formalism I adopt here, reject the notion that determiners serve as heads to the DET + NOUN complex). If this pattern reflects a general property of language production, then the repeated use of a noun as modifier to upcoming heads could lead to a generalized inhibitory effect. This account also takes care of *head priority*: heads are processed quickly to allow time for accessing the set of modifiers. We learn to impose a processing delay to facilitate the typical syntactic demands of the word in context. Another explanation involves aspects of the task. Syntactic formalisms typically require an access point into the structure of any stand-alone unit (e.g., the independent clause). In dependency grammars, this access point is known as the ROOT relation. ROOT is a headless dependency which takes a modifier. The modifier of ROOT is considered the syntactic 'seed' of the remaining structure, and stands at the head (beneath ROOT) of the full dependency graph. For a single-word utterance, there are two possible analyses. Either a full sentential context is generated but suppressed, in which case ROOT attaches to a zero-morph serving as the (elided) sentential seed; or ROOT may initiate non-clausal utterance frames, including single words. If the latter is true, then the bare-noun production task requires the participant to formulate a noun as modifier to ROOT. That is, the speaker must choose from among the possible modifier relations. However, since the speaker is instructed not to produce any modifier for the noun, no selection is made with respect to that noun's possible headship relations. The choice to anchor the noun to ROOT may therefore reflect a straightforward interference effect: to make the correct selection, a number of competitors (whose competitive strength is indexed by modifier diversity) must be overcome, leading to a processing delay. No interference effect is found for headship diversity because there is no competition. Instead, the facilitation for heads may reflect excitatory feedback from the syntactic network, unencumbered by the intention of the speaker to attach a modifier to the head noun.

The above discussion is necessarily speculative, but the findings do have more substantial consequences. First, methodologically speaking, the effects reported here add to the literature which belies the non-syntactic characterization of the bare-noun picture-naming task (Cubelli et al., 2005; de Simone & Collina, 2015; Duràn and Pillon, 2011; Gregory et al., 2012). Previous studies in this vein differ from the present study in that they have (a) always employed additional manipulations (blocking/priming, interference) and (b) considered only categorial information. The present study adds to this literature by showing that fine-grained syntactic-distributional information also influences single-word production. Second, the results challenge any model of lexical access that (a) limits activation of syntactic information to cases of full syntactic encoding (e.g., La Heij et al., 1998) or (b) that proposes unidirectional spreading of activation from words to syntax (Caramazza, 1997; Levelt, 1989; Levelt, Roelofs, & Meyer, 1999). Third, the difference in direction of effects attributable to head and modifier functions poses new questions for non-structural models of language processing, especially *naïve discriminative learning* (Baayen et al., 2011) and *uniform information density* (Jaeger, 2010). The former class of models has been quite successful in accounting for distributional effects on small scales (using relatively short-lag Markov chains), but has yet to produce *different* effects for contextual information within the same small-scale Markov distances for the *same task*. The latter class of models have so far relied heavily on surface-level sequences, and have likewise tended to predict a single direction of effect for informational load (that is, higher informational load translates into increased temporal delay). It is not yet clear how such theories will account for the fact that some syntactic information is facilitative while other syntactic information is inhibitory.

Proposed Research

The present study suggests several questions for future research. These suggestions fall into three broad categories: improving estimates of syntactic diversity, measuring these effects in different processing contexts, and applying these findings to other, typologically distinct languages.

First, I hope to scale up the measures introduced so far to encompass syntactic ramifications beyond the binary relationships captured by the current approach. As a first step, I propose a measure of the *contextualized* syntactic complexity of nouns. Rather than measuring the diversity of individual syntactic relationships, we can take the aggregate structural implications of a noun, even beyond its immediate phrasal scope. The measure involves first inducing a probabilistic context-free grammar (PCFG) from the set of trees in a given treebank that have at least one terminal symbol (*leaf*) corresponding to a target noun. The grammar could be 'trimmed,' such that the target is the only leaf, all other branches ending in pre-terminal (lexical category) nodes. Once we induce the grammar, we can easily compute its entropy. The entropy of the grammar thus calculated reflects broader correspondences between syntactic structures and the given noun. While this measure loses some of the granularity of the dependency-based measures introduced above, it may present a more realistic picture of the syntactic space, covering both headship and modifiership simultaneously. An extension of this measure might be to limit the PCFGs to the material occurring to the right or left of the target while also being subsumed under the maximal phrasal projection (highest NP of a series of uninterrupted NPs) of the noun.

A perhaps more pressing improvement of the existing measures would be to calculate the syntactic diversities over the lemmas, rather than the surface forms (i.e., counting dependencies with both singular and plural forms of the target nouns). I could then contrast the lemma diversities with the putatively morphological *inflectional entropy* (entropy of plural-singular distribution). Inflectional entropy has been known to have an inhibitory effect on production latencies (Bien, Baayen, & Levelt, 2011).

Second, I explore the effects of these measures in comprehension, word naming, and naturalistic production. For comprehension, I look at lexical decision data, simple (British Lexicon Project and English Lexicon Project) and primed (Semantic Priming Project). To explore the possibility of syntactic-distributional priming, I take the Jensen-Shannon divergence (JSD) between the target and prime in the database. JSD measures the point-wise distance between two distributions, and so should correlate positively with response times (more distant primes should facilitate access less than more similar primes). SPP also contains varied inter-stimulus intervals (ISI), meaning that we can approximate the time-course of any syntactic priming effect: does it only apply at short ISI, or does it persist over longer periods? Preliminary evidence suggests that the diversity measures are generally facilitative to comprehension, though I have yet to apply the directional measures. For simple word naming, I use the English Lexicon Project. For primed word naming, I use the SPP. Word naming is known to by-pass semantic processing (cf. Shibahara et al., 2003), helping us to rule out a fully semantic account of the observed effects. However, word reading is also assumed to tap into phonological encoding directly, and so to by-pass syntactic processing, as well. Therefore, I do not expect to find an effect of these measures on word naming. Finally, I look at naturalistic production. In particular, I predict that the order of coordinate nouns will be influenced by their relative syntactic diversities. Speaking to this point, I find that when speakers produce coordinate NPs of the form *the* N1 *and the* N2, they tend to place words with higher headship diversity in N1 position. Coordinate phrases of this kind are thought to be free from semantic constraints; they are also known to allow for parallel processing of the two nouns (Allum & Wheeldon, 2007). Thus, the race is won by nouns that the experimental data suggest to be more efficiently processed.

Third, the close relationship between hierarchy and word order revealed by these pilot studies suggests that other languages might show different effects. In particular, head-final languages like Japanese might tune processing to facilitate leftward integration. In that case, I would expect H_{lh} to dominate over H_{rh}. Predictions for modifiership diversity are less clear; for instance, they may stay the same (interfering with the selection of a rightward head). I also have picture naming data and treebanks for Dutch, Japanese, Italian, Spanish, German, Hungarian, Bulgarian, and Mandarin. While the romance languages have shown inconsistent results, German shows a clear facilitation for H_h similar to English.

The proposed research can be summarized in the following steps:

1. Compute existing diversity measures using lemmas rather than surface forms.
2. Compute new diversity measures using the one-leaf PCFG approach (OLPA).
3. Compute inflectional entropy for the targets.
4. Compute relative entropies (Jensen-Shannon Diversity) between primes and targets in the Semantic Priming Project database.
5. Collect control measures for visual lexical decision and word naming (*age of acquisition, orthographic neighborhood size, OLD20, character length, properties of the initial segment*) from norming databases.
6. Analyze the English Lexicon Project and British Lexicon Project visual lexical decision data.
7. Analyze the Semantic Priming Project primed visual lexical decision data.
8. Check correlation of the syntactic spaces with the semantic spaces (are words that are closely related in the Latent Semantic Analysis space also closely related in our syntactic spaces?).
9. Analyze ELP word-naming database.
10. Analyze SPP primed word-naming database.
11. Compute diversity measures for Dutch, Japanese, Italian, Spanish, German, Hungarian, Bulgarian, and Mandarin.
12. Analyze the cross-linguistic production data.
13. Analyze the cross-linguistic comprehension data.

I propose the following organization for the final dissertation (I omit intoductory and concluding chapters, as their content will be determined on the basis of analyses that have yet to be completed):

Chapter 1: Processing at the syntax-lexis interface
This chapter will survey the state-of-the-art in psycholinguistic modeling of lexical access with a special focus on the relationship between words and syntax. This review will also of necessity dip into the sentence processing literature. It will establish the importance of the question of how words are linked to abstract structures, and how the answers to these questions have implications for both processing and representation, hence for both psychologists and linguists.

Chapter 2: From morphology to syntax: Paradigmatic effects on lexical access
This chapter review findings from the morphological processing literature regarding the role of inflectional paradigms on processing. I then report an analysis (some of which has been published as Lester & Moscoso del Prado Martín, 2015) that extrapolates the inflectional approach to syntax. This analysis reveals that nouns with more diverse distributions across prepositions in the English prepositional phrase are recognized faster. New aspects of the analysis include a measure of the average distance between syntactic relata which demonstrates that increased distance diminishes – but does not overturn – the facilitatory effect. This analysis serves as a bridge from fully lexical to fully abstract-syntactic distributions.

Chapter 3: Diversity in syntactic paradigms I: A first approximation
This chapter will outline the original three diversity measures (total, headship, and modifiership diversity). It will describe the data, collection procedures, and operationalization of diversity. It will report an initial analysis of the BLP lexical decision data mirroring the analysis presented in Chapter 2. It will also report the pilot study on production using the IPNP data (some of which has been published as Lester & Moscoso del Prado Martín, 2016). Results will be compared across the two processing modalities..

Chapter 4: Diversity in syntactic paradigms II: The role of word order
This chapter introduces the directional extensions of the basic diversity measures. It will be structured identically to Chapter 3, with analyses of comprehension and production data. The data will be the same, but the new diversity measures will be substituted for the old. This chapter will also report two replication studies. First, I will replicate the visual lexical decision effect using the ELP data. Then, I report an experiment replicating the Bates et al. (2003) methodology. The pairs of experimental results will be compared.

Chapter 5: Representation of syntactic paradigms: A priming-based perspective
This chapter reports priming in visual lexical decision and word naming. It introduces a new measure, the Jensen-Shannon Divergence, which captures the distance between two distributions. I apply this measure to each of the distribution types and compare the performance of the resulting distances on priming. I specifically pit these distance measures against a semantic distance measure (cosine in the Latent Semantic Analysis space) to test whether they are indeed reducible to semantics. I further investigate the correlation between the syntactic and semantic spaces to demonstrate their independence.

Chapter 6: Syntactic diversity in the cross-lingusitic perspective
Here I explore the typological implications of the syntactic diversity measures. In particular, the disadvantages uncovered for head-final diversity in English (a predominantly head-initial language). I

compute the relevant measures for Dutch, Japanese, Italian, Spanish, German, Hungarian, Bulgarian, and Mandarin. I then apply those measures as predictors of the corresponding production latencies in the IPNP and (for selected languages) of recognition latencies.

References

Allum, P. H., & Wheeldon, L. R. (2007). Planning scope in spoken sentence production: The role of grammatical units. Journal of Experimental Psychology: Learning, Memory, & Cognition, 33, 791-810.

Baayen, R. H. (2008). Analyzing linguistic data: A practical introduction to statistics using R. Cambridge: Cambridge University Press.

Baayen, R. H. (2013). languageR: Data sets and functions with Analyzing Linguistic Data: A practical introduction to statistics. R package version 1.4.1. https://CRAN.R-project.org/package=languageR

Baayen, R. H., Milin, P., Filipović-Đurđević, D., Hendrix, P. & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. Psychological Review, 118, 438-482.

Baayen, R. H., Levelt, W. M. J, Schreuder, R. and Ernestus, M. (2008). Paradigmatic structure in speech production. Proceedings of the Chicago Linguistics Society, 43, 1-29.

Baayen, R. H., van Rij, J., de Cat, C. and Wood, S. N. (to appear). Autocorrelated errors in experimental data in the language sciences: Some solutions offered by Generalized Additive Mixed Models. In Speelman, D., Heylen, K. and Geeraerts, D. (eds), Mixed Effects Regression Models in Linguistics. Berlin, Springer.

Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., Herron, D., Lu, C. C., Pechmann, T., Pléh, C., Wicha, N., Federmeier, L. Gerdjikova, I., Gutierrez, G., Hung, D., Hsu, J., Iyer, G., Kohnert, K., Mehotcheva, T., Orozco-Figueroa, A., Tzeng, A., & Tzeng, O. (2003). Timed picture naming in seven languages. Pscyhonomic Bulletin and Review, 10, 344-380.

Bien, H., Baayen, R., and Levelt, W. (2011). Frequency effects in the production of Dutch deverbal adjectives and inflected verbs. Language and Cognitive Processes, 27, 683-715.

Bloomfield, L. (1933). Language. Chicago: University of Chicago Press.

Bonin, P., Peereman, R., Malardier, N., Méot, A., & Chalard, M. (2003). A new set of 299 pictures for psycholinguistic studies: French norms for name agreement, image agreement, conceptual familiarity, visual complexity, image variability, age of acquisition, and naming latencies. Behavior Research Methods, Instruments, & Computers, 35, 158-167.

Bresnan, J. (2001). Lexical-functional syntax. Oxford: Blackwell.

Bürki,A., Sadat, J., Dubarry, A-S., & Alario, F-X. (2016). Sequential processing during noun phrase production. Cognition, 146, 90-99.

Caramazza, A. (1997). How many levels of processing are there in lexical access? Cognitive Neuropsychology, 14, 177–208.

Chang, F. and Fitz, H. (2014) Computational models of sentence production: A Dual-path approach. In Ferreira, V. S., Goldrick, M., and Miozzo, M. (Eds.), The Oxford Handbook of Language Production (pp. 70-87), New York: Oxford University Press

Charniak, E., & Johnson, M. (2005). Coarse-to-fine n-best parsing and MaxEnt discriminative reranking. In Proceedings of the 43rd Annual Meeting of the ACL (pp. 173-180), Ann Arbor: University of Michigan.

Chao, A., Wang, Y. T., & Jost, L. (2013). Entropy and the species accumulation curve: a novel estimator of entropy via discovery rates of new species. Methods in Ecology and Evolution, 4, 1091–1110

Chomsky, N. (1957). Syntactic structures. The Hague: MIT Press.

Chomsky, N. (1970). Remarks on nominalization. In Jacobs, R. & Rosenbaum, P. (Eds.), Reading in English Transformational Grammar (pp. 184-221). Waltham: Ginn.

Chomsky, N. (1986). Knowledge of language. New York: Praeger.

Cubelli, R., Lotto, L., Paolieri, D., Girelli, M., & Job, R. (2005). Grammatical gender is selected in bare noun production: Evidence from the picture-word interference paradigm. Journal of Memory and Language, 53, 42–59.

de Marneffe, M-C., MacCartney, B., & Manning, C. D. (2006). Generating Typed Dependency Parses from Phrase Structure Parses. In LREC 2006.

de Simone, F., & Collina, S. (2015). The picture-word interference paradigm: Grammatical class effects in lexical production. Journal of Psycholinguistic Research, DOI: 10.1007/s10936-015- 9388-9.

Dell, G. S., Oppenheim, G. M., & Kittredge, A. K. (2008). Saying the right word at the right time: Syntagmatic and paradigmatic interference in sentence production. Language and Cognitive Processes, 23, 583-608.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., García Marquéz, J. R., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2012). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. Ecography, 35, 1-20

Duràn, C. P. & Pillon, A. (2011). The role of grammatical category information in spoken word retrieval. Frontiers in Psychology, 2, 1-20.

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

Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. Journal of Memory and Language, 46, 57–84.

Ferreira, V. S. (1996). Is it better to give than to donate? Syntactic flexibility in language production. Journal of Memory and Language, 35, 724–755.

Goldberg, A. E. (1995). Constructions: A Construction Grammar approach to argument structure constructions. Chicago: University of Chicago Press.

Gregory, E., Varley, R., Herbert, R. (2012). Determiner primes as facilitators of lexical retrieval. Journal of Psycholinguistic Research, 41, 439-453.

Hale, K. (1982). Preliminary Remarks on Configurationality. Proceedings of the North Eastern Linguistic Society, 12, 86-96.

Hendrix, P., Bolger, P. and Baayen, R. H. (2016). Distinct ERP signatures of word frequency, phrase frequency, and prototypicality in speech production. Journal of Experimental Psychology LMC.

Hudson, R. (2007). Language networks: The new Word Grammar. Oxford: Oxford University Press.

Hwang, H. & Kaiser, E. (2014). Having a syntactic choice is not always better: The effects of syntactic flexibility on Korean production. Language, Cognition, & Neuroscience, 29, 1115-1131.

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. Cognitive Psychology, 61. 23-62.

Langacker, R.W. (1987). Foundations of Cognitive Grammar, Vol I: Theoretical prerequisites. Stanford: CSLI.

La Heij, W., Mark, P., Sander, J., & Willeboorsde, E. (1998). The gender congruency effect in picture word task. Psychological Research, 61, 209–219.

Lester, N. A., & Moscoso del Prado Martín, F. (2015). Constructional paradigms affect visual lexical decision latencies in English. In Noelle, D. C., Dale, R., Warlaumont, A. S., Yoshimi, J., Matlock, T., Jennings, C. D., & Maglio, P. P. (Eds.), Proceedings of the 37th Annual Meeting of the Cognitive Science Society (pp. 1320-1325). Austin, TX: Cognitive Science Society.

Levelt, W. J. M. (1989). Speaking: From intention to articulation. The Hague: MIT Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. Behavioral and Brain Sciences, 22, 1–75.

Liu, H. (2010). Dependency direction as a means of word-order typology: A method based on dependency treebanks. *Lingua*, 120, 1567-1578

Marchini, J. L., Heaton, C., & Ripley, B. D. (2013). fastICA: FastICA algorithms to perform ICA and projection pursuit. R package version 1.2-0. https://CRAN.R-project.org/package=fastICA

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. Behavior Research Methods, 44(2), 314-324.

Mel'čuk, I. (1988). Dependency syntax: Theory and practice. Albany: The SUNY Press.

Miller, G. A. (1995). WordNet: A Lexical Database for English. Communications of the ACM, 38, 39-41.

Miozzo, M., Costa, A., & Caramazza, A. (2002). The absence of a gender congruency effect in romance languages: A matter of stimulus onset asynchrony? Journal of Experimental Psychology: Learning, Memory, & Cognition, 28, 388-391.

Moscoso del Prado Martín, F. (2014). Grammatical change begins within the word: Causal modeling of the co-evolution of Icelandic morphology and syntax. In P. Bello, M. Guarini, M. McShane, & B. Scasselatti (Eds.), Proceedings of the 36th Annual Conference of the Cognitive Science Society (pp. 2657–2662). Austin, TX: Cognitive Science Society.

Moscoso del Prado Martín, F. (2016). Vocabulary, grammar, sex, and aging. Cognitive Science, 1-26.

Moscoso del Prado Martín, F., Kostić, A., Baayen, R. H. (2004). Putting the bits together: An information theoretical perspective on morphological processing. Cognition, 94, 1-18.

Myachykov, A., Scheepers, C., Garrod, S., Thompson, D., & Fedorova, O. (2013). Syntactic flexibility and competition in sentence production: The case of English and Russian. The Quarterly Journal of Experimental Psychology, 66, 1601-1619.

Nivre, J. 2005. Dependency grammar and dependency parsing. Technical Report MSI report 05133, Växjö University: School of Mathematics and Systems Engineering.

Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: A model of cumulative semantic interference in speech production. Cognition, 114, 227-252.

Reppen, R., Ide, N., & Suderman, K. (2005). American National Corpus (ANC) Second Release LDC2005T35. DVD. Philadelphia: Linguistic Data Consortium, 2005.

Schiller, N.O., & Caramazza, A. (2003). Grammatical feature selection in noun phrase production: Evidence from German and Dutch. Journal of Memory & Language, 48, 169-194.

Schriefers, H. (1993). Syntactic processes in the production of noun phrases. Journal of Experimental Psychology: Learning, Memory, and Cognition, 19, 841-850.

Schriefers, H. & Jescheniak, J. D. (1999). Representation and processing of grammatical gender in language production: A review. Journal of Psycholinguistic Research, 28, 575-600.

Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. Journal of Memory and Language, 29, 86-102.

Shannon, C. E. (1948). A mathematical theory of communication. The Bell System Technical Journal, 28, 379-423.

Shibahara, N., Zorzi, M., Hill, M. P., Wydell, T., & Butterworth, B. (2003). Semantic effects in word naming: Evidence from English and Japanese kanji. *The Quarterly Journal of Experimental Psychology, 56A*, 263-286.

Simon, H. A. (1956). Rational Choice and the Structure of the Environment. Psychological Review, 63, 129–138.

Stoll, S., Bickel, B., Lieven, E., Paudyal, N. P., Banjade, G., Bhatta, T. N., & Rai, N. K. (2012). Nouns and verbs in Chintang: Children's usage and surrounding adult speech. Journal of Child Language,39, 284–321.

Tesnière, L. (1959). Éléments de syntaxe structurale. Paris: Klincksieck.

Wasow, T., Jaeger, T. F., & Orr, D. M. (2011). Lexical variation in relativizer frequency. In H. J. Simon & H. Wiese (Eds.), Expecting the unexpected: Exceptions in grammar (175-196). New York: de Gruyter.

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., & Smith, G., M. (2009). Mixed effects models and extensions in ecology with R. New York: Springer.