ABSTRACT

Title of Dissertation: A WORLD WITHOUT WORDS:
A NON-LEXICALIST FRAMEWORK FOR
PSYCHO- AND NEURO-LINGUSITICS

Alexandra Krauska
Doctor of Philosophy, 2024

Dissertation Directed by: Associate Professor Ellen Lau
Department of Linguistics

In standard models of language production or comprehension, the elements which are retrieved from memory and combined into a syntactic structure are "lemmas" or "lexical items". Such models implicitly take a "lexicalist" approach, which assumes that lexical items store meaning, syntax, and form together, that syntactic and lexical processes are distinct, and that syntactic structure does not extend below the word level. Across the last several decades, linguistic research examining a typologically diverse set of languages has provided strong evidence against this approach. These findings suggest that syntactic processes apply both above and below the "word" level, and that both meaning and form are partially determined by the syntactic context. This has significant implications for psychological and neurological models of language processing as well as for the way that we understand different types of aphasia and other language disorders. As a consequence of the lexicalist assumptions of these models, many kinds of sentences that speakers produce and comprehend - in a variety of languages, including English - are challenging for

them to account for. In order to move away from lexicalism in psycho- and neuro-linguistics, it is not enough to simply update the syntactic representations of words or phrases; the processing algorithms involved in language production are constrained by the lexicalist representations that they operate on, and thus also need to be reimagined.

This dissertation discusses the issues with lexicalism in linguistic theory as well as its implications in psycho- and neuro-linguistics. In addition, I propose a non-lexicalist model of language production, the "WithOut Words" (WOW) model, which does not rely on lemma representations, but instead represents that knowledge as independent mappings between meaning and syntax, and syntax and form, with a single integrated stage for the retrieval and assembly of syntactic structure. Based on this, the model suggests that neural responses during language production should be modulated not just by the pieces of meaning, syntax, and form, but also by the complexity of the mapping processes which link those separate representations. This prediction is supported by the results of a novel experimental paradigm using electroencephalography (EEG) during language production, which observes greater neural responses for meaning-syntax and syntax-form mapping complexity in two separate time windows. Finally, I re-evaluate the dissociation between regular and irregular verbs in aphasia, which has been used as supporting evidence for a distinction between the grammar and the lexicon. By training recurrent neural networks and measuring their performance after lesioning, I show that the observed clinical data can be accounted for within a single mechanism. By moving away from lexicalist assumptions, the non-lexicalist framework described in this dissertation provides better cross-linguistic coverage and aligns better with contemporary syntactic theory.

A WORLD WITHOUT WORDS: A NON-LEXICALIST FRAMEWORK FOR PSYCHO- AND NEURO-LINGUISTICS

by

Alexandra Krauska

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2024

Advisory Committee:
Associate Professor Ellen Lau, Chair
Associate Professor Naomi Feldman
Assistant Professor Kate Mooney
Professor David Embick
Associate Professor Robert Slevc

Acknowledgments

I started my Ph.D. in the fall of 2019. As we all know, the five years since then have been pretty turbulent (to say the least), personally, professionally, politically, and pandemic-ly. I have so many people that I need to thank here, where I can say with absolutely no exaggeration that I would not have made it to this point without them.

First, I'd like to thank all of the people who have been my advisors (officially or unofficially) in the past five years. To Ellen, thank you for listening to all of my half-formed ideas and helping me transform them into something beautiful and impactful. To Naomi, thank you for challenging me to think outside the box and look for ways to build bridges. To Dave, thank you for always answering my random questions and helping me come up with great experiment ideas. And to Omer, thank you for the "burn it all down" energy that provided the spark for this dissertation.

To my other committee members, Kate and Bob, thank you for your insightful comments and helpful feedback. To the other faculty at UMD, particularly Alexander, Bill, Jeff, Colin, and Tonia, thank you for all of the interesting discussions over the years, and thank you for all of your tough questions!

To Allison and Jenna, thank for making collecting the EEG data much more enjoyable. There's no one I'd rather gel a cap with! And to Dana, thank you for your work putting together the German and Dutch datasets for Chapter 6, and for helping me to think about the project in

new ways.

To my cohort, office mates, friends, and former roommates (many of whom fall into multiple of these categories), thank you for keeping me sane and making the past five years much more bearable. To my academic friends and collaborators, especially Masaya, for getting me into this in the first place, and to William, Jeremy, Nayoun, Jiayi, Sladjana, and others, thank you for your enthusiastic support over the years.

To anyone who has ever given me coffee, particularly Shevaun and Caitlin at the Language Science Center, thank you for keeping me going.

To Drs. Crosby, Witten, and Navarrete, the iWalk 3.0, and my whole care team at Aurora Healthcare in the summer of 2021, thank you for taking apart my foot and putting it back together again.

To my family and friends, thank you for your unconditional love and support, for always believing in me even when I wasn't sure I believed in myself, and for reminding me that work-life balance is important and necessary.

Finally, to Sarah, the love of my life, the list of things I have to thank you for would be as long as this dissertation. Thank you for moving across the country for me (soon to be multiple times), thank you for listening to me talk about my research so much that you memorized the elevator pitch, thank you for being my #1 fan, and thank you for putting up with me for the past 5+ years. None of this would have been possible without you.

Table of Contents

List of Figures

Chapter 1

Introduction

For many years, people have been pondering the puzzle of how language is produced and comprehended; how do we get from a conceptual representation of what we want to say, to a series of articulatory gestures that make up speech or sign? When we perceive a series of such articulatory gestures, how do we interpret that signal to get the intended meaning? As an accident of history, many of the original researchers interested in this problem in the modern psycholinguistics era (beginning in the 1960s and 70s) spoke European languages, particularly English and Dutch. For these researchers, the problem of language production should involve a few intermediary steps: once a concept has been generated, how do we retrieve the corresponding words from memory? After that, how do we build a syntactic structure from those words and put them into the correct linear order? In creating models to answer these questions, those researchers were often making an unnoticed commitment about how language works, centered on a notion of wordhood informed by their own language experience. Dominant theories of syntax at the time - also largely developed based on European languages - assumed that words were the units of combination, and that everything happening below the word level belonged to a separate domain, morphology. In this kind of theory, the word acts as a bridge between meaning, syntax, and form. The psycholinguistic and neurolinguistic models that were built incorporated this understanding of syntax and wordhood into both the representations and algorithms of those models. This is the lexicalist approach.

In recent decades, much linguistic work, relying on a broader set of cross-linguistic data, has argued against lexicalist theories of language. Chapter 2 of this dissertation defines some of the key properties of lexicalist theories, summarizes several of the key arguments used against them, and describes the kind of linguistic phenomena that they struggle to account for because of their lexicalist assumptions.

Despite the many theoretical arguments against lexicalism, these assumptions are pervasive in psycho- and neuro-linguistics, and have continued to influence the way we think about language comprehension and production. Chapter 3 describes the key ways that lexicalist assumptions are manifested in models of language processing, and the ways that those assumptions are problematic in accounting for the production of various linguistic phenomena.

By eliminating "words", "lemmas", and "lexical items" from our models of language processing, a very significant question remains: what is the alternative? Given that the retrieval and combination of lemmas has typically served to drive language production processes, what kinds of processes are necessary in their absence? This is not just a change in terminology, but a major shift in both the kinds of representations and algorithms that appear during language production. Chapter 4 proposes the WithOut Words (WOW) model of language production, which emphasizes separate representations of meaning, syntax, and form which are linked to each other through independent mapping processes (between meaning and syntax, and between syntax and form). In this model, there are no "lexical" representations, and no area of the brain devoted to "lexical" processing. Instead, the neural architectures are argued to be performing transformations between complex data structures. This model also emphasizes the role of prosody in linearization and producing fluent speech.

The WOW model generates several diverging predictions from previous models of language processing, one of which is that the complexity of the mapping processes between different representations should incur processing costs, rather than just involving costs associated with retrieving and combining specific representations. Chapter 5 describes an experiment which tests this prediction, using EEG to measure neural responses during the production of phrases with different levels of complexity in the mapping between meaning and syntax (Experiment 1), and in the mapping between syntax and form (Experiment 2). The experiment provides preliminary evidence that mapping complexity does modulate neural responses independent of syntactic complexity, providing a proof of concept for one of the key predictions made by the WOW model.

The WOW model also generates some predictions about the properties of language disorders such as aphasia. The aphasia literature often assumes a grammar-lexicon distinction, an extension of the lexicalist approach, with double dissociations often providing key evidence to support such a view. For example, it has been argued that people with "agrammatic" aphasia exhibit a deficit for the regular past tense, which is performed via a grammatical rule, whereas people with "fluent" aphasia exhibit a deficit for the irregular past tense, which is retrieved from the lexicon. Chapter 6 describes a set of recurrent neural networks that perform the past tense inflection in English, German, and Dutch, simulating the pattern of deficits observed for people with agrammatic aphasia after lesioning, where verb frequency plays a role in how well those forms are preserved; this evidence supports a non-lexicalist account, where both types of verbs are inflected by a single mechanism that – when damaged – can disproportionately impact one set of verbs over another, without requiring them to be distinguished in the grammar. This model also provides a starting point for a connectionist implementation of the WOW model, which performs transformations between complex data structures.

The non-lexicalist framework described in this dissertation provides an important perspective for bridging the gaps between multiple fields, drawing insights from theoretical linguistics, psycho- and neuro-linguistics, and computational approaches. By moving away from lexicalist assumptions, this model aligns better with contemporary syntactic theory and provides better cross-linguistic coverage. This dissertation is just a starting point for creating a "world without words" as the focus of psycho- or neuro-linguistic inquiry.

Chapter 2

Issues with lexicalism in linguistic theory

Lexicalism has been around for a long time in linguistics, and many of the foundational theories of syntax analyzed words as the minimal units in syntactic computations. Though the Lexicalist Hypothesis was first introduced in Chomsky's *Remarks on Nominalization* (Chomsky, 1970), lexicalism is not one single cohesive theory, but rather an approach taken by a variety of linguistic theories which rely on some or all of the following assumptions: firstly, that sub-word and supra-word processes are different in kind, such that the rules that apply in the syntax cannot apply within lexical items, and conversely, that the processes or relations that hold among derivationally related lexical items are different in kind from those that hold among derivationally related syntactic structures; secondly, that wordhood is the domain of listedness; and thirdly, that lexical items necessarily include triads of sound, meaning, and syntax. However, there has been a great deal of development, especially relying on cross-linguistic data, which has challenged these assumptions. This section will describe these claims in detail, summarize several of the key theories which rely on them and how they were used, and provide evidence which calls each one into question.

2.1 Syntactic and morphological computations are different in kind

Under this assumption, morphology (or other sub-word operations) and syntax (or other supra-word operations) are fundamentally different in kind; each has their own sets of atoms and rules of formation, and neither can be used to account for the same phenomena (Di Sciullo and Williams, 1987). This notion of "atomicity" is closely related to the lexical integrity hypothesis, the strong lexicalist hypothesis (Lapointe, 1980), or the lexicalist hypothesis (Chomsky, 1970; Williams, 1981). These hypotheses hold that, though words do have features, those features do not correspond to structure, and the features that are relevant to word-internal or sub-word operations cannot be accessed by supra-word operations. According to Di Sciullo and Williams (1987), morphology and syntax are as different as "history and forestry," i.e., the two are sciences concerned with different objects.

Under a lexicalist approach, morphological rules and syntactic rules operate over different units. Syntactic rules operate over phrases and categories, such as in examples (1a-b). Morphological rules, though superficially similar, operate over roots, stems, and affixes, such as in examples (1c-d).

(1) a. S $\rightarrow$ NP Infl VP

b. NP $\rightarrow$ (Det) (Adj) N (PP)

c. stem $\rightarrow$ stem affix

d. word $\rightarrow$ affix word

Given these syntactic and morphological rules, atomicity is a natural consequence; the syntactic rules cannot analyze morphological objects because they do not have the necessary means to do so - they can only make reference to a restricted set of grammatical components (Di Sciullo and Williams, 1987). Structure below the word level is not accessible to syntactic operations, and morphological rules cannot operate on syntactic configurations, because those two sets of rules do not apply to both sets of units. The two sets of rules should not be used to account for the same phenomena simply because they cannot both operate over the units that compose those phenomena.

Some interaction needs to exist between syntax and morphology, namely in verbal inflection, though lexicalist theories argue that the interaction is limited and functions in such a way that the two sets of rules and operations are not intermixed, and that only certain components can become part of the shared "vocabulary". For example, grammatical category can be encoded as a feature or set of features (such as +N, +V, etc), along with argument structure and selectional restrictions. Feature percolation, a mechanism which allows those lexical features to "percolate" up in order to become accessible to the syntax, creates some indirect interaction between affixes and the syntactic environment, but no more. For example, in order for verb agreement to occur in a sentence like "students enjoy linguistics", the +3 and +PL features of the subject "students" percolate up in order to be accessible to the syntax, which are then passed to the verb in order to get the correct +3PL verb inflection on "enjoy". In this way, the word level is a "bottleneck" of information between morphology and syntax.

One piece of evidence that has been used to argue that morphology and syntax are definitionally different in kind is that the lexicon is less productive than the syntax; the lexicon may still be productive, but that productivity is more restricted than what is seen in the syntax. This belief is held to varying degrees, with Fabb (1984) suggesting that productivity is a definitional property of syntax and not morphology, while Di Sciullo and Williams (1987) argue that productivity is not a criterial difference, as both productive and unproductive phrase types and word types exist. They concede that some morphemes may be productive, but only for a particular class of words (such as *-ion* for latinate words; *instruct* ∼ *instruction; calculate* ∼ *calculation*).

Finally, another piece of evidence that lexicalist theories use to support this claim is in the domain of meaning. According to Di Sciullo and Williams (1987), words are "generic" in meaning in a way that phrases are not. A word like "robber" denotes a permanent property of being a thief; in a sentence like (2b), it cannot mean that "John is robbing a bank at this very moment", but rather that he has the permanent characteristic of being a person who steals from banks. In contrast, in a phrase such as (2c), it becomes temporary (time-bound) property. Di Sciullo and Williams (1987) conclude that time references must come from INFL, not NP, and thus time references must be accessible to syntactic rules in order to be interpreted. As a result, words, being syntactically atomic, cannot be time-bound until they enter into an event.

(2) a. Robber — *generic, denotes a permanent property*

b. John is a bank robber — *not time-bound, denotes a permanent property*

c. The man who is robbing a bank — *time-bound property*

To summarize, lexicalist approaches argue for a distinction between the morphology and the syntax because syntactic rules and morphological rules operate over different units, there is limited interactivity between syntax and morphology, syntax is fully productive while morphology exhibits only limited productivity, and because words are generic in ways that phrases are not.

Some of this evidence is just a product of the limited theoretical machinery that was available at the time; given the advancements in morphosyntactic theory, it is unnecessary to posit a constraint such as syntactic atomicity based on incompatible sets of rules, because frameworks have been developed which allow syntactic rules to operate over sub-word units. Meanwhile, some of the other claims have turned out to simply be empirically false, in light of new evidence which has arisen in the past 40 years.

To begin with, investigations into the structure of polysynthetic languages have provided many insights into the properties of language and the nature of wordhood. Classic work on polysynthetic languages by Baker (1985) indicated that the relationship between syntactic structure and morphological structure is much more systematic than would be predicted by lexicalist approaches. Baker observed that the ordering of affixes within words in polysynthetic languages mirrors the order of the isolated elements that realize the same semantic categories in non-polysynthetic languages, and that the order of the morphemes within the word can also have an effect on grammatical functions and semantic interpretation. In Quechua, for example, the order of the morphemes within a word can have a significant impact on its meaning, even when they contain a very similar set of morphemes, as shown in example (3) below:

(3) a. Maqa **-naku** -ya **-chi** -n.
beat -recip -dur -caus -3S
'He$_j$ is causing them$_i$ to beat each other$_i$.'

b. Maqa **-chi** **-naku** -rka -n.
beat -caus -recip -pl -3S
'They$_i$ let someone$_j$ beat each other$_i$.'

In (3a), the semantic subject of the verb root "beat" and the direct object are in a reciprocal relationship (*they are beating each other*). In (3b), in contrast, the causer is interpreted to be the direct object of the beating (*they are causing the beating of themselves*). The key difference between these two sentences is the relative order of the reciprocal morpheme and the causative morpheme. (3a) shows the reciprocal morpheme closest to the verb, which binds the object of the verb to the underlying subject ("beat each other"), eventually followed by the causative morpheme ("cause to [beat each other]"); meanwhile, in (3b) the causative morpheme is closest to the verb, which binds the object of the verb to the causer ("cause to beat"), followed by the reciprocal morpheme ("[cause to beat] each other").

Based on this observation, Baker concludes that the processes involved in generating these effects - theta roles, the reciprocal, causative, passive, and agreement, which are assumed to be assigned structurally - must simultaneously have effects on the morphological structure, as well as the syntactic structure and grammatical functions, and semantic interpretation. If the morphology and the syntax are two separate domains, there is no reason to expect any particular relationship or pattern to emerge between the two sets of operations. The Mirror Principle shows instead that there is an extremely systematic relationship between them, where the order of affixes within the word exactly mirrors that of their hierarchical relation in the syntax. This could only emerge if the rules of word formation were subject to the same principles as syntactic structure building, and if rules of syntax and morphology can operate over the same fundamental units. This provides powerful evidence that the lexical integrity hypothesis - the claim that lexical items have features but not structure - is false. Furthermore, the highly productive morphemes involved in the generation of novel words in these languages directly challenges the idea that productivity should be a definitional property of syntax and not morphology. If this were the case, lexicalist approaches would have no way to account for the fact that a speaker of a polysynthetic language is just as likely to produce a word that has never been used before as a speaker of English is to produce a sentence that has never been used before. This would not be possible if morphology cannot be used productively in the same way as phrasal syntax.

Additional evidence that there needs to be a significant amount of interaction between syntax and morphology comes from the causative morpheme in Japanese. As shown in (4a), the causative morpheme *-sase* attaches to the verb, though its form may be conditioned on the verb that it attaches to. In this example, *Taroo* is marked with nominative case and is interpreted to be the causer, while *Hanako*, the logical subject of the root verb, is marked with accusative or dative case and is interpreted to be the causee. Harley (2008) points out that because the V+CAUS combination constitutes a single phonological word, which - according to a lexicalist approach - would be considered to be a single terminal node, then it should head a single syntactic verb phrase, and thus should behave like a monoclausal[1] construction, such that *-sase* should only scope over the verb that is attached to. It does behave like a monoclausal verb with respect to case, tense, and NPI licensing. However, the productive V+CAUS combinations exhibit a number of properties that instead point towards a biclausal structure, especially related to binding, scope, control, and disjunction. As shown in (4b) below, the causative morpheme can appear on both verbs in a disjunction ("cause to [clean the house]" and "cause to [pay rent]"), and the interpretation is that there is a possibility of either "causing" event ("Hanako either caused Masao to clean the house, or she caused him to pay rent"). (4c), in contrast, illustrates that there is the availability of disjunction of two VPs under a single causative morpheme, where the causative scopes over the disjunction ("Hanako caused Masao to either clean the house or pay rent"). This example is an issue for lexicalist approaches, because disjunction is treated as a syntactic operation, not a morphological one.

[1] *Monoclausal* here means that there is only a single clause, as in the English sentence "Zachary loves dogs". *Biclausal*, in contrast, means that there are multiple clauses, as in the English sentence "It is surprising that Zachary loves dogs".

(4) a. Taroo-ga Hanako-o ik-ase-ta
Taro-N Hanako-A go-CAUS-PST

'Taro made Hanako go'

b. Hanako-ga [[Masao-ni uti-o soozis-aseru]-ka [heya-dai-o haraw-aseru]] koto ni sita
Hanako-N Masao-D house-A clean-CAUS-OR room-rent-A pay-CAUS that D do

'Hanako decided to make Masao clean the house or she decided to make him pay room rent'

Reading: OR scopes over CAUS; Masao won't have a choice

c. Hanako-ga [[Masao-ni uti-o soozisuru]-ka [heya-dai-o haraw]]-aseru koto ni sita
Hanako-N Masao-D house-A clean-OR room-rent-A pay-CAUS that D do

'Hanako decided to make Masao clean the house or pay room rent'

Reading: CAUS scopes over OR; Masao has a choice

If syntax can only operate over "lexical" units which are formed through non-syntactic processes, and the kind of disjunction observed above is necessarily a syntactic operation, the example in (4) provides an ordering paradox; word formation, specifically the *-sase* suffixation outside the disjunction, would need to be fed by the syntax which can only operate on the output of word formation processes. Alternatively, lexicalist approaches would have to assume that binding relations, scope, and adverbial control are morphological or lexical operations, rather than syntactic ones. Given the other syntactic properties of these operations, this conclusion is undesirable. The multiclausal properties of these causative constructions could only arise if syntactic operations and morphological operations are able to operate over the same kinds of units, and not distinct sets of rules.

Finally, the argument that words are generic in ways that sentences are not seems to be an arbitrary statement that exhibits many exceptions. In example (5), the phrase "John is a bank robber" seems to indicate a temporary property, in contrast to example (2b) above where it seemed to denote a permanent or generic property. To add further complication to this argument, adjectives such as "former" require some reference to time; it is difficult to derive the meaning of "former president" without involving some temporal properties. In addition, some languages - such as Kwakiutl, Guaraní, Potawatomi, and Sirionó, among others - exhibit relatively robust temporal inflection on nominals, as discussed in Nordlinger and Sadler (2004).

(5) Today, John is a bank robber and Bill is the getaway man *time-bound*

Furthermore, even if the generalization were correct, Di Sciullo and Williams (1987)'s observations may be captured via other principles that do not rely on a distinction between words and phrases. For example, an alternate hypothesis could be that some entities or relations require a syntactic structure in order to be semantically interpretable ("robber" requires the syntactic context of "John is a bank robber" in order to get a time-bound interpretation), whereas others do not ("former president" does not require a verb tense in order to be interpreted based on its temporal properties). The syntactic context can also provide alternate interpretations, as suggested by example (5). Given this, these observations of genericity may not be considered conclusive evidence that syntax and morphology need to be different in kind.

If we take seriously the evidence discussed above, we are led to the conclusion that a theory of syntax should not assume that there is a split between morphology (or other sub-word operations) and syntax (or other supra-word operations); there must be one syntax which operates over all morphosyntactic objects above and below the "word" level (whatever that may correspond to cognitively). The structure of words in polysynthetic languages provides clear evidence that there needs to be a great deal of interactivity between operations above and below the word level, and that productivity cannot be an exclusive property of supra-word operations. Japanese causatives provide additional evidence that morphology and syntax need to operate over the same kinds of units in order to generate the correct scope interpretations exhibited in (4). Finally, the generalization that genericity is a unique property of words is subject to a number of exceptions, so it should also not be considered to be a definitional distinction between words and phrases.

2.2 Wordhood and Listedness go hand-in-hand

The idea that wordhood is the domain of listedness (used here to mean listed in a "dictionary", the lexicon, or long-term memory) is another implicit assumption in many lexicalist theories, taken to varying degrees. That is, the idea that the set of words in a language is not only finite but also listed; from these finite means, an infinite number of sentences can be generated. Producing a novel sentence, therefore, is unremarkable. Producing a novel word, however, adds a new entry to the "dictionary" of listed things.

There is some variability in the definition and constraints on listedness across lexicalist theories, but in general, the lexicalist claim is that the things which are listed in the lexicon are words, but not things smaller or larger than a word. Di Sciullo and Williams (1987) argue that listed things include all things that must be memorized in the language, such as words and idioms and their meanings, though wordhood may not perfectly align with listedness. According to Jackendoff (1975), when a person produces a novel word, they are adding that word to their lexicon as a listed unit, even if that word can be composed via regular rules. In this way, there is grammatical significance to the use of a "new" word, but not a new sentence. Aronoff (1976), meanwhile, argues that the set of listed things includes "actual" words, while "potential" words are generated from word-formation rules. Meanwhile, there is no parallel notion of "actual" and "potential" sentences.

Heim and Kratzer (1998) also codify the listedness assumption in their theory of formal semantics, based particularly on the lexicalist assumption that syntax does not extend below the word level, and therefore any item in a terminal node must correspond to a word, which then must be listed in the lexicon, as shown in (6)[2]:

(6) Terminal Nodes (TN)

if α is a terminal node, $[\![\alpha]\!]$ is listed in the lexicon.

The distinction between listed and unlisted things, and the argument that wordhood corresponds to listedness, has several consequences in these lexicalist theories. For Aronoff, this is a key component of his account of why some words like "gloriosity" cannot exist, while "gloriousness" does - "gloriousness" is a listed word in the lexicon, so it blocks "gloriosity" from occurring, even though it could exist given the word-formation rules of English. Meanwhile, because the meaning of "terrific" is listed as "excellent", alternate meanings are blocked, even though it could mean something like "having the property of being terrifying" given the word formation rules of English (as with *horror* $\sim$ *horrify* $\sim$ *horrific*, *terror* $\sim$ *terrify* $\sim$ ***terrific***). In contrast, blocking does not occur between phrases. This account of blocking relies on a listed-unlisted distinction - where words are the listed things - in order to be coherent.

However, these blocking effects do not seem to be restricted to the word level, both because

[2]An additional consequence of this rule is that syntactic terminal nodes must have their own semantic interpretation in order to be listed in the lexicon. As will be discussed in section 2.3 (specifically in reference to examples 12 and 13), this may not be the case. Even if one were to argue that the terminal nodes can be smaller than the word level (as the non-lexicalist approach would suggest), the relevant semantic interpretation may need to be over larger pieces of syntax that contain several terminal nodes.

words can sometimes fail to show blocking effects (as in *curiousness* and *curiosity*, where both forms are permitted), and because phrases can sometimes block words, and vice versa (*taller* seems to block *more tall*, while *more colorful* seems to block *colorfuller*). Blocking effects should not be taken to be conclusive evidence that listedness is definitional only of words and not phrases.

Several pieces of evidence clearly indicate that the domain of listedness should not necessarily correspond to the word level, that the word level must be both greater and smaller than what needs to be listed in a language. Firstly, as discussed by Di Sciullo and Williams (1987) idioms can have complex syntactic structures with non-compositional meanings. The idioms shown in (7) are not just "strings of words" - they can undergo syntactic transformations (7a-b), incorporate non-idiomatic modifiers (7c), and involve open syntactic positions (7d-e). These suggest that the domain of listedness must be larger than the word level, and should include the phrase's syntactic structure.

(7) a. Let the cat out of the bag

b. If the cat was let out of the bag, then the trick would be exposed *Passivization*

c. The cat was well and truly out of the bag *non-idiomatic intervener*

d. Keep one's cool *Open position receives genitive case*

e. The senator kept his cool under the intense line of questioning

Conversely, highly productive morphemes with consistent meaning, such as *re-* in English, suggest that the domain of listedness must be smaller than the word level. The examples shown in (8) illustrate the consistent and productive meaning of *re-*.

(8) a. paint ~ repaint *to paint again*

b. send ∼ resend *to send again*

c. follow ∼ refollow *on social media, to follow someone again after unfollowing*

Again, another indication that the domain of listedness must be smaller than the word level comes from polysynthetic languages where many productive morphemes are frequently combined to create novel words, as discussed in example (3). Arguing that what is listed in the lexicon of a polysynthetic language are the composite words (or that they become listed the first time they are used) seems untenable.

To conclude this section, for some languages, monomorphemic words may be *at least* the things which are listed, but there must be other listed things as well. At this point, it is important to clarify a distinction between listedness and storage. The things which are listed in the lexicon should be all of the idiosyncratic things: any linguistic object whose properties cannot be comprehensively and predictively calculated from properties of its subparts. Storage, meanwhile, is a process of holding onto common phrases such as “kick the ball”, “walk the dog”, or “I don’t know” in long-term memory. The probability that something would be stored depends on its frequency of use, its utility (both in a single context and across a variety of contexts), and how often it is rehearsed. Storage and listedness should not be conflated; listedness is a property of the competence grammar, while storage is the product of a language system that becomes optimized for use. This ties in with language-specific optimization processes that will be discussed in Chapter 3.

2.3 Lexical items necessarily include triads of sound, meaning, and syntax

According to many lexicalist theories, lexical items include semantic, syntactic, and phonological information, tied together in some unified way (Di Sciullo and Williams, 1987; Aronoff, 1976; Jackendoff, 1975). This may be another implicit assumption that was taken for granted in early syntactic theories, though not supported empirically, but it has persisted throughout the decades in some form or another. The triad can be represented as a triplet set of meaning, syntax, and form, such as in Figure 2.1 below, representing the triad for the English word "kick". The meaning representation in some models is abstract, or it may involve conceptual features of some kind (the object of "kick" needs to be something concrete and kick-able; the object of a verb like "eat" needs to be eat-able). The syntactic representation may just include syntactic category in some theories, but it may include other features or larger segments of hierarchical syntax in other theories. The phonological form is often represented as a string of phonemes. Under this view, for every piece of syntax that has a stored meaning, it also must have a stored phonological form, and conversely, for every piece of syntax that has a stored phonological form, it must also have a stored meaning.

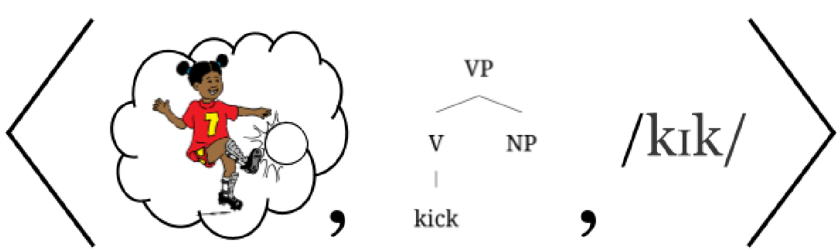

Figure 2.1: A visual representation of the triad for "kick"

If we were to assume that the lexicon is organized in these triads of meaning, syntax, and form, we would expect that every syntactic unit should have a corresponding form and meaning. It would be possible to have homonyms and synonyms, but each would constitute an independent triad, as illustrated for "bank" in Figure 2.2. In this kind of model, these simple ambiguities are trivial.

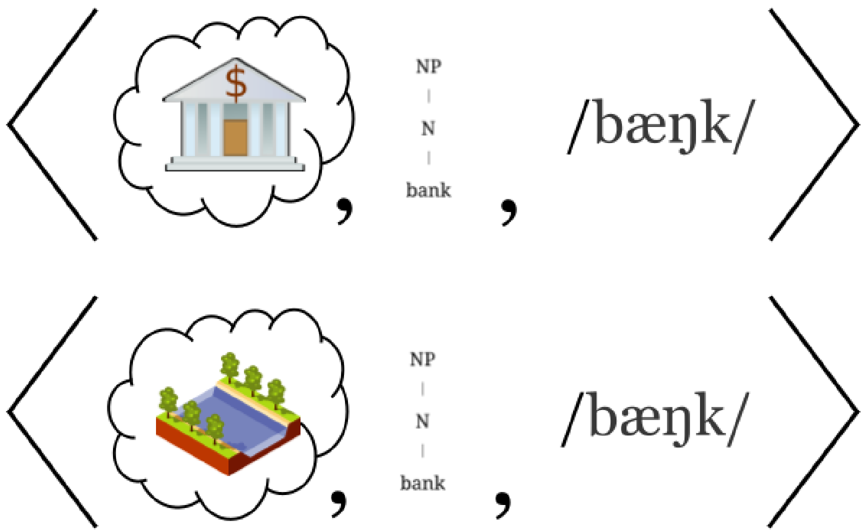

Figure 2.2: A visual representation of the ambiguous word "bank"

If we make this assumption, we would not expect there to be cases more complex than these simple ambiguities, ones where meaning, syntax, and form would be mapped to one another in more complicated ways, or instances where the syntactic context would impact the form or meaning of individual words. However, there is a wide array of evidence against the idea that lexical items can be represented neatly as triads of meaning, form, and syntax. Some of the key evidence concerns the status of roots, which, it turns out, cannot be identified directly from their form nor directly from their meanings, and must instead be identified in purely syntactic terms. These roots and their syntactic context can be mapped to sets of units of meaning and sets of units of form, but are not fully specified for either meaning or form. In contrast to the "triad", the mappings between syntax and meaning and syntax and form may not be “symmetrical” - a single piece of syntax which is individuated in terms of meaning may not be individuated in terms of form, and conversely, a single piece of syntax which is individuated in terms of form may not be individuated in terms of meaning. Figure 2.3 below illustrates the “symmetric” and “asymmetric” mappings in the idiom *went off* (as in “the alarm *went off* at 7:15 A.M.”) In a theory which requires symmetric mappings, the individual phonological words *went* and *off* must map to pieces of syntax, which must map to individual meanings. Given this, it is challenging to give reasonable meanings to *went* and *off* that would make it possible to derive a meaning like “rang”. In a theory which allows asymmetric mappings, the meanings of “past” and “ring” can map to sets of syntactic units, which can then be mapped to different phonological words; the sets of meanings and phonological units do not necessarily need to be aligned.

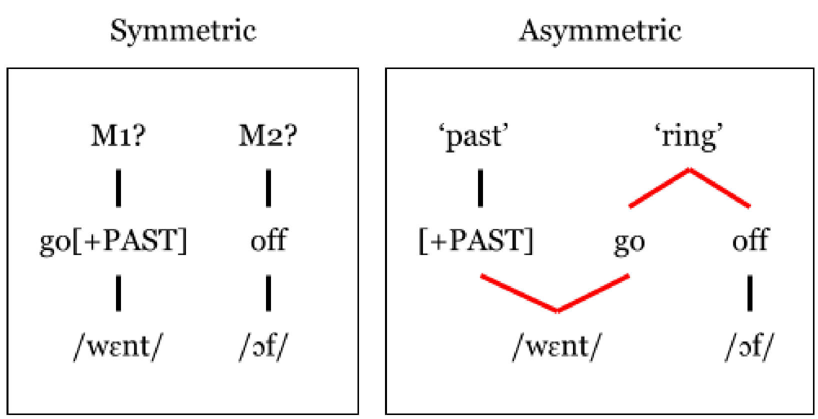

Figure 2.3: Symmetrical vs. asymmetrical representations of “went off”

Another reason to move away from the triad is that the mapping between syntactic units and form may not be fixed, and may not be determined by those syntactic units alone. Harley (2014) argues that roots do not have a fixed form, and that the phonological form must be conditioned by the syntactic context that it occurs in. The primary evidence for this comes from suppletion. In English, there are only a few restricted examples of suppletion (*go* ∼ *went, bad* ∼ *worse, person* ∼ *people*), which might be considered to be semantically "light", with meanings that may be bleached, bordering on the status of "grammatical morpheme." However, in languages such as Hiaki[3], there are many more examples of suppletion for verbs with singular and plural subjects, as well as singular and plural objects, a subset of which are shown in (9):

(9)
a. vuite ∼ tenne — *run.sg* ∼ *run.pl*

b. siika ∼ saka — *go.sg* ∼ *go.pl*

c. weama ∼ rehte — *wander.sg* ∼ *wander.pl*

d. kivake ∼ kiime — *enter.sg* ∼ *enter.pl*

e. vo'e ∼ to'e — *lie.sg* ∼ *lie.pl*

f. weye ∼ kaate — *walk.sg* ∼ *walk.pl*

g. me'a ∼ sua — *kill.sg**Obj*** ∼ *kill.pl**Obj***

Suppletion of this sort indicates that phonological form cannot be determined by the root alone, but instead requires reference to the syntactic context that the root occurs in. If those roots were inserted into the syntactic structure already specified for their phonological form, as the lexicalist approach would suggest, suppletion like this could not occur; as shown below in Figure 2.4a-b, only one phonological form could be inserted. Even with something like feature percolation, where the +PL or +SG features could be made accessible to the verb indirectly, verbs like (9g) that involve suppletion based on the number of the object would be especially problematic; the features of the object (as well as the subject) would need to be able to percolate

[3] Hiaki (also referred to as *Yaqui* or *Yoeme*) is an Uto-Aztecan language spoken in the states of Arizona (USA) and Sonora (Mexico).

up, but only for that particular set of verbs that involve object-based suppletion. A lexicalist account might also argue that the verb in (9g) is not a case of true suppletion, but that it is actually multiple verbs that have different argument structures that select for singular or plural objects[4]. This would suggest that a single meaning would appear in multiple triads that have different syntactic and phonological content, as a case of synonymy; this would not account for the distributional facts that are characteristic of suppletion (i.e., *vuite* never appears with a plural subject, and *tenne* never appears with a singular subject). Alternatively, a context rewriting rule could help to generate the correct form, but that would be using a unit of form that is larger than the units of meaning that were inserted into the structure initially, a deviation from the lexicalist triad. These conclusions are not very appealing, partly due to their lack of parsimony, and partly because a lexicalist approach would not expect these kinds of forms to appear in the language in the first place; these are "post-hoc" explanations. Because this is an instance where the form of the verb is conditioned on its morphosyntactic context, as suggested by Figure 2.4c below, a syntactic theory which relies on the "triad" notion of lexical items would struggle to account for this data.

[4]It is important to note here that this is about the syntactic feature of number, not the semantic feature. Something being semantically plural does not necessitate that it is syntactically plural, and vice versa. For example, "scissors" is syntactically plural, while being semantically singular, while "furniture" is syntactically singular while being semantically plural. Agreement is always with the syntactic features, not the semantic ones. Some verbs like "juggle" seem to require a plural object (# John juggled the task), but the sentence is still syntactically well-formed (contrast with a sentence like "the sharp scissors cuts the paper easily", which involves agreement mismatch). As a consequence of this, it cannot be the case that the features necessary for agreement are available at a conceptual level to help determine the form of the verb.

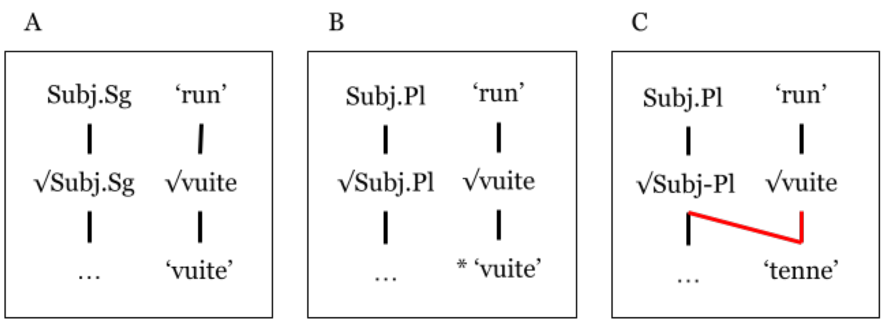

Figure 2.4: Symmetrical vs. asymmetrical representations of Hiaki suppletion, where the output of B would be ungrammatical.

The lexicalist approach to inflection more generally, even in English, is problematic for their own theory. Suppletion for the past tense inflection in English (as in *go* ∼ *went*) is not seen as an issue in the lexicalist approach because the inflected forms can be stored together in the same lexical entry, and the correct form can be inserted into the structure because the PAST feature is made available through feature percolation. However, even though those that take the lexicalist approach are happy with this solution, that solution does not support the notion of a triad. The issue is that the context relevant to choosing among these forms (the choice between *go* and *went*) is a context assembled by the syntax (the PAST morpheme is or is not combined with the root for *go* depending on the syntactic structure which has been built). If the form associated with a verb root cannot be determined unless and until the syntactic structure is complete, it cannot be the case that the lexicon lists pairings of form and meaning, or triads of form, syntax, and meaning, independent of the syntactic computation involved.

In addition, Harley (2014) also discusses evidence against the triad view on the meaning side, arguing that units of meaning can be both smaller and larger than individual units in the syntax. Some Hebrew roots, for example, do not have a consistent meaning across different morphosyntactic contexts, as shown in (10). Several studies have shown that the triconsonantal roots in semitic languages are a morphologically real element of the grammar (Ussishkin et al., 2015); these should not be taken to be an unrelated set of homophonous consonants. However, the meanings of each of the words in (10) are very hard to unify under a single meaning rubric. There are diachronic reasons that these meanings came to be, but those reasons are not accessible to the Hebrew-learning child. Because these individual roots only receive a meaning when combined with specific templatic morphology, these cases illustrate that there is a structured unit of form "k-b-ʃ" for which one cannot identify the unit of meaning that would be associated with it, and therefore the units of meaning in this case must be larger than the individual units in the syntax.

(10)
a. keveʃ — *'gangway, step, degree, pickled fruit'*
b. kviʃ — *'paved road, highway'*
c. kviʃa — *'compression'*
d. kviʃan — *'furnace, kiln'*
e. maxbeʃ — *'press, road roller'*
f. mixbaʃa — *'pickling shop'*
g. kavaʃ — *'to conquer, subdue, press, pave, pickle, preserve, store, hide'*
h. kibeʃ — *'to conquer, subdue, press, pave, pickle, preserve'*
i. hixbiʃ — *'subdue, conjugate'*
j. kavuʃ — *'subdued, conquered, preserved, pressed, paved'*
k. kvuʃim — *'conserves, preserves'*
l. mexubaʃ — *'pressed, full'*

Similarly, in English, there are several examples of "cranberry" morphemes, morphemes that have no meaning outside of their morphosyntactic context, as shown in (11). Aronoff (1976) uses these "cranberry" morphemes to argue that the minimal unit in the lexicon must be the word, rather than the morpheme, because these morphemes do not themselves constitute Saussurean signs. It is important to note here that appealing to the word level does not resolve the issue, as in the case of words such as *cahoot* (*in cahoots*) or *shrift* (*short shrift*), which lack meaning except when in a very particular context. These are examples of "cranberry words", in that they are words that have no meaning (for many speakers) outside of their morphosyntactic context. Furthermore, based on the discussion in Section 2.1 and Section 2.2, morphemes should also be able to be listed in the lexicon; given the behavior of these cran-morphemes, we need to reject the idea that the lexicon necessarily consists of Saussurean signs.

(11) a. *-ceive*: *deceive, receive, conceive, perceive*; all undergo the *-ceive/ception* alternation

b. *-pose*: *oppose, suppose, depose, compose, repose, propose*; all undergo the *-pose/position* alternation

c. *-port*: *comport, deport, report, import, support*

d. *-here*: *adhere, inhere*

Especially in the case of *-ceive* and *-pose*, where there are uncommon irregular alternations that systematically apply across the group of words[5], it seems that there must be a common root among each set of verbs, one which lacks a consistent meaning across different morphosyntactic contexts, but which can participate in predictable phonological alternations. In the case of words like *cahoot*, there is only one morphosyntactic context which can provide its meaning. If these

[5]The *-ceive* ∼ *-ception* alternation is highly unpredictable, compared to other phonologically similar verbs like *sieve* or *believe*, and given the range of other nominalizing affixes such as *-ance*, *-ment*, *-tion*, or *-al*. The *-pose* ∼ *-position* alternation is similarly unpredictable, even though it may seem similar to *-tion*.

were in fact triads of meaning, syntax, and form, we would expect each of these groups of words in (10) and (11) to have consistent meanings that would correspond to the meaning of the root. If, instead, these are fully abstract roots that are supplied a meaning based on the morphosyntactic context, the behavior of these morphemes is not surprising.

Verb-object idioms provide additional evidence that the meaning of a root must be provided by its morphosyntactic context. Examples such as those in (12) indicate that the meaning of verbs like *kill*, *pass*, and *take* can be dependent on the semantic content of its object, while remaining indifferent to that of its subject. Meanwhile, the object has a fully transparent, non-idiomatic meaning. Idioms like these are very common, and are used in a variety of registers; it would be inadvisable to treat these as exceptions, or as simple cases of lexical ambiguity.

(12) a. Kill: *kill a bottle, kill an evening, kill the clock, kill the music*

b. Pass: *pass a test, pass a law, pass a kidney stone, pass the hat*

c. Take: *take a photo, take a nap, take a bus, take a chance*

If the meaning of *kill*, *pass*, or *take* was already specified in the lexicon, we would not expect any of these verb-object idioms to emerge with these idiosyncratic meanings. In these cases, and in many others that were not listed in (12), the meaning of the verb can only be determined when it is interpreted in its morphosyntactic context. Based on this evidence, it must be concluded that roots must be individuated purely abstractly, not specified for meaning or form.

Along the same vein, Noyer (1998)'s discussion of Vietnamese idiomatic collocations suggest that "lexical items" can include multiple syntactically separable pieces. These examples are presented in (13) below. In this case, lexicalist approaches would treat these as two "words", because they are syntactically separable. However, their meaning cannot be interpreted independently.

(13) a. Tôi xay **nhà** **cửa** → Tôi xay **nhà** xay **cửa**
I build house door → I build house build door
'I build a house'

b. Tôi không muốn **đèn** **sách** → Tôi không muốn **đèn** không muốn **sách**
I NEG want lamp book → I NEG want lamp NEG want book
'I do not want to study'

c. Tôi lo **vườn** **tược** → Tôi lo **vườn** lo **tược**
I care.for garden XX → I care.for garden care.for XX
'I take care of gardens'

This data is challenging to a lexicalist approach, because syntactic rules must be able to operate across the boundary between morphology and syntax in order for these to be syntactically separable. If atomicity were to be preserved, one could assume that *nhà cửa*, *đèn sách*, and *vườn tược* were all composed of two words, but the idiomatic reading of each phrase and unavailable interpretation of *tược* would provide an additional complication that the lexicalist approach could not account for. For a theory which allows them to be syntactically complex (similar to tree-adjoining grammars; Joshi and Schabes, 1997) where the stored pieces of syntactic structure are combined with other stored syntactic structures, these examples suggest that it cannot be as easy as "clipping" these idioms into a syntactic structure - there must be an additional process of adjunction or movement which occurs to allow this separation, suggesting that the lexicalized structures of the elementary trees cannot be truly atomic.

Based on all of the evidence discussed above, "lexical items" as they are used in lexicalist literature cannot exist, at least not as triads of meaning, syntax, and form. For this reason, I adopt an approach in which syntactic units are fully abstract, lacking meaning or form, and become associated with meaning or form via mapping rules from sets of syntactic units to sets of units of meaning or form, as argued by Preminger (2021). I prefer a fully abstract, late-insertion approach over one that specifies a list of forms in the syntax (e.g., Bresnan, 1978); with both approaches, where there is either a list of forms that is included in the syntax, or there is a process which links those abstract objects with a form later, the system must be able to modify forms based on syntactic context after the syntax has been generated. To demonstrate that forms are present in the syntax, one would have to show that phonological forms can modulate syntactic relations (e.g., binding, theta roles, etc.) independent of their syntactic configuration (e.g., the fact that "what" begins with *wh-* is what causes wh-movement, and not because of its syntactic properties). Even if some cases suggest that form does influence linear order, this presents another set of options: one could either treat those phenomena as syntactic, meaning that syntax has to be sensitive to more than just syntactic properties (but only in a limited way, not universally; this then requires an account of why not all phonology modulates the syntax), or treat them as post-syntactic or prosodic, where those phenomena are handled by a mechanism that is external to the syntax. By making syntax fully abstract (no meaning, form, or linear order), syntax is not required to be as powerful to account for those kinds of phenomena. Instead, it is only required to encode abstract hierarchical structure, acting as a "hidden layer" between meaning and form. The abstract syntax becomes associated with form via mapping rules from sets of syntactic units to sets of form units, and they become associated with meaning via mapping rules from sets of syntactic units to sets of meaning units. Those mappings can also be represented in a more probabilistic format, as a calculation over larger or smaller pieces of syntax. The mappings to meaning and form also need to be independent to allow for asymmetric mappings, rather than operating as one unified set that recreates the "triad" with different terminology, since that would run up against the same set of issues just enumerated.

2.4 In summary

The data presented here suggest that lexicalist theories of syntax are unable to account for many phenomena in a variety of languages, including English. The theory of morphosyntax should not assume that there is a split between morphology (or other sub-word operations) and syntax (or other supra-word operations); there must be one morphosyntax which operates over all morphosyntactic objects. Wordhood does not definitionally correspond to listedness, because many things both larger and smaller than the word level should also be listed. Lastly, "lexical items" as triads of meaning, syntax, and form cannot exist; they should instead be represented as mappings from meaning to fully abstract syntactic atoms, and mappings from fully abstract syntactic atoms to form.

As will be discussed in the next chapter, these lexicalist assumptions appear not only in syntactic theory, but also in psycholinguistic and neurolinguistic theories, either implicitly or explicitly. Many models of language production and comprehension, language disorders such as aphasia, and neural models of the language system often assume separate subdomains for "syntax processing" and "morphological processing", and often assume that the lexical items that are retrieved from memory and combined in the syntax are words which are stored as triads of form, meaning, and syntax. They may also assume that syntax is computed word-by-word, as an incremental process. These assumptions may be very appealing, seeming like innocent simplifying assumptions, helping to idealize over some of the "peculiarities" of natural language. However, they carry with them a great deal of theoretical baggage, and models built around them should exhibit the same vulnerabilities as those discussed above, not just relating to syntax, but also in other aspects of the processing theory.

This is partly an issue of linguistic diversity; much of the data discussed above and in the following sections comes from a typologically diverse set of languages, illustrating some of the phenomena that existing psycholinguistic models may struggle to account for. However, many of these phenomena can also be found in English, and these non-lexicalist conclusions should hold for any language, given the assumption that all languages utilize the same underlying cognitive processes. There is a space here for additional hypotheses about language-specific, storage-based processing optimization that could resemble some (but not all) properties of lexicalism, but those hypotheses should be discussed intentionally and empirically tested. This possibility will be discussed in more detail in Section 3.1.4.

At this point, one might be asking, what is a "word" then, if not a triad of meaning, syntax, and form? It is true that as language users, we seem to have intuitions about wordhood, about what constitutes a single word and what does not (even if those intuitions may vary). However, those intuitions are hard to formulate into a coherent hypothesis about linguistic units, as discussed by Haspelmath (2017). Wordhood should not be defined as a "unit of meaning", because things which are "intuitive words" may not be meaningful (as in the expletive *it* in the sentence *it is raining*), because a single unit of meaning may not correspond to an intuitive word (as is the case for idioms and some compounds), and because an intuitive word may not correspond to a single unit of meaning (a verb often includes tense morphology which encodes additional meanings, and the same holds for contractions[6] and morphologically complex intuitive words). We

[6]It should be noted that contractions are a case where speakers seem to have less clear intuitions about wordhood (e.g., whether "we've" should be counted as one word or two).

also cannot ground intuitive wordhood in being a “unit of syntax”, because syntactic operations can apply to units smaller than intuitive words, as illustrated by the Inuktitut and Vietnamese examples. It also does not seem that we can ground intuitive wordhood in being a “unit of phonology”, because there exist many phonological words (which define the domain of phonological operations) which are not intuitive words, such as *“dyawanna”* (“do you want to”) in English, and cases where an intuitive word does not correspond to an independent phonological word, as is the case for many function words in English (Selkirk, 2014). It could be that most of our intuitions about wordhood are in fact grounded not in natural spoken language, but in orthography, among literate communities whose writing system make use of white spaces as separators. For readers of such orthographies, “word” could serve as a useful term for the things between white spaces, which might well define processing units for the reading modality. However, many other writing systems have not made use of this convention, and it is notable that those speakers often have much less developed intuitions about wordhood (Hoosain, 1992). In summary, it is hard to see how speakers’ intuitions about wordhood systematically correspond to any representational or processing unit of natural spoken language, although they could correspond to units of certain written languages. Ultimately, it may not correspond to anything useful for our theory of language or our model of language production after all.

Chapter 3

Issues with lexicalism in psycho- and neuro-linguistics

In standard models of language production or comprehension, the elements which are retrieved from memory and combined into a syntactic structure are “lemmas” or “lexical items”. Such models implicitly take a lexicalist approach, by assuming that those lexical items store meaning, syntax, and form together, that syntactic and lexical processes are distinct, and that syntactic structure does not extend below the word level. As summarized in the previous chapter, linguistic research examining a typologically diverse set of languages has provided strong evidence against this approach. These findings suggest that syntactic processes apply both above and below the “word” level, and that both meaning and form are partially determined by the syntactic context. This has significant implications for psychological and neurological models of language processing as well as for the way that we understand different types of aphasia and other language disorders. As a consequence of the lexicalist assumptions of these models, many kinds of sentences that speakers produce and comprehend - in a variety of languages, including English - are challenging for them to account for.

In this chapter, I examine how lexicalism has influenced psycho- and neuro-linguistics, discuss the consequences for the models that make lexicalist assumptions, and argue that a non-lexicalist approach is necessary for constructing more accurate models of language processing.

In particular, I focus on models of language production as a sort of case study, but I encourage readers to reflect on their own approaches using this case study as a model. The critiques of lexicalism and its effects in these models should apply to any kind of model or theory of language and language processing which makes either of these lexicalist assumptions, including sentence processing and single-word lexical processing, both in comprehension and in production. The discussion in this chapter will focus on two key properties of lexicalist models of language production:

1. **Syntactic and morphological processes are different in kind:** Under this assumption, morphology (or other sub-word operations) and syntax (or other supra-word operations) are fundamentally different operations. Each has their own sets of atoms and rules of formation; syntactic rules operate over phrases and categories (NP, V, etc.), while morphological rules operate over roots, stems, and affixes. This establishes words as the "atoms" of syntax (Lapointe, 1980; Chomsky, 1970; Williams, 1981). Some interaction needs to exist between syntax and morphology, such as in verbal inflection, but lexicalist theories argue that the interaction functions in such a way that the two sets of rules and operations are not intermixed, and that only certain components can be referred to in both sets of rules.

2. **Lexical items include triads of sound, meaning, and syntax:** According to this assumption, everything which can be syntactically individuated has its own context-independent meaning and form. This creates a "triad", where each lexical item links a single meaning representation to a piece of syntax and a single form representation. The size and complexity of the piece of syntax can vary across theories; in some accounts, the syntactic component only contains a single syntactic terminal or a set of features (Jackendoff, 1975;

Aronoff, 1976; Di Sciullo and Williams, 1987; Pollard and Sag, 1994), while in other accounts the syntactic component can be a “treelet” or “construction” that is morphosyntactically complex, thereby rejecting the first assumption above but retaining lexicalist properties (Kempen and Hoenkamp, 1987; Vosse and Kempen, 2000; Matchin and Hickok, 2020, among others).

Lexicalist assumptions have played a central role in the development of models of language processing, either explicitly or implicitly. Many models of language production assume something like a lemma or lexical item, which functions as a stored triad of form, meaning, and syntax, also codifying a distinction between morphology and syntax. These models also create a division between lexical and syntactic processes, treating morphology as a different system from syntax. I introduce specific phenomena in several different languages, which are meant to represent a variety of phenomena across human languages. These phenomena are not isolated instances that can be treated as outliers, but rather common occurrences in human language that also need to be accounted for in models of language processing.

The issues discussed here are partly related to linguistic diversity in model development. Using one’s own language to generate models of language in general is not necessarily an issue - if you want to know how language *in general* is processed, a good place to start is to look into how *one* language is processed. However, a phenomenon which is deemed to be “exceptional” in one language - and thus exempt from the usual steps in linguistic processing - may be commonplace in other languages. Given the assumption that all languages utilize the same underlying cognitive processes, our models also need to account for those kinds of data.

3.1 Lemmas and other lemma-like things

Many models of language production rely on the notion of "lemmas" (Kempen and Huijbers, 1983; Levelt, 1989; Levelt et al., 1999; Bock, 1995). According to the Levelt model, a lemma is a representation which stores syntactic information, and also points to a conceptual representation and a phonological form, bridging the Conceptual Stratum, Lemma Stratum, and Form Stratum. In this model, there is a lemma for every "lexical concept", and once a lemma has been selected for production, the lemma activates the phonological codes for each of its morphemes. These models commonly assume that the lemma is a terminal node in the syntactic structure (Levelt, 1992). Syntactic frames for these lemmas can specify how semantic arguments - such as "theme" or "recipient" - should be mapped onto syntactic relations - such as direct or indirect object (Levelt and Indefrey, 2000). Syntactic structure is built by combining multiple lemmas which have been retrieved from memory, according to their selectional restrictions and syntactic frames that are provided.

The diagram in Figure 3.1 of the lemma for the word *escorting* (from Levelt et al., 1999) illustrates how the lemma uniquely identifies a lexical concept in the Conceptual Stratum. The lemma has a number of "diacritic parameters" which need to be specified, including features such as number, tense, aspect, and person. These features may be prepared at the conceptual level or at the point of grammatical encoding. The lemma and its given features point to the phonological form of the stem *escort* and its suffix *-ing*, along with the metrical structure of the word. For morphologically complex words like *nationalize* and compounds like *afterthought*, the lemma model assumes a single simplex representation at the lemma stratum which maps to several form pieces in sequence at the form stratum (Roelofs et al., 1998). There are slight variations in the

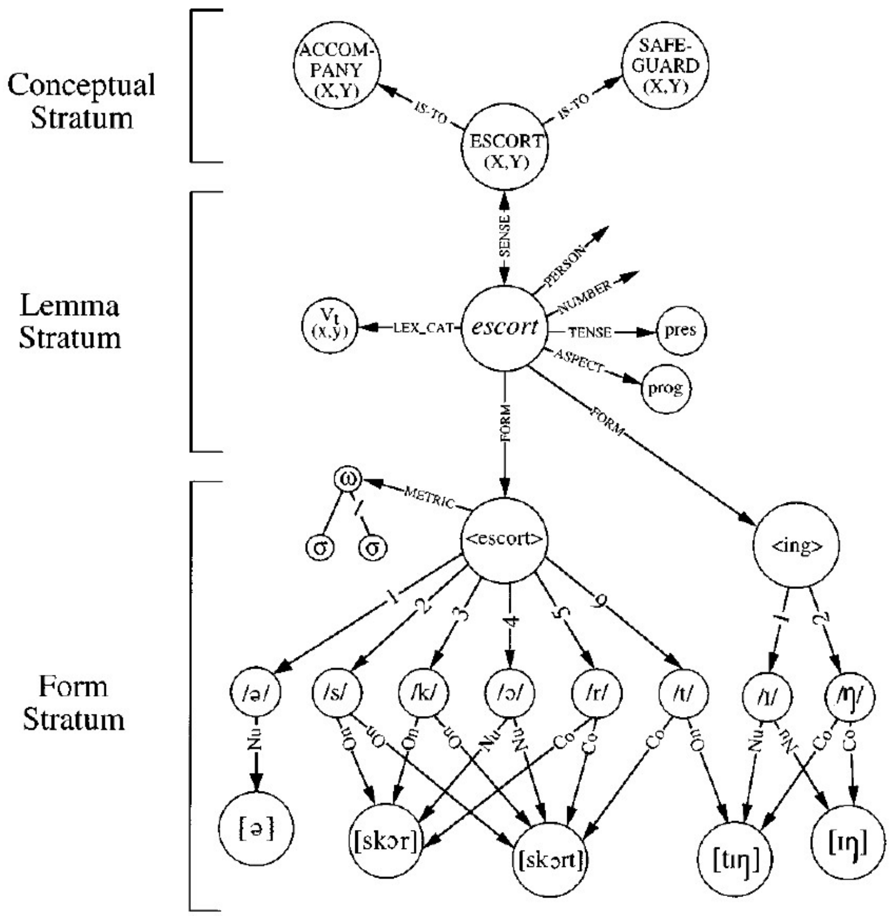

Figure 3.1: Lemma representation of the word "escorting", from Levelt et al. (1999) A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22, 1-38, reproduced with permission.

assumptions made by different lemma models of language production; for example, according to Levelt and Indefrey (2000), function words have their own lemma, while in the Consensus Model (Ferreira and Slevc, 2007), they do not. Some production models refer instead to "lexical items", but these are usually given similar attributes as lemmas and embody the same lexicalist assumptions.

3.1.0.1 Lemmas encode a distinction between lexical and syntactic processes

The lemma codifies a fundamental distinction between morphology and syntax. Morphologically complex words are taken to embody complexity in lexical representations and retrieval processes, rather than syntactic complexity. Because inflectional morphology and derivational morphology is stored within lemmas, and syntactic properties of the lemma are only represented by features obtained through indirect interaction, the lemma creates a "bottleneck" between morphology and syntax. For English, this might seem reasonable, but for languages with richer morphology and inflectional paradigms, the lemma becomes increasingly unwieldy. For example, in polysynthetic languages, a single word can be composed of many productive morphemes, representing complex meanings. In order to represent those words as lemmas, each lemma would have to correspond to very complex lexical concepts, with many redundant lemmas, to represent all of the possible morpheme combinations in that language; alternately, each lemma would have to incorporate a massive set of features in order to have a "complete" inflectional paradigm.

Along a similar vein, the idea that lemmas only exist for words and their inflections and derivations, reinforces the idea that it is only *complete* words that are stored in the lexicon, rather than pieces smaller or larger than a word. We can take as an illustration the commonly cited myth that “Eskimos have 150 words for snow”, which has been debunked several times over (Martin, 1986; Pullum, 1989; Kaplan, 2003). As polysynthetic languages, Eskimoan languages such as Inuktitut have several main “snow” root morphemes (*aput*, “snow on the ground”; *qana*, “falling snow”, *piqsirpoq*, “drifting snow”; *qimuqsuq*, “snowdrift”) which can be combined productively with a wide array of other morphemes to create a massive number of words relating to snow: types of snow, quantities of snow, adjectival forms such as “snow-like”, verbs involving snow, verbs where snow is the object, and so on. One could describe this situation by saying that Inuktitut has a tremendous number of “words” for snow and snow-like things, but this would be a bit like noting that English has a tremendous number of phrases or sentences about snow - it is simply not a very useful description of the language.

Because the lemma model assumes that morphological structure and syntactic structure is fundamentally different, and that derivational and inflectional morphology is stored within the lemma (and not built on-line like syntactic structure is), the individual morphemes within each word cannot exist independently of the lemmas that they appear in. Consequently, the lemma model has two options. One is to assume that each derived form in Inuktitut constitutes a separate lemma, and thus that there are 150+ different lemmas for each derived form of “snow”; this creates a great deal of redundancy, since each lemma would list the same root morpheme separately. The other option is to assume that there is a single lemma for *snowflake* stored with a massive inflectional paradigm that can generate all the derived forms that include the snowflake morpheme. This same dilemma would arise for every root in the language, of which there are thousands. For these languages, the lemma - as it is currently defined - is not a useful construct.

Let’s look at a few examples from Inuktitut[1] to appreciate the challenges polysynthetic

[1]These examples come from the Utkuhiksalingmiut dialect of Inuktitut, which is currently spoken in the Inuit

languages pose for lemma models of production (examples from Briggs et al., 2015; Cook and Johns, 2009):

(14) a. nivak -tuq
shovel.debris -PTCP.3S
'She shovels debris, old snow [out of the door]'

b. uqaalla -qattaq -tunga
say -often -PTCP.1S
'I say that sometimes'

c. havauti -tuq -ti -taq -niaq -tara
medicine -drink -cause -frequently -going.to -PTCP.1S/3S
'I'm going to give her medicine frequently'

The sentence in (14a) is a good example of a case that the lemma model can handle with the same machinery used for English and Dutch inflectional morphology, as illustrated for "escorting" in Figure 3.1. The *nivak* lemma could simply be specified with inflection diacritics for mood, person and number, agreeing with the (null) subject. If we turn to the sentence in (14b), perhaps the lemma representation could remain simple as in (14a), and the complexity could be limited to the form level as the sequence of forms, *uqaalla*, *-qattaq*, and *-tunga*, similar to how the model represents compounds and other derived forms. However, since the lemma model assumes that each lemma corresponds to a single stored "lexical concept", this case would require assuming that speakers store atomic lexical concepts like "I say that sometimes". A case like (14c) appears more challenging yet to represent as a single inflected lemma. How might the lemma model try to represent the many different units used to generate this single complex word?

One possibility, following (14b), would be to assume that there is a single stored lexical concept that corresponds to the entire meaning "I'm going to give her medicine frequently", and

communities in Gjoa Haven, Baker Lake, and formerly in the Black River area of Nunavut.

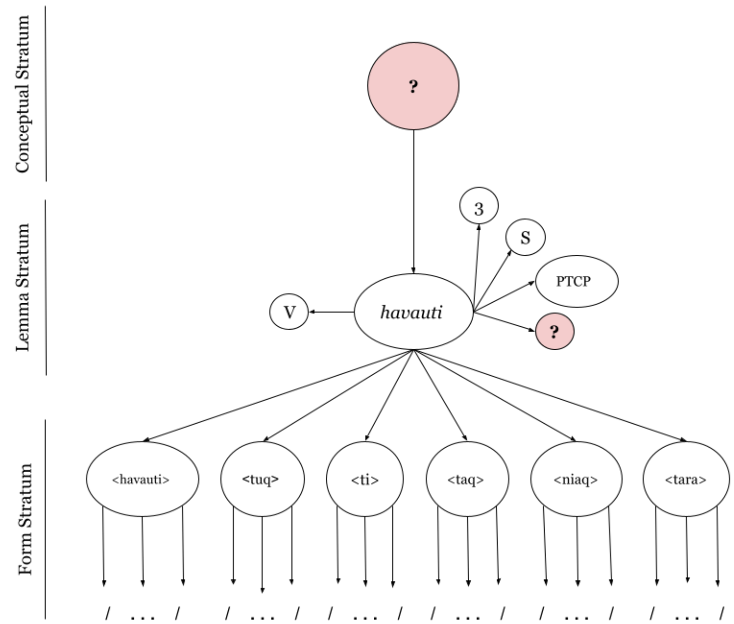

Figure 3.2: A possible lemma representation for the Inuktitut example in (14c), *havautituqtini-aqtara*, "I'm going to give her medicine frequently".

thus a single corresponding lemma, with complexity at the form level only. This seems implausible. This would mean storing as separate full lexical concepts the meanings corresponding to every similarly-structured word that speakers produce (e.g., "I'm going to give her vitamins frequently"), and would put pressure on the theory to provide a systematic account of how this multiplicity of lemmas containing productive derivational morphology was created in the first place.

An alternate approach to (14c) would be akin to the inflectional morphology case, to assume a core lemma for the lexical concept "medicine", and then generate the complex utterance in (14c) from a set of diacritics on the lemma, as illustrated in Figure 3.2. But this would lead to another question about what kinds of diacritic features could possibly represent each of those morphemes, especially if they would go unused in the majority of cases (the morpheme for "drink", *-tuq*, would appear relatively rarely, seemingly not enough to justify its status as a feature in the lemma representation, in contrast to features like tense or number), and considering that they can be used productively. Furthermore, the relationship between the morphemes within a lemma is only one of linear order, so this would mean that no non-linear structured relations between the elements of (14c) could be represented by the lemma. This would be problematic given the large body of evidence from polysynthetic languages for non-linear (hierarchical) relations between the elements within morphologically complex words [2], as shown in (3).

If one sticks with the core idea of the lemma model, that lemmas are defined such that a single lemma corresponds to a single lexical concept, intuitively the best solution to (14c) is to assume that the individual morphemes within the word like those for "medicine", "drink", and "frequently" have their own stored lemmas. This means giving up a view of production in which stand-alone words always correspond to stored lemmas, and instead adopting the non-lexicalist assumption that morphologically complex "words" can be constructed in the course of production in the same way that sentence structure is. Although the need for this move is most obvious in the case of the production of languages with rich morphology, assuming a processing model in which lemmas can be combined to form structured words provides a needed account of productive morphological word formation in languages like English or Dutch as well.

As discussed in Section 2.3, there is additional evidence that syntactic rules must be able to operate across the boundary between morphology and syntax, challenging the lexicalist notion of the "atomicity" of words, that words are the units of syntactic combination. Idiomatic colloca-

[2] There is a wide array of evidence that morphemes are hierarchically structured, both from lexicalist and non-lexicalist accounts. For example, the English word "unlockable", can either mean "able to be unlocked" or "not able to be locked"; the ambiguity in meaning can be analyzed as a structural ambiguity between [[un - lock] - able] and [un - [lock - able]]. The debate here is not whether morphemes are hierarchically structured, but whether that hierarchical structure is morphosyntactic or purely morphological in nature. Baker (1985) and other non-lexicalist approaches argue for the former, while lexicalist accounts argue for the latter. Morphemes only being linearly ordered is a more general issue for the lemma model, not just because of their lexicalist assumptions.

tions in Vietnamese are composed of several morphemes, which in some cases are syntactically separable, as shown in (13) and repeated below in (15), where the collocations preserve their idiomatic interpretation when separated by other syntactic material (often used in Vietnamese for stylistic effect or affect):

(15) a. Tôi không muốn **đèn** **sách** → Tôi không muốn **đèn** không muốn **sách**
I NEG want lamp book → I NEG want lamp NEG want book
'I do not want to study'

b. Tôi lo **vườn** **tược** → Tôi lo **vườn** lo **tược**
I care.for garden XX → I care.for garden care.for XX
'I take care of gardens'

According to the lemma model, these idiomatic collocations would need to constitute single lemmas with multiple morphemes. Each collocation would correspond to a single lexical concept because of their idiosyncratic meanings - and in some cases, parts with unavailable meanings of their own (indicated by "XX" in the gloss). Furthermore, in (15a), though *đèn* ("lamp") and *sách* ("book") are nouns individually, when used together they function as a verb; because syntactic category is a property of lemmas and not morphemes, this provides further evidence that they must constitute a single lemma. However, if a sequence like *đèn sách* corresponded to a single lemma with separate pieces at the form level only, then the two pieces of the collocation could only appear adjacently and would not be syntactically separable, no different from *escort* and *-ing* in Figure 3.1.

Some work in the lemma tradition has tried to develop an alternative approach to deal with phrasal idioms. Sprenger et al. (2006) and Cutting and Bock (1997) argue that idioms have a "hybrid" representation, where there is a lexical concept node or "superlemma" for the idiom which also activates the lemmas of its constituents (i.e., the superlemma for "kick the bucket" would activate the simple lemmas for "kick" and "bucket"). One of the key assumptions of these accounts is that each of the constituents of the idiom must have its own lemma representation that can be activated. Because all lemmas must have an associated lexical concept, this assumes that every idiom would have a literal interpretation which is overridden by the idiomatic interpretation. However, for the Vietnamese idiomatic collocations, and example (15b) in particular, this claim would be problematic. The morpheme *tược* has no interpretation outside of the idiomatic collocation, so it could not correspond to a lexical concept independent from the idiom; thus, there could not be a *tược* lemma which could be activated. Furthermore, Kuiper et al. (2007) argues that the superlemma specifies only phrasal functions between simple lemmas (constituting a VP or NP, for example), rather than sub-word pieces or a single syntactic category. This would be a problem for the *đèn sách* ("study") example, where two nouns are compounded to form a verb; a VP requires a verb head, but neither element would be able to serve that function (in contrast to English phrasal idioms like the VP "kick the bucket", or the NP "kit and caboodle").

These examples challenge one of the key assumptions of the lemma model, following the lexicalist approach, that syntax and morphology are separate operations that cannot interact. In order to account for these kinds of examples, the only solution would be to assume instead that the *đèn* and *sách* morphemes within the "study" lemma are themselves syntactic objects that can interact with the syntactic structure. This means giving up a view of syntactic structure where words or lemmas are the units of combination, and instead adopting the non-lexicalist view that morphology and syntax are part of the same system. The evidence from Inuktitut and Vietnamese indicates that, not only do we need to move away from a view of production in which stored lemmas correspond to words, but we also need to give up the idea that the units of language production are syntactically atomic by definition.

3.1.1 Lemmas function as a stored triad

The lemma is defined as grouping together form, meaning, and syntax, creating the "triad" illustrated in Figures 2.1 and 2.2. The lemma maps between meaning, syntax, and form in a "symmetrical" way, where for every element that is syntactically individuated, it is also individuated in terms of meaning and form, and it is not dependent on other lemmas or the syntactic context to determine meaning or form. Even if the phonological form is not stored within the lemma itself, the mapping between lemma and form is deterministic and context-independent. If we make this assumption, we would not expect there to be cases where meaning, syntax, and form would be mapped to one another in more complicated ways, or instances where the syntactic context would impact the form or meaning of individual words.

One place in which the phonological form seems to be conditioned by the broader syntactic context that it occurs in is suppletion. Existing models have a way to account for some kinds of suppletion, such as what is seen for a few English verbs, based on tense (*go* ∼ *went*) and agreement with the subject (*is* ∼ *am*). However, it is harder for this kind of model to account for suppletion based on a larger piece of syntax, where the form is not determined by a single syntactic object or a limited set of features, but by the larger syntactic context. As discussed in Chapter 2.3, example (9), Hiaki exhibits suppletion for some verbs with singular and plural subjects, as well as singular and plural objects (the verb for "run" arises as *vuite* when the subject is singular, and *tenne* when the subject is plural; the verb for "kill" arises as *me'a* when the object is singular, and *sua* when the object is plural).

For the English verb *escorting* above, the diacritics for the person and number of the subject help to determine the inflection on the verb for agreement; for the Hiaki verbs that exhibit

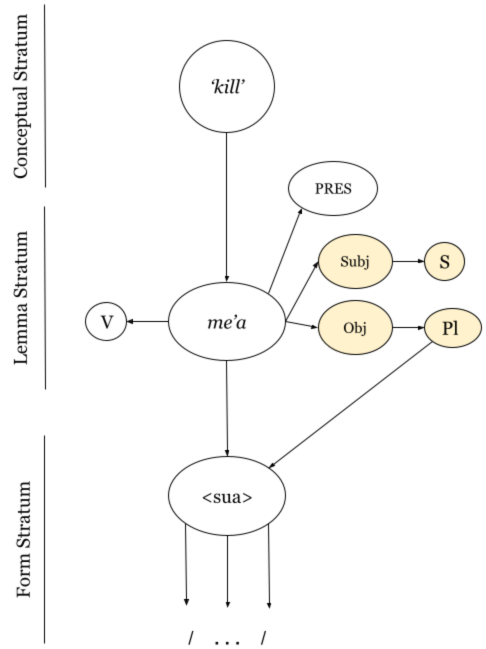

Figure 3.3: A possible lemma representation for the Hiaki suppletive verb *sua*, "kill.plObj". The "Subj" and "Obj" diacritics would indicate the number feature on the subject and object of the sentence, but they would often be redundant, given that regular verbs in Hiaki do not inflect for person or number.

suppletion based on the number of the subject or the object, such as *me'a*, there would need to be two diacritics for number, one for the subject and one for the object, as indicated in Figure 3.3. One issue for this kind of representation is that one or both of these sets of diacritics would always be redundant, especially because Hiaki does not inflect regular verbs - those that do not have suppletive forms - for person or number (the form of the regular verb *aache* ("laugh") is the same for all subjects and objects; Sánchez et al. (2017)).

Verb-object idioms provide evidence that the meaning of a syntactic unit can also be dependent on its morphosyntactic context. Examples such as those in (12) indicate that the meaning of verbs like "pass" (*pass a test*, *pass a law*), "take" (*take a photo*, *take a nap*), "get" (*get the idea*, *get the check*), and "kill" (*kill the music*, *kill time*) can be dependent on the semantic content of its object, while remaining indifferent to that of its subject. Although many architectures treat idioms as exceptions, these kinds of examples are very common, and are used in a variety of registers. The strong and systematic dependence of the verb's meaning on the object in these cases make them unlike simple cases of lexical ambiguity.

If the meaning of each verb was uniquely specified in the lexicon, with no context-dependent interpretations, we would not expect any of these verb-object idioms to emerge with these idiosyncratic meanings. It is not clear that the lemma model can explain this phenomena simply by stating that these verbs are ones that are semantically "light" or "bleached", or underspecified for meaning, because the intended meaning of each verb phrase is clear and specific. In these cases, and in many other cases not listed here, the meaning of the verb is determined by its morphosyntactic context.

One could interpret these cases as homophony, such that there would be multiple lemmas which are pronounced as "take" that correspond to different lexical concepts (one for *steal*, one for *photograph*, one for *sleep*, one for *ride*, and so on). However, on a homophony account, it would be a coincidence that all the lemmas pronounced as "take" have the same irregular past-tense form "took". One could also interpret these cases as polysemy, but this would require an additional mechanism in the conceptual domain to link very different concepts to the same lemma, which would be an issue if lemmas are meant to correspond to single lexical concepts.

Another possibility would be to treat these as idioms with a "hybrid" representation, as proposed by Cutting and Bock (1997) and Sprenger et al. (2006), where the superlemma or lexical-conceptual representation of the idiom "take a nap" would activate the lemmas for *take* and *nap*, so the idiosyncratic meaning would be associated not with the *take* lemma itself, but rather with the superlemma. This account would suggest - contrary to the lexicalist approach - that a single conceptual unit can be mapped to a syntactic complex, and not just to a single syntactic atom. Furthermore, this also suggests that stored linguistic representations can be syntactically complex, involving both morphological and syntactic structure. I argue that both of these are important steps in the right direction, though I discuss the advantages and disadvantages of "treelet-based" approaches of this type in more detail in Section 3.1.3.

To summarize, lemmas are a manifestation of both of the lexicalist assumptions discussed above: they codify a distinction between syntax and morphology, and establish themselves as a stored "triad" of form, syntax, and meaning. As a result, there is a large amount of data that the lemma will struggle to model, including (but not limited to) inflection and morphological structure, suppletion, and idioms, phenomena which are fairly widespread throughout human languages. These phenomena suggest that syntax and morphology need to be able to interact fully, not just by sharing a limited set of features, and that the form and meaning of a syntactic object is partially determined by the syntactic context, not just by the syntactic object itself.

3.1.2 Incrementality and lexical units

A central concern for models of language production, going back over a century, is incrementality: how much of the preverbal message and linguistic encoding is planned before the speaker starts talking? To what extent can planning and articulation processes be interleaved? If not all of it is planned in advance, how can speakers ensure that all the linguistic dependencies and word order requirements of the language are satisfied? Over the years, one common suggestion of highly incremental production models is that both preverbal and syntactic representations can be planned and updated in "lexically sized units", as proposed by Brown-Schmidt and Konopka (2015) and Dell et al. (2008). However, it is often not explicitly recognized how crucially these planning models thus depend on lexicalist assumptions about the units which are being incremented over. The reason is that an assumed one-to-one mapping from meaning to syntax to form makes it such that each increment of planning at one level can be matched by exactly one increment at the other levels. Without this assumption, there is no reason to think that the correct selection of a unit at the phonological level could be done by looking at a single unit at the meaning or syntax level, which is what maximal incrementality would require.

The cross-linguistic examples above that challenged the one-to-one mapping can be used to illustrate the parallel issues for lexically-based incremental production models. If a lexical unit corresponds to a single unit of meaning, then a fully incremental model would struggle to produce the two pieces of a Vietnamese idiomatic collocation in different, non-contiguous parts of a sentence. If the lexical unit corresponds to a single unit of syntax, then the two pieces of the idiomatic collocation would have to be separate units (as they are syntactically separable), and thus the incremental model would struggle to generate any pieces of the collocations that do not have independent interpretations, such as in (15b). If a lexical unit corresponded to the phonological word, that would suggest that a whole sentence in Inuktitut would be represented as a single lexical unit, again ignoring the productivity of morphology in polysynthetic languages. These incremental models would also struggle if the lexical units correspond to syntactic units but the meaning and form are determined solely by the lexical unit itself, for the same reasons discussed above for Hiaki verb suppletion and English verb-object idioms. For example, for the Hiaki verbs which exhibit suppletion based on the number of the object, such as *me'a*, a fully incremental model would retrieve the meaning and syntax for "kill" correctly, but could not correctly condition its phonological wordform on the number of the object because at the time that the verb was being produced, the following object would not have been planned yet. As will be discussed further in Section 4.2.3.3, the order in which phonological elements appear cannot directly correspond to the order in which the corresponding syntactic or meaning components were planned.

3.1.3 Treelet-based approaches - a step in the right direction

Many models of language production have taken steps to provide a more detailed account of the syntactic representation of lexical items, especially in regard to the separation of morphology and syntax in the representation of words. For example, Kempen and Hoenkamp (1987), Vosse and Kempen (2000), Ferreira (2013), and Matchin and Hickok (2020) (among others) propose models where the syntax of lexical items are represented as lexicalized "elementary trees" or "treelets". These models allow for the syntactic properties of a lexical item - such as argument structure - to be represented as syntactic structure, rather than a limited set of features or as sentence frames. Because the treelets are composed via syntactic rules, and then undergo a process of lexicalization in order to be stored as treelets, syntactic and lexical representations are thus not definitionally distinct, thereby rejecting the first lexicalist assumption, that syntax and morphology are separate systems that cannot interact. As long as the tree-based model assumes a syntactic theory which can accommodate the kinds of phenomena described above, it will be able to represent them as treelets. One could easily adopt a non-lexicalist theory of syntax, where even a single treelet could involve highly complex morphological structure, as is needed for Inuktitut and other polysynthetic languages, and for the structure of idioms, while still using the same basic operations and preserving the same architecture of the processing model.

However, these models are also clear examples of why it is insufficient to simply update the syntactic representations of the treelets without also reconsidering the criteria for lexicalization, and how the meaning and form of the resulting treelets are represented. These are all lexicalist approaches in that treelets correspond to stand-alone words or phrases, rather than pieces of syntax that are smaller than stand-alone words. Meaning and form are only specified for treelets, in a context-independent way, so the triad persists. Here, Inuktitut words pose the same kind of issue as they did for lemmas; a treelet would need to be stored in the lexicon for each possible stand-alone word in the language, some of which would constitute entire sentences. If these models were to argue that the treelets can be smaller than a stand-alone word in order to account for this data, then these models could not be considered fully lexicalist; however, they would still struggle to capture phenomena which are beyond the triad, because meaning and form would be specified for most treelets. Hiaki verbs that exhibit suppletion based on the number of the subject or object are still problematic, because there would need to be separate treelets for the same verb depending on the number feature of the subject or object. In addition, for treelet-based models which assume that the treelets are "atomic" in the sense that their sub-parts cannot participate directly in the syntactic structure outside of the treelet, they will struggle with Vietnamese idiomatic collocations and other similar phenomena as well.

More broadly, I agree that storage and retrieval of composed structures may turn out to be a central property of language processing, but they should not be defined in the lexical representations of words. The intuition captured by these treelets might be better understood not as a *representation* but perhaps as a byproduct of the *implementation* in a highly adaptable neural system. Furthermore, I see no reason that this property should be restricted to things at the "word" level; it should apply equally to phrases as well as sub-word pieces.

3.1.4 Language-specific optimization

So far in this section, I have argued against the claim that the system of language production requires lexical knowledge to be formatted in terms of lemmas or lexical units as an organizing principle. However, for things that do have a 1-to-1-to-1 mapping between meaning, syntax, and form (where a single syntactic object has a consistent meaning and form across a variety of contexts), it would be entirely plausible that lemmas - or something like them - could arise as a byproduct of language-specific optimization, where it would be faster or more efficient to represent meaning, syntax, and form in that way, even if it is not an architectural principle. In these cases, it is possible that the translations which are performed for that word can treat the word as if it were atomic (i.e., the calculation to determine the form for the word does not need to refer to any other elements in the syntactic context), as is suggested by the lemma model. This kind of symmetry might occur more often in some languages, so linguistic behavior may appear to be more "lemma-like" than it would for other languages. To be clear, this would be a consequence of optimization at the *implementation* level, rather than the *representation* or *algorithm* level (Marr, 2010); it should be the case that speakers of all languages have the same underlying mechanisms which can become specialized depending on the frequency and complexity of processes that are involved in the language, but that the lemma representation is not the default.

The possibility of "lemmatization" may not hold for every piece of syntax in a single language, even in English, but it is an interesting empirical question which is only made possible under a non-lexicalist approach - under what circumstances would a "lemma" be formed, if at all? It seems entirely plausible that a neural system which is searching to optimize and reduce resource use wherever possible would store frequently used linguistic objects in some way, and it is possible that something like a lemma could arise for some items in a language. A central commitment of lexicalist theories is that there is a principled divide between what kinds of representations can be stored in the lexicon and what has to be generated online. In contrast, non-lexicalist approaches that do not assume such a divide are free to predict that frequently generated relations *of any kind* could be stored, if this would facilitate future production operations. This could include commonly-used phrases (such as “kick the ball” or “walk the dog”), Multi-Word Expressions (as discussed by Sag et al., 2002; Bhattasali et al., 2019), or groups of words with high transition probabilities. The same considerations that apply to whether or not a complex word like “nationalize” is stored will also apply to whether or not a common phrase is stored. Depending on the properties of a particular language, storage of different sized pieces may optimize production, allowing wide variation cross-linguistically in the size of the stored pieces even if the underlying grammatical architecture is assumed to be the same.

3.2 Division between lexical and syntactic processes

As I touched on in the discussion of incrementality above, assumptions about processing algorithms are deeply intertwined with assumptions about the units of representation. In the case of language production, the lexicalist assumptions that characterized the lemma units led to models which made a fundamental division between the process of lexical selection and the process of syntactic structure building. Much of the same data discussed above presents a clear challenge for models that work this way. This means that moving to a non-lexicalist production model is not just a matter of updating the representation of stored linguistic knowledge.

In the Levelt and Indefrey (2000) model, the lexical concepts for the sentence are first selected, and then the corresponding lemmas are retrieved from memory. The syntactic structure is built incrementally as lemmas are retrieved, according to the syntactic frames of each lemma, and subsequent lemmas are inserted into the syntactic structure. For example, to produce the sentence “Maria kicked the ball”, the lemmas for *Maria*, *kick*, *the*, and *ball* would be retrieved. The verb *kick* has a syntactic frame which specifies its arguments and the thematic roles that they have in the sentence, so *Maria* would be inserted into the subject position because she is the agent, and *the ball* would be inserted into the object position because it is the patient. In this way, every lemma (except the first one which initiated the structure building) is inserted into a “slot” in the syntactic structure as it is being built. Once the syntactic structure has been built, the morphophonological code for each of the lemmas is retrieved, followed by phonetic processes and articulation.

However, idiomatic collocations in Vietnamese are difficult to explain in a model in which lemma selection and syntactic structure building are separate processes, because they demonstrate that syntax needs to operate across the boundary between syntax and morphology. Because each idiomatic collocation corresponds to a single lexical concept, it would be represented by a single lemma. In the Vietnamese sentence for “I do not want to study”, in example (15a), the *want* lemma would be retrieved, and because the *study* lemma is its complement, it would be inserted into the syntactic structure as a single lemma with two morphemes; the two pieces of the collocation would only appear adjacently. In order to have them appear separately, one would either have to argue that there is an additional step of post-insertion movement which allows the pieces to appear in separate positions in the structure, even though lemmas are treated as being syntactically atomic, or that the idiomatic collocation corresponds to two lemmas that are retrieved independently and inserted into their respective positions in the syntax - in which case, the idiosyncratic meaning could not arise without the involvement of an additional mechanism. Another possibility is that there are two lemmas for the same idiomatic collocation, one for when the two pieces are adjacent to each other, and one when they are not, each with a different syntactic frame to specify how the structure is built around it; again, this would not explain how both lemmas would get the same idiosyncratic meaning. None of these possibilities are available in the current lemma model.

Looking now at the the Consensus Model proposed by Ferreira and Slevc (2007), shown in Figure 3.4, which likewise operates over lemmas, the main difference in this model is that the process of lemma selection and morphophonological retrieval ("content processing") is done in isolation from the syntactic composition ("structure building"), as two separate subprocesses. As a consequence, the building of the syntactic structure is driven by conceptual properties and thematic function rather than the selectional restrictions of individual elements in the syntax. To produce the sentence "Maria kicked the ball", the message would first be encoded in terms of semantic meaning - the entities and concepts that are involved in the sentence - and relational meaning - how those entities relate to one another in the sentence, as agents and patients, and so on. On the "content" side, the lemmas for *Mary*, *kick*, and *ball* would be selected (function words and morphemes do not have their own lemmas), while on the "structure" side, the syntactic structure would be created for the sentence. When the morphophonological code for each lemma has been retrieved, they would be inserted into their position in the syntax (though the authors concede that the problem of how exactly those forms are inserted into the correct position does not currently have a solution).

This division between structure building and content processing poses several problems for the cross-linguistic phenomena reviewed here. Firstly, this model would have trouble generating

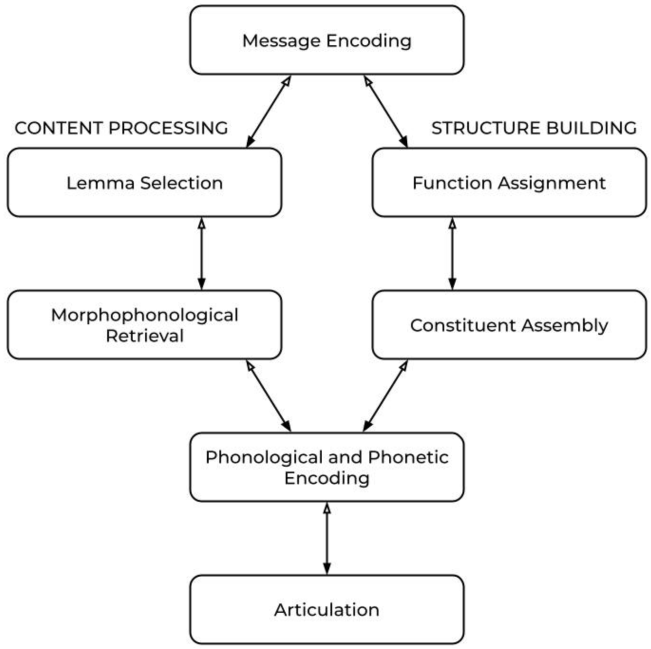

Figure 3.4: The Consensus Model of sentence production (Ferreira and Slevc, 2007)

Hiaki verb suppletion conditioned on the object, because morphophonological retrieval happens in isolation from constituent assembly; the "relational meaning" of the object (how the object relates to the other entities in the sentence, as the agent or patient of the verb) would be available, as would the conceptual representation of the object as singular or plural, but the syntactic structure and syntactic features would not be[3]. At the point of morphophonological retrieval, none of those features would be accessible to the *me'a* lemma.

The production of Hiaki suppletion could be accomplished if there are connections be-

[3]It is important to note here that this is about the syntactic feature of number, not the semantic feature. Something being semantically plural does not necessitate that it is syntactically plural, and vice versa. For example, "scissors" is syntactically plural, while being semantically singular, while "furniture" is syntactically singular while being semantically plural. In the cases where there is a mismatch between the syntactic and semantic features, agreement always occurs with the syntactic features, not the semantic ones. Furthermore, if there is a mismatch in conceptual number features but not syntactic number features, the sentence will be grammatical, even if it is semantically odd. Some verbs like "juggle" seem to require a plural object at a conceptual level (# *John juggled the task*), but the sentence is still syntactically well-formed (contrast with a sentence like "the furniture are in the living room", which involves agreement mismatch). As a consequence of this, it cannot be the case that the features necessary for agreement or argument structure are necessarily available at a purely conceptual level.

tween "lemma selection" and "structure building", and between "morphophonological retrieval" and "constitutent assembly", as is assumed in Eberhard et al. (2005). This framework allows for syntactic structure building to have an influence on a lemma's morphophonological form, assuming that there is a mechanism by which the features of the object lemma could be indirectly shared with the verb lemma. However, for both the Consensus model and the Eberhard et al. (2005) model, separation between structure and content (or the syntax and the lexicon) will cause problems in other cases where the lemmas would need to interact with the syntax beyond just sharing features, such as in the Vietnamese idiomatic collocations, where elements of the collocation can be syntactically separated.

In this discussion so far, a paradoxical problem seems to arise relating to the order of operations. In the discussion of the Levelt and Indefrey model, I argued that there will be issues if lemmas are inserted into a syntactic structure which was built *before* they were retrieved, in order to account for the production of Vietnamese idiomatic collocations. In the discussion of the Consensus Model, I argued that the syntactic structure should not be built *at the same time as* - but separately from - the lemma retrieval process, in order to account for the production of Hiaki verb suppletion conditioned on the plurality of the object, as well as instances where "lexical items" need to interact with syntax beyond sharing a limited set of features. It also should not be the case that the syntactic structure is built entirely *after* the lemma retrieval process, or there may be issues with verbal arguments not being satisfied[4]. Part of this problem stems from the ordering issue - at what point the lemmas are retrieved relative to the building of the syntactic structure - but also due to the commitment to the lemma as an atomic unit. These issues would

[4] If the syntactic structure is built only after the lemmas have been retrieved, and the speaker wants to use a verb such as *devour*, they may not have selected the lemma for the object even if one would be required, given that the syntactic requirements of the verb may not correspond to semantic or conceptual arguments. Because the model is serial, there would be no way to "go back" and retrieve the missing lemma.

not be resolved by adopting a tree-based approach, which uses syntactically-complex treelets, but assumes a similar model architecture. The non-lexicalist solution to this conundrum is that syntactic structure building and the retrieval and insertion of morphemes is a fully interactive process. There should be no stage at which the processes occur in isolation. Thus, rather than treating these as two separate processes, in the non-lexicalist approach we can treat them as a single unified process of syntactic structure building.

3.2.1 In summary

The evidence raised in this chapter, coming from a set of typologically diverse languages, demonstrates that the lexicalist approach is problematic not just in syntactic theory, but also for models of language production. Lemmas - and other things like them - encode lexicalist assumptions about the organization of the language system, either implicitly or explicitly, and the models which use them encode those assumptions in their algorithms. As a result, there are many phenomena that those models of language production will struggle to account for, not just in Inuktitut, Vietnamese, and Hiaki, but in languages like English and Dutch as well. The kind of model change that these considerations require cannot be satisfied by updating the terminology; the representations and algorithms involved in the model need to be fundamentally different, operating over different kinds of units and performing different calculations.

I have argued here that in order to achieve broader coverage, models of language should not assume a split between morphology and syntax, or assume that there are “lexical items” which function as triads of meaning, syntax, and form. This knowledge should instead be represented as mappings from meaning to syntactic atoms, and mappings from fully abstract syntactic atoms to form, without relying on lemmas or lemma-like representations. Models that adopt such a framework are better able to capture cross-linguistic data and generate clearer predictions for linguistic behavior in those languages. Within these models, there is space for language-specific optimization processes based on the reliability of mappings between different representations, and the number of licit possibilities for that mapping, which may vary across individual words or phrases and between different languages.

Chapter 4

WithOut Words, a non-lexicalist model of language production

Many models of language processing were built around the assumption that sentences are composed of “words”. However, as discussed in previous chapters, there are many linguistic phenomena across many languages which are not so clearly combinatorial, cases where semantic, syntactic, and phonological domains do not align, and where the boundary between things above and below the “word” level (traditionally the fields of syntax and morphology) is blurred. Several types of phenomena are particularly challenging to word- or lemma-based models of language production, summarized in (16) below. Models that assume that words or lemmas are the core representation of language production are either unable to account for these phenomena, or they would have to account for these data in ways that are highly undesirable and/or cognitively implausible.

(16) a. Cases where combinatorial or syntactic operations operate on units smaller than a “word”, as seen in polysynthetic languages that have highly productive morphological systems, and on units larger than a “word” as seen for phrasal idioms

b. Cases where morphology participates in phrasal syntax, as seen in Vietnamese idiomatic collocations (Noyer, 1998), Mandarin separable verbs (Yu et al., 2024), and German verbs with separable prefixes

c. Cases where meaning or form is dependent on a larger syntactic context, as seen in many languages, such as English verbs where the meaning is dependent on its object (*kill the music*, *crash the computer*, etc.), and in Hiaki verbs where the form is suppletive based on the number of its subject or its object (Harley, 2014)

d. Cases where the phonological word corresponds to a syntactic object which is not a constituent, such as the English "dyawanna" ("do you want to"), or as seen in grammaticalized "blended" signs in signed languages (Sandler, 1999), as well as cases where syntactic structure diverges from phonological structure (Selkirk, 1995)

e. etc.

In this chapter, I propose a non-lexicalist neurolinguistic model of language production, the "WithOut Words" (WOW) model, in order to better account for these kinds of phenomena. This model diverges from other current models in that it rejects semiotics (form-meaning pairings) as a fundamental principle, and that the format of the "lexicon" contains no meaning-syntax-form triads, but rather two independent mapping processes between meaning and syntax, and between syntax and form. In addition, The WOW model assumes that prosody is divided into two mechanisms - pre-syntactic and post-syntactic prosody - and that these mechanisms have a much larger role in linearization processes than assumed by previous models. Given that this model assumes a significantly different account of the representations and algorithms involved in language production, the WOW model also generates alternative analyses for the functional localization of linguistic mechanisms in the brain and acquired language disorders such as aphasia.

Importantly, in order to adopt a fully non-lexicalist approach in language processing, it is not enough to simply update existing word-based models with more nuanced accounts of morphosyntactic structure. Word- or lemma-based models encode lexicalist assumptions in their algorithms as well as in their representations, so even if one were to update the representations of the model in a non-lexicalist fashion, the model would not have the machinery required to accommodate cases of context-dependent form or meaning or interactions between morphosyntactic operations above and below the lexical level. By moving away from lexicalist assumptions, the WOW model provides better cross-linguistic coverage and aligns better with contemporary work in syntactic theory which has observed that syntactic and morphological processes cannot be distinct, that there are no good criteria to empirically define wordhood, and that representations of meaning and form do not always align.

4.1 Representational Assumptions

In the non-lexicalist approach adopted in this model, the basic units of processing are not "words", but rather separate, independent representations of meaning, syntax, and form (inclusive of speech, sign, and orthography). These representations are linked by two sets of context-dependent mappings, between meaning and syntax and between syntax and form, as shown in Figure 4.1. These mappings are not always "symmetrical", such that a single object in one representation does not necessarily correspond to a single object in the other representations. Consequently, what is often thought of as the lexicon would be distributed over these two sets of mappings, containing "lexical" representations. Processes which many people think of as lexical retrieval or lexical access are captured instead by these two mapping processes (meaning maps to syntax which maps to form) that are not constrained by the "word" level.

What does it mean for these representations of meaning, syntax, and form to be fully inde-

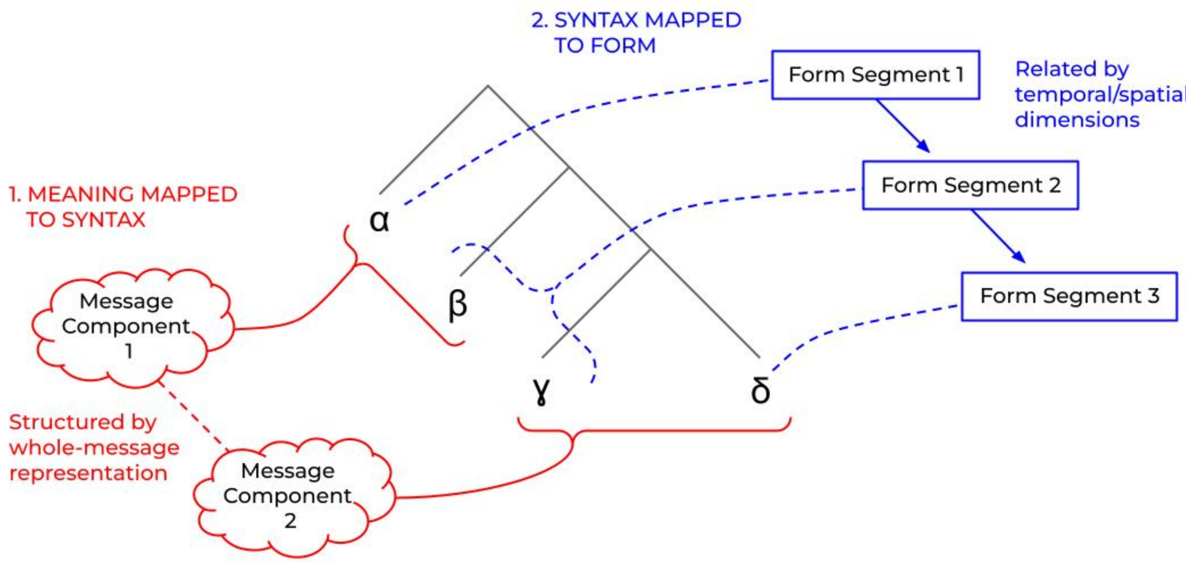

Figure 4.1: An illustration of the non-lexicalist approach, with separate mappings from (1) meaning to syntax and (2) syntax to form.

pendent? In this model, the representations are independent in the sense that each set of representations is stored separately, and that the operations performed for one set of representations do not directly inform the representations at the others. For example, syntactic rules do not depend on the meaning or the phonological form of the syntactic elements which they operate on, only their syntactic features (wh-questions do not undergo movement because they start with a /w/ or /h/ sound or because they have a question meaning, but because they have particular syntactic features which allow or require them to move in certain contexts). Furthermore, the properties of one representation do not have any direct bearing on the properties of the other representations, independent of the mapping processes that link specific representations. For example, grammatical category does not correspond to a single semantic or conceptual category, but proficient speakers of English know that the message like "kick" corresponds to a verb "kick"; grammatical gender in a language like Spanish is often realized as *-o* (masculine) and *-a* (feminine), but there are a variety of exceptions such as "mano" (hand) which are grammatically feminine despite having a form with a masculine-like ending.

As a consequence of the independence of representations, as well as the independence of the two mapping processes, a given syntactic object (a terminal, a set of terminals, or a structured arrangement of terminals) may have a mapping to meaning but is not required to (as has often been argued for the case of the expletive *it* in “it is raining”), and it may have a mapping to form but is not required to (as in the case of ellipsis, pro-drop, or null morphemes). Importantly, the two sets of mappings do not need to refer to the same syntactic object(s), where the meaning and form of a given syntactic object may be calculated based on the larger syntactic, pragmatic, or discourse context in which it appears. In addition, these mappings also do not have to apply over contiguous pieces, such as in Vietnamese idiomatic collocations and Mandarin separable verbs (there needs to be a theory about the constraints on non-contiguous mappings, so as to not over-generate beyond what is observed in natural languages; however, that is beyond the scope of the current work).

Thus, rather than representing *kick* as a stored complex which includes meaning, syntax, and form, as illustrated in Figure 4.2, the stored linguistic knowledge (which is traditionally thought to be stored in the lexicon) includes each representation stored individually, the relationship between the *kick* concept and the syntactic object [V kick], and the relationship between the syntactic object [V kick] and the form /kɪk/; notably, the meaning-syntax mappings and the syntax-form mappings do not need to apply over the same syntactic objects. This may seem like redundancy in the simple case, in that it provides essentially the same information but is represented over multiple “lexical entries” (per se); however, it also allows for much more flexibility. This is particularly important when one side of the mapping needs to be context-sensitive, such as in the idiom “kick the bucket”, where the meaning of the syntactic complex [VP kick the bucket]

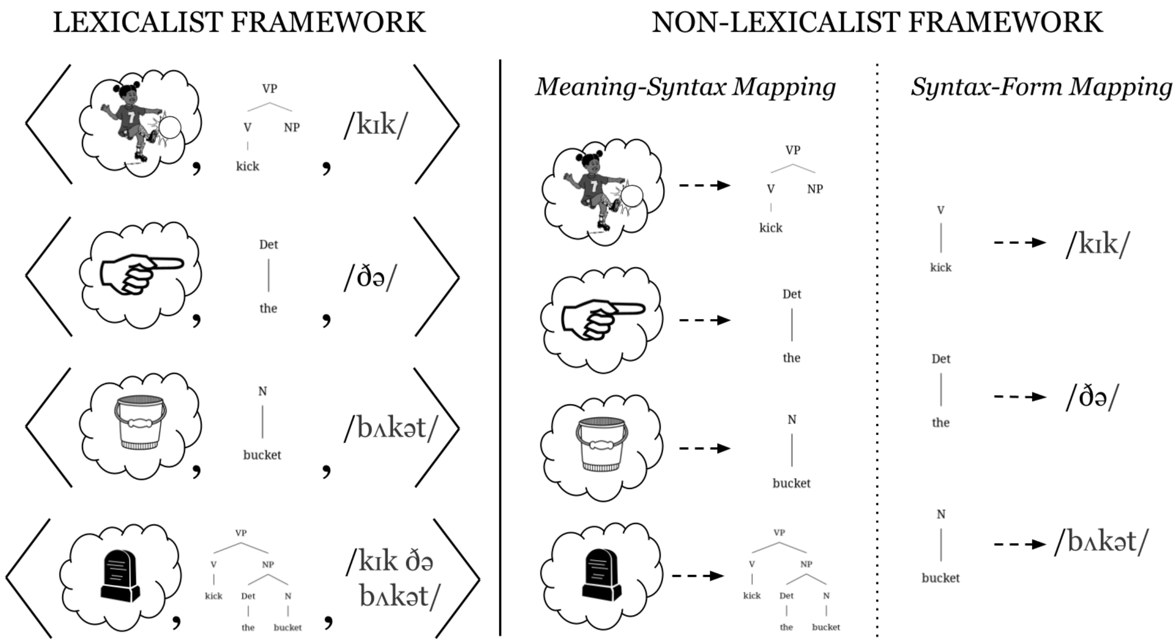

Figure 4.2: Representations a lexicalist vs. non-lexicalist framework. In the non-lexicalist framework, independent representations of meaning, syntax, and form are linked by mapping associations, rather than strict pairings or triplings.

can map to something other than the literal, compositional interpretation in the appropriate context. The systematic syntactic properties of the idiomatic and non-idiomatic readings of "kick the bucket" arise naturally as a product of the independence of syntax and meaning; the two phrases have the same syntactic elements in the same syntactic configuration, and those elements are mapped to the same set of phonological objects, but they happen to have a different relationship to meaning (dependent on discourse context or the larger syntactic context). This may appear to be more complex than word-based models of language production, but this move is critical in order to capture the kind of data that the lexicalist approach struggles to account for.

Another feature that this approach emphasizes has to do with the process of linearization, in translating the hierarchical syntactic structure into a linear string, as well as the active role that prosody plays in that process. It is worth mentioning here that - on most recent views - the hierarchical syntactic representation is not ordered; conventions in the syntax literature (and the 2-D domain of print literature) require some order to be imposed on the elements in the syntax, but the structure itself is more like a spinning mobile than a 2-D tree (Berwick and Chomsky, 2016; Idsardi and Raimy, 2013; Chomsky, 2007). In this way, the process of linearization is not as easy as just "going left to right" in the tree, but also requires a process of determining how to order the elements in the unordered structure so that it can be articulated as speech or sign, modalities that have physical limitations on how much can be articulated at a single time point. Beyond that, there are other processes that influence the order of elements in a sentence, such as Heavy-NP shift and other prosodic effects (Wagner, 2010; Anttila, 2016; Shih and Zuraw, 2017), as well as working memory resources and other domain-general mechanisms which are involved in language production (Martin and Slevc, 2014; Geranmayeh et al., 2014; Novick et al., 2010, a.o.).

This model assumes the role of prosody in linearization to be earlier - and more active - than what has been suggested by traditional, lemma-based models of language production. It has been observed that prosodic weight can influence many ordering phenomena, including the use of a double object construction or a prepositional dative construction in English ditransitive verbs (Anttila et al., 2010). As shown in the example below (where parentheses demarcate the prosodic phrases, and accents indicate lexical stress), the choice in construction is dependent on the prosodic weight or stress of the two NP objects. Pronouns like *her*, *him*, *us*, and *it* do not receive lexical stress, so pronoun themes are generally avoided in double object constructions. Sentences such as (17d) are ill-formed because one of the prosodic phrases contains only the unstressed pronoun *it*, while (17c) is acceptable because the lexically unstressed pronoun him is not alone in its prosodic phrase. It is important to note here that these constraints cannot be syntactic in nature, given that syntactic rules do not apply over phonological strings, but abstract syntactic units; no syntactic rule should be sensitive to the phonological features which are mapped to those syntactic units.

(17) a. Pat (gave Chris) (some fóod)

b. Pat (gave Chris) (fóod)

c. Pat (gave it) (to him)

d. * Pat (gave Chris) (it)

Furthermore, for verbs which tend to avoid double object constructions, such as *lower*, *mutter*, *donate*, or *return*, the double object construction is improved for lexically unstressed themes, as shown in the example below. Anttila et al. (2010) attributes this effect to a "stress clash" between the verb and the lexically stressed NPs (*John*, *Rachel*, *the charity*, and *the government* in the examples below).

(18) a. * I lowered **John** the box

b. Buddha lowered **him** the silver thread of a spider

c. * Susan muttered **Rachel** the news

d. She muttered **him** a hurried apology

e. * John donated **the charity** money

f. They can get the gullible ones to donate **them** money

g. * John returned **the government** the money

h. Judas returned **them** the money

In contrast to syntactically-driven effects, these judgements may be more gradient and sensitive to focus contrast and the prosodic contour of the utterance. The comparison of weight between two phrases is also relative; some people judge both sentences below to be well-formed and produce them themselves.

(19) a. ? Pat (gave me) (it)

b. Pat (gave it) (to me)

Based on these examples, it is necessary to conclude that prosody plays an important role in linearization at a relatively early point in the production process. Given that it is not particularly effortful to produce double object constructions, this should not be considered the product of "reanalysis", where a prosodically ill-formed utterance (e.g., stress clash) is generated, and when the error is detected the speaker starts the production process over from scratch with another syntactic structure, as would be required by the Levelt et al. (1999) model[1]. Prosody's effect on linearization needs to be an on-line process that can easily rearrange elements in the utterance to meet prosodic well-formedness constraints, rather than being a more peripheral mechanism as suggested by earlier models.

4.2 The model

The development of the WOW model involved identifying a new set of mechanisms used in language production in the absence of a "lemma" or "lexical" representation, as well as a reanalysis of the experimental and clinical evidence traditionally used to support the lexicalist

[1]In the Levelt et al. (1999) model, there is a "self-perception" loop that allows the speaker to monitor their own speech either before or after articulation, but it is parsed and passed to "conceptual preparation" much in the same way that one would when listening to another speak; there is not a mechanism which allows a produced utterance to be revised without starting the whole production process over.

models. Critically, the mechanisms of this model are not involved in “activating” specific stored representations, but instead in translating between different kinds of structured representations. Rather than having a mechanism for “syntactic structure building”, as is seen in many models of language production, this model has a mechanism which takes a structured “message” representation as its input, and outputs a structured syntactic representation, as an active calculation, with no separate process for mapping the “pieces” and generating the “whole”, such that there is a single integrated mapping process which deals with representations of all sizes. Models where brain areas are devoted to representations (rather than processes) reinforce a lexicalist notion that the relationships between the meaning, syntax, and form are trivial - that given a certain meaning, a single syntactic object will be deterministically activated, and given that syntactic object, a form will be deterministically activated. My approach gives much more credit to the kinds of transformations that need to happen to link between each kind of data structure (and consequently, the quantity of neural tissue devoted to those calculations is rationalized).

The WOW model is composed of seven mechanisms, described below, split into two sets (“Representation” and “Linearization”). The first set of mechanisms is responsible for generating and mapping between the structured representations of meaning, syntax, and form, predominantly through circuits in the left temporal lobe. The second set of mechanisms, localized predominantly in the left frontal lobe, exert influence on the translations between representations and work to organize those representations into a linear (temporal) order and generate a prosodic representation. Each mechanism is fully interactive with the ones that are connected to it, allowing feedback within each process. In some cases, the feedback loops are critical to on-line processing, while others are predominantly involved in language learning (such as the link between phonological representations and articulation).

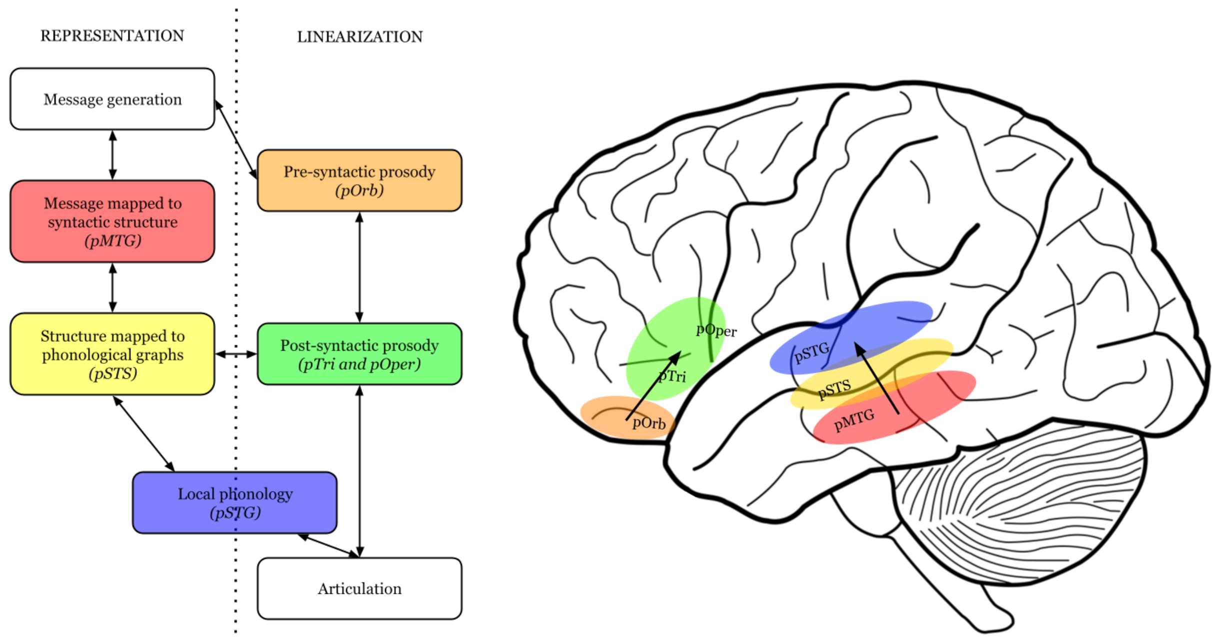

Figure 4.3: A diagram of the WOW model of language production

For each mechanism, tentative neural localizations have been provided, based on reanalyses of findings from previous literature through a non-lexicalist lens. Consequently, this model generates predictions for aphasia, a language disorder that arises after a brain injury - such as stroke, tumor, or head injury - that impacts parts of the brain associated with language processing. Many of the existing models of aphasia were built using lexicalist assumptions, so the predominant interpretation of the clinical data and the conclusions drawn from them may be incompatible with the non-lexicalist approach. The proposed model provides the means to reassess those findings with a non-lexicalist lens.

The goal of the WOW model is to create a unifying neural model of language production that brings together insights from a variety of fields and traditions. While it is committed to a non-lexicalist view of language, based on the growing consensus that appealing to a "word level" is insufficient, this model does not make a strong commitment about the combinatorial properties of the syntactic representation, and thus should be compatible with several different kinds of non-lexicalist syntactic traditions, including Distributed Morphology, Nanosyntax, the Exoskeletal Model, the Parallel Architecture, and perhaps some varieties of construction grammar, non-exhaustively. The differences between these theories may generate specific predictions in the context of this production model which would be empirically testable; this discussion is outside the scope of this paper, though it would constitute an important line of research into the representations and algorithms involved in syntactic structure building in the brain.

The sequence of events in this model – where a message is generated first, then the message is mapped to a syntax, and then syntax is mapped to a form – diverges from what is commonly assumed in (generative) competence theories of the grammar. In the Y-model in generative linguistics (Chomsky, 1993, 1982; Chomsky and Lasnik, 1993), it is assumed that the starting point of a sentence is the D(eep)-Structure which undergoes a series of transformations to create the S(urface)-Structure, at which point it can be spelled out in PF (Phonetic Form), and after other transformations, it can be spelled out to LF (Logical Form). As a theory of linguistic performance, however, it would not seem very plausible to assume that a speaker first generates a syntactic structure, then spells it out as form, and then decides what meaning to assign to that structure. The approach taken in the current model provides a more realistic implementation of language production, with the hope that the intuitions and judgements captured by the Y-model can be preserved as a competence theory of the grammar, represented in the kinds of relationships between elements within the syntactic structure as well as the mappings between syntax and meaning or form. What is traditionally thought of as the "grammar" in generative linguistics can be split into three kinds of knowledge in this model: pieces of syntax and the way that they can combine with other pieces of syntax, mapping rules that link pieces of syntax with meaning, and mapping rules that link pieces of syntax with form. The notion of "grammaticality," then, includes well-formedness in any of those areas, including syntactic well-formedness (where all of the pieces of the syntactic structure combine appropriately according to the constraints of the language), well-formedness of the meaning-syntax mapping (where all of the pieces of the syntactic structure that need to be interpretable are able to map to a meaning, and that meaning is well-formed according to any semantic/conceptual rules that may exist), and well-formedness of the syntax-form mapping (where all of the pieces of the syntax that need to be realized are able to map to a form, and that form is well-formed according to phonological and prosodic constraints). The grammar can also be thought to include structures within meaning representations and structures within form representations, but these may be qualitatively different from the structures of syntactic representations.

4.2.1 Temporal lobe structures ("Representation")

4.2.1.1 Message generation

In the first stage, many different sources of information and data types are consolidated into a message, which notably go beyond asserting facts about the state of the world. This message incorporates the conceptual representations for the entities involved in the utterance as well as their associated properties, event structure and thematic roles, information structure, and discourse and social context, among others. Given the constraints of the linguistic modality, the message cannot include all of the conceptual details that would be available to a speaker to represent a given situation, and not all conceptual details are worth articulating, as Bock and Ferreira (2014) discuss by distinguishing the "message" from a "notion". For example, if someone were to witness one child pushing another child on the swing, the message would need to make a commitment about what the event is - what the action is, which entities are performing the action, and which entities are the ones being acted upon. The same scene could be represented by any of the sentences below, but any one of those commitments would be sacrificing some degree of detail compared to the full conceptual representation that the speaker may have of a scene. In this way, the message acts as a bottleneck between the full conceptual representation of a scene and what is actually communicated through the utterance.

(20) a. Ellie pushed Carl on the swing.
  b. Carl asked Ellie to push him on the swing.
  c. The young boy was swinging.
  d. The kids are pushing each other on the swing.
  e. The kids are swinging in the backyard.
  f. ...

The message is generated with context in mind, including how to name the different entities or objects in the utterance. For example, in a journal of veterinary medicine, the term “canine” is more appropriate than “doggy”, but when talking to a young child, the opposite is true. This also applies to names and nicknames, where some people whose legal name is “William” may be referred to as “Bill” in some contexts. The choice of name an entity or object is dependent not only on the level of formality of the context, but also on the preferences of the person that is being labeled; knowing which “Williams” are known to go by “William” and which ones are known to go by “Bill” is a critical part of the contextual selection stage, allowing the language system to link multiple individuals to different (sometimes non-overlapping) sets of names[2]. The key claim

[2]The choice between “William” and “Bill” should not be thought of as an optionality in phonological processes,

here is that “William”, “Bill”, “canine”, and “dog” are each a separate message component, each mapping to a separate syntactic element, and that the choice between them is dependent on context and world knowledge[3].

The message generation stage also includes the decision of whether to say something in the first place, or to say something even if it conveys no new information (e.g., talking about the weather when both participants in the discourse are standing outside). The decision to say something or nothing can carry social weight, with its own social meanings and implications.

The message generation stage cannot function without some reference to the syntactic structure, because the types of features that are grammatically encoded in that language must also have some influence on the elements of the message which are prepared. For example, a variety of languages, including Turkish, grammatically encode evidentiality: whether the speaker personally witnessed the event, or if the information is second-hand (Slobin, 1996; Aikhenvald et al., 2003; Ünal and Papafragou, 2016). For the same event represented as an utterance in another language, the evidentiary status of the utterance may not need to be encoded at all. Speakers of English can encode evidentiality lexically, with adverbs or additional clauses (“They said that the dog was running”, or “I heard that the boy fell”), but these are not obligatory components of the grammar. For these speakers, the choice to include the information source may be a pragmatic one, whereas for speakers of Turkish, the evidential must be included because it is grammatically obligatory[4]. The decision of what needs to be included in the message is at least partly based on

unless one stipulates that all speakers of English have productive knowledge of the phonological rules for medieval rhyming nicknames, and that the phonological system is able to apply that rule in a way that is sensitive to the social context. Furthermore, it seems that “William” and “Bill” should be thought of as involving different syntactic objects; in the phrase “here’s a Bill, and there’s another *one*”, *one* can only refer to someone who goes by “Bill” and not someone who only goes by “William” (in contrast to other shortenings; “here’s a ‘roo, and there’s another *one*”, where *one* can equally refer to “‘roos” or “kangaroos”.)

[3]This is separate from sociolinguistic variation in pronunciation, e.g., “running” and “runnin’”, which can be accounted for by context-dependent applications of phonological processes.

[4]It is important to note here that grammatical differences between two languages should not be attributed to

the syntactic properties of the language, and not exclusively on the conceptual representation of the event or the communicative goals of the speaker.

For this reason, the elements that are required by the syntactic structure can be incorporated into the message through an interactive feedback loop between the message generation process and the subsequent process of mapping that message to syntactic structure. In a case such as Turkish evidentiality, a message that lacks that component in the message would not map to a complete or well-formed syntactic structure; that "error" signal would be passed back to the message generation mechanism, which would then supply that component of the message. This feedback loop would also be useful for components such as verb argument structure that requires a specific number of entities (which may not correspond transparently to the conceptual representation; for example, the verb "devour" requires an object while "eat" does not, a difference which cannot be attributed to their slight difference in meaning), so that a message that has an incorrect number of entities or relations for the selected verb could be revised in order to map to a well-formed syntactic structure.

Not all components of the message will necessarily be represented in the utterance in the same way. For example, in the utterance "you're wearing pajamas?", the truth conditional meaning is "the addressee is wearing pajamas", but there is an additional meaning conveyed by the prosodic contour, a rising tone to indicate surprise, disbelief, or judgment. The speaker has the option to communicate both aspects of the message non-prosodically (e.g., "I am shocked that you are wearing pajamas because that is inappropriate given the context"), but that may have different affective properties and could result in a different response from the listener (they may

---

differences in conceptual representations between individuals, i.e., it should not be argued that Turkish has evidential suffixes because Turkish speakers believe that that information is more necessary to communicate, compared to English speakers.

be offended, rather than gently corrected). I assume a broad definition of the "message" generated in this mechanism, one which includes all of the information that the speaker intends to communicate, and where not all of this information will be expressed syntactically.

Message generation thus involves consolidating those many different sources of information, each of which involves its own kind of computations and decisions, and formatting it in such a way that it can be represented in an utterance, and determining how each aspect of the message need to be distributed across prosodic, syntactic, or gestural representations to create the intended interpretation (both literally and socially). How those features are distributed depends on whether those features can be encoded grammatically or by other means; for example, formality is grammatically encoded in languages such as Korean and Japanese, but may be encoded as prosody or choice of phrasing in languages like English.

The message does not need to be fully generated in one step, and may be built incrementally, such that the speaker may begin subsequent processing stages and begin articulating an utterance before the message is fully "complete" (corresponding to a complete proposition). For example, if a speaker's ultimate goal is to produce a sentence like the (20a) above, they might initially generate the part of the message that contains the subject and the verb, "Ellie pushed. . . ", with the intention that more of the message would be compiled at a later point in time. The message could also correspond to an utterance that is larger than a single sentence, given that people are prone to producing run-on sentences despite them being prescriptively incorrect.

This stage has two primary interfaces. The first is with the mechanism described below, which maps the message to a syntactic structure. This interface provides a key feedback loop to generate a complete syntactic structure that aligns with the speaker's intended message and interpretation. The second interface is with the pre-syntactic prosody mechanism in the frontal lobe, which generates the correct prosodic contour for the components of the message that can be expressed prosodically, discussed in Section 4.2.2.1.

4.2.1.2 Message mapped to syntax

In the next mechanism, the message which was constructed and consolidated in the previous stage is mapped to a complex syntactic structure. This mapping process links the parts of the message that can be grammatically encoded to a grammatical representation. As discussed above, this process is fully interactive with the message generation stage in order to ensure that the syntax generation mechanism has access to all of the meaning information that the grammar of that language requires to be expressed. Features such as evidentiality or formality may not be central to the intended message, but if the syntactic structure lacks those features, it may be ungrammatical in that language. The building of the syntactic structure may not necessarily be completed in a single step, and may need to query the message level for missing information. Importantly, the derivation of the sentence does not "fail" if there are missing pieces of the syntactic structure, because they can still be provided through an interactive updating process.

A key feature of this model is that the pieces of syntax that are mapped to each unit of meaning can be as small as a single morpheme (e.g., the prefix "re-" in English), or as large as an entire syntactic complex (i.e., if the message is mapped to a negative polarity item such as "anymore", there will also need to be a licensor at another point in the structure, even if there is intervening syntactic content). There is no architectural constraint[5] on the size or structure of the pieces of syntax which can correspond to a single unit of meaning, and there is no distinction

[5] As previously mentioned, a grammatical theory of how these mapping processes may be constrained has not yet been fully detailed. Potential accounts could consider locality domains (e.g., to explain why the mapping from meaning to syntax does not link syntactic elements that are separated by multiple clauses), or perhaps learnability and processing constraints that operate with limited working memory resources.

between the kind of structure above and below the “word” level. Another key non-lexicalist feature of this mechanism is that there is not a separate stage before or after this one where “lexical items” are retrieved from memory independently from the syntactic structure building processes they participate in. In this model, the process of retrieving stored pieces of syntactic structure is integrated with the generation of novel syntactic structures, performed at the same time by the same mechanism, following the same set of syntactic principles. Consequently, there may be some optionality in determining whether to use a stored piece of syntax or whether a new piece of syntax has to be generated; this may depend on the degree of compositionality and transparency in the translation between meaning and form.

Every complex utterance has a syntactic representation, and must route through this mechanism. However, there may be shortcuts developed that link a meaning to a form directly. For example, in an object naming task, there could be a direct association between the conceptual representation of an object and the phonological representation that labels it. Furthermore, some types of idiom-like phrases may lack a syntactic representation altogether. For example, a monolingual English speaker can know that the phrase “che será será” means “what will be will be” and use it as such, but they may not have a grammatical analysis of it in the same way that a speaker of Italian would.

As a consequence of the architecture of the model, given that the mapping processes will always engage meaning and syntax or syntax and form, it is not possible to truly isolate “syntactic processing” from the representations that it interfaces with in terms of neural responses or functional localization; the mechanism is not “doing syntax” on its own, but rather transforming a complex message data structure into a complex syntactic data structure. This observation should not be taken to mean that the role of syntax in language processing is minimal, especially given the numerous examples of the combinatorial properties of language where there is no readily apparent explanation based on meaning or form. This observation is also not a contradiction to the main point about the independence of meaning and syntax; although they are separate representations in the mind, where we can descriptively identify phenomena that are purely syntactic in nature, a processing system which links representations together would make it such that those representations are not independently accessible within that algorithm (for further discussion, see Matchin, 2023; Călinescu et al., 2023; Fedorenko et al., 2020).

4.2.1.3 Syntax mapped to form

Once the syntactic structure has been built, the next challenge is how that structure can be mapped to segments of form in order to meet the constraints of the articulatory modality. The phonological word may correspond to a single syntactic unit (such as monomorphemic words in English), or it may correspond to a larger segment of the syntactic structure, even pieces that do not compose a single constituent. In English, contractions such as *I'll*, *she's*, *let's*, or *dyawanna* ("do you want to") involve a single phonological word that spans over a set of syntactic terminals that are not constituents. Similarly, in signed languages, multiple syntactic terminals - which can constitute their own signs in other contexts - can be represented as a single "blended" sign, as seen in Israeli Sign Language (Sandler, 1999).

Thus, the output of this mechanism is a set of phonological units which may not transparently reflect syntactic constituency. Various PF movement operations that occur in the interaction between syntax and phonology also happen during this stage (see Embick and Noyer (2001) for more details).

As discussed above, one puzzle that this mechanism helps to solve is related to order, given that syntactic structure does not inherently encode linear order. Because of that, a large component of this mapping process is determining which syntactic objects need to be linked to the same piece of form, given that there is no inherent order. For example, in English, tense can appear on the main verb, as in “John went to the store”, it could appear on the auxiliary verb, as in “John did go to the store”, or it could appear as an independent element, as in “John will go to the store”.

The form which is mapped to the structure may not be a string of phonemes, but rather a structured phonological representation or “directed graph” as discussed in Papillon (2020), Idsardi (2022), and Raimy (2000). Compared to strings, these graphs are simpler (requiring only a precedence relation, and containing fewer stipulations), and allow for much greater coverage of less string-like phenomena such as templatic and nonconcatenative morphology as seen in Semitic languages, circumfixation, morphologically-conditioned allomorphy, reduplication and other non-linear phonological operations, and more. The examples in Figure 4.4 demonstrate the basic operations of these directed graphs for a) the English plural “cats”, b) infixation in Atayal animate action focus, c) reduplication in the Indonesian plural, and d) Arabic templatic morphology, where the graphs on the left side of the arrow exhibit those morphophonological modifications, with serialization algorithm transforming them to the graphs on the right.

Furthermore, these graphs may provide a better representation of the different components of signed languages (hand shape and orientation, movement, location, and non-manual signs such as facial expressions) are coordinated in sequence, especially when components of one sign can often overlap in time with other signs.

The calculation of the phonological form is done in a context-dependent, probabilistic way,

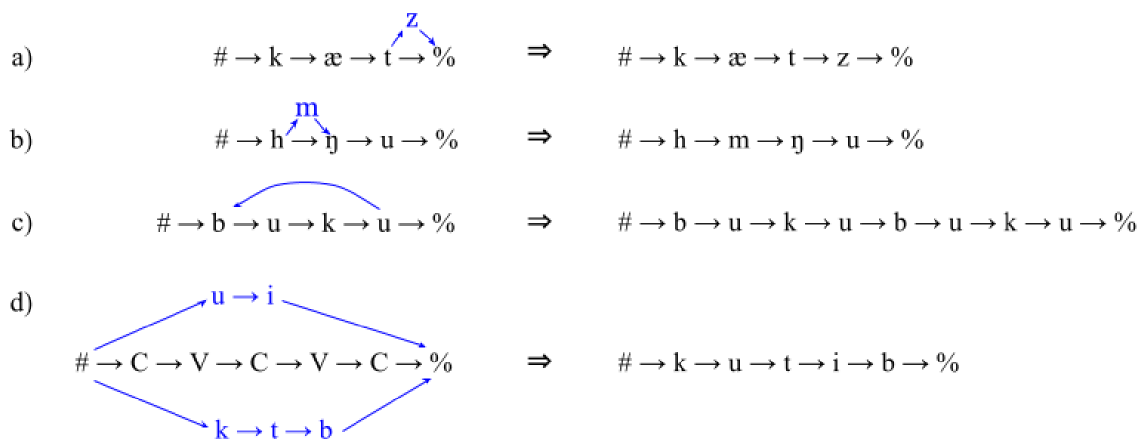

Figure 4.4: Directed graphs illustrating affixation, infixation, reduplication, and templatic morphology (shown in blue), before and after serialization.

where for a given syntactic object there may be multiple forms that are considered, but context and likelihood plays a large role in deciding between those alternatives. The interaction between this mechanism and the previous is also important as feedback, to ensure that the phonological form being linked to the given piece of syntax would be interpreted as having the intended meaning and affect; for example, using contractions in English would result in a different formality level and affect compared to using full phrases. Furthermore, if the syntax generates a representation that does not link to a viable piece of form, there is an opportunity to verify or revise the syntactic structure in such a way that it does.

4.2.1.4 Local phonology

The last "representation" mechanism moves the proto-utterance closer to a linear string. After the phonological graphs have been specified in relation to the syntactic structure and provided an order (in tandem with the "linearization" mechanisms in the frontal lobe), more local phonological constraints apply to the forms provided by the phonological graphs. In contrast to the previous stage which specifies the phonological forms in relation to syntactic objects, this mechanism applies phonological processes in the domain of the phonological word, syllable, and immediate phonological context. This would include phonotactic processes like assimilation (e.g., the plural morpheme in English is pronounced -/s/ as in "cats", or -/z/ as in "dogs", based on the voicing of final consonant) and vowel harmony (e.g., suffixes in Turkish match the front/back feature of the vowels in the stem, as in /jel-i/, wind.ACC, vs. /kol-u/, arm.ACC), syllabification and syllable-based constraints, and so on. This mechanism also acts as a buffer, holding the output in memory and releasing components of the form for articulation at the correct time.

4.2.1.5 Neural Implementation

Message generation

Given that this stage constitutes a number of complex sub-processes, and many types of representations which are considered, it is likely that the mechanisms serving this function are broadly distributed across a number of different brain regions corresponding to the various data types involved in the message. Given the diversity of these sub-processes, there may be several "hubs" which consolidate the aspects of the message which are represented as syntax, prosody, or gesture. How those different aspects of the message are sorted into these streams may result in different connectivity patterns or variations in localization, allowing for a great deal of language-specific procedures and optimization depending on how the different components of the message are represented in that language. The primary interfaces of this set of mechanisms are the "message mapped to syntax" mechanism, as well as pre-syntactic prosody.

Message mapped to syntax

This mechanism is predicted to be localized to the posterior middle temporal gyrus (pMTG), consistent with several neuroimaging studies of morphosyntactic complexity that implicate that brain area (e.g., Matar et al., 2021), as well as the finding that syntactic comprehension is impaired in paragrammatic aphasia, which is associated with injuries to that region (Matchin et al., 2023, 2020; Yeaton, 2023). In contrast to previous models, this model emphasizes that syntactic generation is crucially a mapping from meaning to syntax (or from syntax to meaning, in the context of a comprehension task). As such, this mechanism does not perform structure building in isolation during this mapping process, instead performing a translation between different complex data structures (the message and the syntactic structure).

Much of the literature investigating "syntax in the brain" has indicated Broca's area (BA44/45) as the primary localization of syntactic computations, but this literature exhibits a number of confounds that likely lead to misleading results, implicating a number of brain areas that may not be directly involved in syntactic processing. Firstly, many of these studies use stimuli that vary in syntactic complexity, such as in the example below; the first condition is a well-formed, coherent sentence, while the second condition uses the same words presented in a random order. The main issue with comparisons like these is that the more complex structure also has a more complex form (more components of the tree to be linearized) compared to the word list condition.

(21) a. The red boat floated in the sparkling bay

b. Boat sparkling the floated bay in red the

For this reason, it is not surprising that the inferior frontal gyrus (IFG) is implicated in many of these studies as being involved in syntactic processing, because it plays a critical role in linearizing syntactic structures and is a primary interface with this mechanism (see the discussion on "post-syntactic prosody" below). As a piece of syntax gets more complex, so too does the linearization process.

Syntax mapped to form

This mechanism can be localized to the posterior superior temporal sulcus (pSTS). This brain area has also been associated with social perception and a number of other cognitive mechanisms (audio-visual integration, recognition of facial expressions, and the perception of biological motion) (Hein and Knight, 2008; Liebenthal et al., 2014; Venezia et al., 2017, a.o.), which may also be factors that are involved in this mapping process. The selection of the correct output of the phonological graph based on the syntactic context is assisted by the post-syntactic prosody mechanism, supported by functional connectivity to the inferior frontal gyrus (IFG), the motivation for which is discussed below.

Local phonology

This mechanism can be localized to the posterior superior temporal gyrus (pSTG). The pSTG has been argued to play an important role in storing phonological categories, providing the phonological targets to be used in language production (Yi et al., 2019; Chang et al., 2010; Buchsbaum et al., 2001; Hickok and Poeppel, 2000), and in error correction during speech (Meekings and Scott, 2021), as well as functioning as a phonological buffer utilized in speech comprehension and verbal working memory (Acheson et al., 2011). The primary interfaces of this mechanism are the "syntax mapped to form" mechanisms, as well as the distributed array of motor mechanisms which control articulation.

4.2.2 Frontal lobe structures ("Linearization")

This second set of mechanisms is primarily involved in generating the linear representation of the utterance, and in coordinating the articulatory gestures that make up speech and sign. The heterogeneity of prosodic and linear phenomena has led me to split these operations into two mechanisms based on the kind of information that is relevant for the different phenomena. Thus, each mechanism operates over a different kind of structure, and is selectively sensitive or insensitive to different kinds of information generated by the temporal lobe structures (meaning, syntax, or form).

	Sensitive to message	Sensitive to syntax	Sensitive to form
Pre-syntactic prosody	✓	X	X
Post-syntactic prosody	X	✓	✓

As will be discussed in more detail below, the pre-syntactic prosody mechanism is sensitive to meaning in that, as a consequence of some aspect of the message, a prosodic representation is generated or a prosodic operation is applied (e.g., a rising tone at the end of the sentence to indicate a question or surprise; contrastive focus applied based on the speaker's theory of mind for other discourse participants). This mechanism is insensitive to syntax and form in that neither the syntactic representation nor the phonological representation of the utterance have an influence on how that set of prosodic representation is generated; the same prosodic contour is compatible with many syntactic structures.

The post-syntactic prosody (control) mechanism is sensitive to syntax in that linear prosodic operations are applied as a consequence of the syntactic structure (linking different parts of the structure together so that they can be mapped to the same segment of form), and is sensitive to form in that a prosodic operation or constraint is applied as a consequence of some aspect of the form (e.g., heavy NP shift; avoiding “‘light” pronouns as the second object in double object constructions, * *John gave his friendly neighbor it*, cf. *John gave it to his friendly neighbor*; Irish pronoun postposing (Elfner, 2011)), but this process has no bearing on the meaning of the utterance.

The third and final component of this second set of mechanisms is involved in articulation, and coordinating in time the many different gestures that make up speech and sign.

This section will not be able to address all of the literature on prosody, but it is important that the model include these mechanisms given that these are real components of natural language and relevant to language processing. The goal of these components of the model is not to provide a complete mechanistic account of prosody production, but rather provide a sketch of the kinds of components that will need to be included in a complete model of language production.

4.2.2.1 Pre-syntactic prosody

At the point when the message has been generated, the speaker will already know some things about the form that the utterance will take, even without knowing the exact syntactic structure or phonological form. The message will specify whether the utterance is a question or an assertion, as well as the social dynamics and discourse conditions involved. For example, in English, if an utterance is a question, it will often exhibit wh-movement (a syntactic phenomenon) or question prosody, which often involves rising pitch (a temporally-bound phenomenon). Information structure is often reflected in the syntactic and prosodic structure of the utterance, such as with fronting, focus and topic marking, and contrastive focus. Because the syntactic structure does not encode temporally-bound or linear properties, those components of the message cannot be mapped to a syntactic representation. Thus, in order for those components of the message to be realized, there must be an early process of encoding some linearized components separately from the syntactic structure being generated in the temporal lobe before or during the mapping between message and syntactic structure. The subsequent processes in the frontal "linearization" stream will sustain this representation and expand on it as more linearizable information becomes available.

A division between the "linear" components and the syntactic and phonological "representation" components is consistent with a variety of previous findings that suggest that prosodic representations are operate somewhat independently from other articulatory content during sentence production (Turk and Shattuck-Hufnagel, 2014; Pierrehumbert, 2000). Evidence from prosodic contours show that this mechanism is engaged early in the language production process, allowing a speaker to already know a great deal about the utterance even at the beginning of the utterance. Over the course of an utterance, a speaker's tone steadily drops (a process called "downstepping"), and it has been observed that speakers will modify the rate at which their tone declines, or start their utterance with a higher tone initially, depending on the length of their utterance (De Looze et al., 2015). Thus, the speakers may not know exactly what the articulatory content of the utterance will be, but they already seem to have a clear intuition about how long the utterance will be.

The pre-syntactic prosody mechanism's primary interface is with the message generation mechanism, in order to produce the prosodic contour that will have the intended affect and pragmatic interpretation. This mechanism receives information about the message, discourse, and general context, and then integrates those different types of information to determine which contour should be applied. Again, the key contrast between pre-syntactic and post-syntactic linearization operations is that pre-syntactic prosody is associated with a meaning, and makes no reference to the syntax or form of the elements; meanwhile, post-syntactic prosody refers to phonological properties such as lexical stress or prosodic weight, and may have an impact on the linear order of elements in the sentence, with reference to the syntax. The second interface of this mechanism is with the dorsal route of the dual motor speech coordination system proposed by Hickok et al. (2023), which articulates voice pitch.

4.2.2.2 Post-syntactic prosody (control)

I assume that the syntactic structure generated in previous representation stages does not yet include any kind of ordering information, but is just hierarchically-organized relations between syntactic units. Importantly, the linear order of the phonological elements of the utterance cannot always be determined by the syntactic structure alone, and sometimes prosodic or phonological information must be involved in order to determine it. For example, the comparative in English can be realized in two different ways, either morphologically, as a suffix *-er* (as in “smart-er”), or periphrastically, using *more* (as in “more intelligent”); the decision between the two forms is conditioned on the length of the adjective, as well as the interaction between its prosody and phrasal prosody (Adams, 2014), as illustrated in Figure 4.5. In Tagalog, a similar effect has been observed for the relative order of adjectives and nouns, where the choice between the noun-adjective and adjective-noun word orders is conditioned on phonological and morphological well-formedness of the resulting forms (Shih and Zuraw, 2017). Given that these phenomena also apply to novel

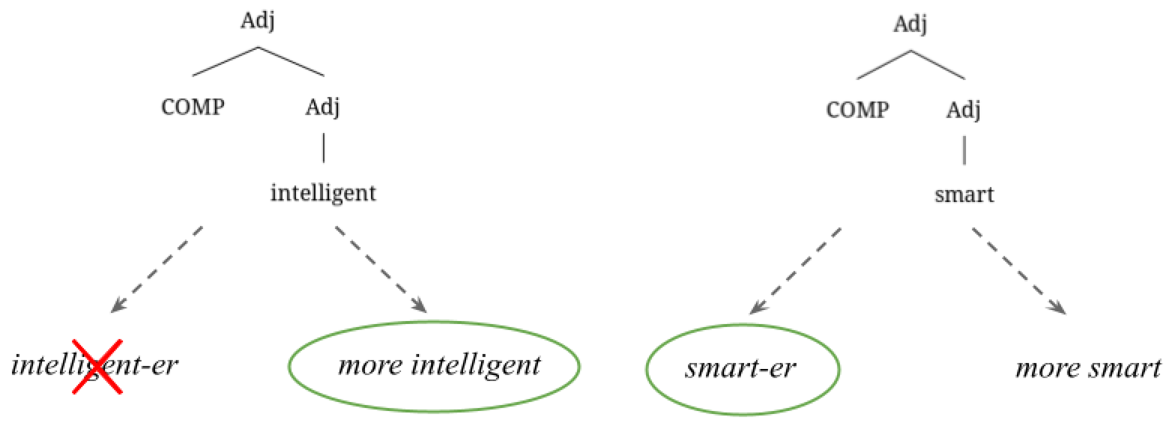

Figure 4.5: The syntax-form mapping process identifies two possible forms of the comparative, the morphological realization and the periphrastic realization; prosody acts as a filter to identify the form that is prosodically better-formed.

terms as well as familiar ones, the choice between different linearization alternatives should not be “lexically-specified”, or stored in the lexicon, but should be an active process of linearization. Therefore, I assume that the mapping of syntactic structure to phonological structure discussed above is additionally modulated by a post-syntactic prosody mechanism.

The post-syntactic prosody mechanism considers competing options for phonological realization of the chosen syntactic structure, and helps to identify the best combination of realizations to optimize the utterance based on prosodic and phonological constraints. This mechanism can be thought of as being related to “cognitive control” - the collection of domain-general processes that help people to complete goal-directed tasks by sustaining activation for the representations required for the task while inhibiting unrelated or distracting ones - but serves a domain-specific function here, holding a given linguistic representation in working memory and deciding between multiple alternatives for linearization. In this way, the post-syntactic prosody mechanism provides the additional attentional and decision-making resources that the syntax-form mapping mechanism needs to identify the correct set of phonological graphs for the given syntactic structure while inhibiting the others, helping to navigate a complicated translation space. This mechanism helps to resolve any issues or points of incompatibility, ensuring that every syntactic object is associated with a prosodically and phonologically appropriate form (which in some cases, may be null), and that every form is complete. Importantly, in contrast to Matchin and Hickok (2020), this mechanism does not store or calculate any form representations itself (it is not doing "linear morpho-syntax"), but only provides the control processes and prosodic constraints to assist the syntax-to-form mapping mechanism in linearizing the syntactic representation.

The process of mapping syntax to phonological graphs may be more complicated for languages that exhibit greater flexibility in terms of where different elements may appear in the utterance, and in cases when there are more elements to be linearized. For example, in English, verb tense morphology always appears on a verb, but it can appear at several different points in a sentence, as shown in Figure 22 below:

(22) a. Sarah walk**ed** the dog — *walk* + PAST

b. Sarah **did** walk the dog — *do* + PAST *walk*

c. **Did** Sarah walk the dog? — *do* + PAST . . . *walk*

In this example, the past tense morphology can appear on the main verb as in (22aa), but it can also appear with *do*, either before the main verb as in (22bb), or at the beginning of the sentence for a question as in (22cc). On top of that, the form of the past tense morphology is different depending on the verb that it attaches to. The placement and the form of the past tense morphology can be calculated in the translation between syntactic structure and linear form, but choosing the correct placement form for the past tense morpheme among competing alternatives may require additional control resources.

4.2.2.3 Articulation

At the final stage, the linearized sequence of articulatory gestures is sent off to various articulatory motor mechanisms, in order to be produced. This mechanism coordinates the activity of different muscle groups in order to produce fluent speech or sign.

4.2.2.4 Neural Implementation

Pre-syntactic prosody

I localize this mechanism to left fronto-basal areas, around the pars orbitalis (pOrb). This is consistent with the findings of previous studies which found engagement of the pOrb during semantic and emotional expression (Mayer et al., 2002; Belyk et al., 2017). It is also consistent with Binder et al. (2009), which observed activity around this area during semantic tasks; considering that this mechanism is sensitive to meaning, this activity would be expected. This brain area also has the functional connectivity with the middle temporal cortex to support this mechanism's primary interfaces (Xiang et al., 2010), given that this mechanism needs to consolidate multiple kinds of information from the message when generating the prosodic contour. In addition, its connectivity with dorsal motor regions is consistent with the Hickok et al. (2023) proposal of the dual speech coordination model, where the pOrb would generate the abstract representation of the prosodic contour, and the dorsal precentral speech area would support the corresponding articulatory processes related to voice pitch.

Though this model focuses on the left hemisphere language network, many studies implicate the right hemisphere in prosodic comprehension deficits and prosodic "flattening" in speech even in the absence of other linguistic deficits, and it may be involved in the production of affective prosody (Darby, 1993; Blonder et al., 1991). There are several components that may be influencing these observations; firstly, that the right hemisphere may have longer envelopes in auditory processing (Boemio et al., 2005), which would result in the comprehension of prosodic contours being somewhat right lateralized; secondly that mechanisms of social cognition and emotion which are critical in determining the prosodic contour are unlikely to be left-lateralized along with linguistic functions, suggesting that there may be some engagement of the right hemisphere in those processes. Importantly, this does not mean that the production of the two prosodic mechanisms in this model have to be right lateralized along with those functions, given that the prosodic mechanisms need to engage multiple kinds of processes and representations; however, it does leave open the possibility that knowledge of prosody associated with social cognition and emotion could be preserved or impaired independently of the processes which generate the prosodic contour.

Post-syntactic prosody

I localize this mechanism to the pars triangularis (pTri, BA45), and the pars opercularis (pOper, BA44). These brain areas have the functional connectivity that one would expect, with significant connectivity to the STS - the localization of the syntax-to-form mapping mechanism - as well as adjacent areas within the STG and MTG (Margulies and Petrides, 2013; Xiang et al., 2010). Although BA 44 and 45 are not completely homogeneous in function, and there may be a gradient of function within the two brain areas (based on connectivity patterns with different temporal lobe areas and their associated "representation" mechanisms in this model) the neuroimaging and clinical literature is not very informative to this point, in part because these two brain areas are often grouped together in analyses as "Broca's area" making it difficult to determine the extent of their independence.

In contrast to the Matchin and Hickok (2020) model, I assume that this brain area is not performing "linear morpho-syntax" itself, but instead supporting linearization by comparing the prosodic and phonological well-formedness of candidate forms generated by the mapping mechanisms localized in the temporal lobe. This is supported by case studies of agrammatic aphasia in Kalaallisut (West Greenlandic), a polysynthetic Eskimoan language, which observed that people with agrammatic aphasia produce the same number of morphemes as healthy controls but fewer (phonological) words (Nedergaard et al., 2020). In this language, both morphemes and words must be linearized, so the distinction would be surprising in an account where the IFG is applying a linear order to all hierarchical structure. However, this observation is not surprising in the WOW model; morpheme order in Kalaallisut is relatively fixed based on the syntactic configuration, so it can be supported by the syntax-form mapping mechanism in the temporal lobe, whereas word order is much more variable, requiring greater engagement of prosodic mechanisms in the frontal lobe in order to determine the ultimate linear order (Berge, 2016).

Articulation

As discussed by Matchin and Hickok (2020), these should be localized to the dorsal precentral cortex (DpreC), the ventral precentral cortex (VpreC), and the Sylvian parietal-temporal region (Spt) (Hickok and Poeppel, 2007; Rauschecker and Scott, 2009; Guenther and Hickok, 2015; Hickok, 2012; Hickok et al., 2023). Interactivity between this mechanism and the Local Phonology mechanism may be key for language learning, where a speaker attunes the phoneme targets in Local Phonology with what is actually produced.

4.2.3 Model consequences and predictions

Based on the architecture of this model, where the primary mechanisms are involved in the translations between different representations - and not in "activating" or composing those representations independent from other representations - the model predicts that processing costs should correspond not just to the number of representations involved or the complexity of their configurations, but also to the complexity of the mappings between the sets of representations at different levels. This cost corresponds to the effort required to perform the active, context-dependent translations as well as to sustain those representations in working memory. Because the model assumes that the meaning-syntax mapping and the syntax-form mapping are fully independent, the level of complexity may not be uniform across the course of utterance planning. For example, the meaning-syntax mapping may be very complex, involving a number of context-dependent elements, while the syntax-form mapping could be much more straightforward with little variability in form or context-dependence for each of the syntactic elements in the structure (the verb in a phrase like "he catches the bus" has an idiomatic meaning that does not straightforwardly map onto the "catch" syntactic object, but each syntactic element maps predictably onto phonological forms); conversely, the meaning-syntax mapping can be very simple while the syntax-form mapping can be highly complex (the phrase "the cat has caught a mouse" has a very straightforward relationship between meaning and syntax, but the realization of tense and aspect modifies the form of the verb and adds additional phonological elements to the utterance). The heterogeneity in complexity across the course of language production would lead to varying processing costs, with potentially different temporal or spatial dynamics in neural or behavioral responses. This cost should not be so great that it prevents people from producing utterances that have that kind of mapping complexity, but it should be identifiable in on-line processing measures. These costs are discussed further in Chapter 4, which tests this prediction using EEG to measure neural responses to mapping complexity during language production.

Similarly, the complexity of the syntactic representation will result in additional processing costs related to linearizing their associated phonological forms; the more elements there are that need to be put in order, and the more information that is required from the syntactic context to perform that linearization process, the more effort it will take to do so. Again, the linear order of the associated forms of elements in a structure is not something that is given based on the syntax alone, but must be actively calculated, based partly on the phonological or prosodic weight of the associated forms.

4.2.3.1 Optimization

An important feature of this model is cross-linguistic optimization, as previously discussed in Section 3.1.4. Given that each language has its own grammatical and phonological properties, the exact strategies that the speaker uses during language production can be optimized to navigate those particular processes more efficiently. For example, languages that use templatic morphology (such as Semitic languages) will have the same underlying mapping mechanisms as polysynthetic languages, but the exact calculations and strategies that the speakers of these languages use during language production become optimized to perform the specific morphophonological operations that their language requires. This optimization plays an important role in language acquisition. Becoming fluent in a language (either first or second language acquisition) is not just about learning how the independent representations of meaning, syntax, and form map to one another, but also about developing strategies to optimize those processes for the target language variety.

This will also result in differences in how languages engage the different mapping mechanisms during language production. Some languages may exhibit less variability or optionality in form or linear order than others; as a result, speakers of those languages may be faster in performing the mapping from syntax to form (and the process may be less effortful) because the syntactic structure serves as a more reliable cue for the utterance's form or order, compared to speakers of languages that exhibit more variability or optionality. In languages where the syntactic structure is a less reliable cue for linear order, other factors will need to be considered when determining the linear order, and more computational effort may be required. Languages will also vary in the number of obligatory grammatical features that they have (e.g., gender, evidentiality, case, etc.), which may result in differences in the time and effort required to prepare the message and map it to a syntactic structure, particularly in cases where the grammatical feature encodes some component of the message (e.g., evidentiality, tense, and aspect are grammatical features that encode a component of the message, whereas grammatical gender does not always correspond to reality).

4.2.3.2 "Lexical retrieval"

Commonly used experimental constructs such as "lexical retrieval" should be reevaluated under the non-lexicalist framework, given that there are no "lexical items" which can be retrieved, and that "words" or "lexical items" can be syntactically complex in the same way as phrases (in some accounts, even a simple noun like "cat" involves at least a root morpheme and a categorizing morpheme). The process of lexical retrieval is often assumed to function like picking up a ring of keys, where if any key is picked up, the rest of the keys will come along with it automatically; if the semantic representation of a lexical item is activated, its syntactic and form representations will be close behind it. In this model, there is no "key ring" to link the different representations together, but instead active, context-dependent mapping processes. Thus, tasks that involve object naming or giving single-"word" responses should engage the same mapping mechanisms as tasks that involve sentence production, but perhaps with lower levels of complexity, given that the semantic, syntactic, and phonological context is much smaller than in a larger phrase or sentence. There is some evidence for such an account, as suggested in Gaston et al. (2023), where auditory word recognition was found to be qualitatively different in continuous speech and in isolation, suggesting that the two tasks engage language processes differently, or engage the same processes but to different degrees across tasks. It should be emphasized that models built around single-word tasks cannot constitute a complete explanation of linguistic behavior; even though some aspects of language processing can be done in isolated, "word-by-word" paradigms, it does not capture the full range of linguistic phenomena that occur in natural, continuous speech.

4.2.3.3 Incrementality

Incremental processing is an area of the psycholinguistics literature which has gained significant attention in recent years, which analyzes linguistic processes in terms of "word-by-word" prediction and interpretation (e.g., Brown-Schmidt and Konopka, 2015; Iwasaki, 2011; Dell et al., 2008; Brown-Schmidt and Konopka, 2008; Brown-Schmidt and Tanenhaus, 2006; Griffin, 2001; Schriefers et al., 1998). Some of the discussion on this topic is sensitive to the idea that there are different levels of representation that may have different planning windows, but those independent representations can often be conflated when much of the discussion focuses on "words" or "lexical items" as the unit which is incremented over, with the accompanying assumption that those words correspond to a single representation at levels of meaning, syntax, and form.

As observed by Momma (2016), strict incrementality in production is not conducive to languages like Japanese, where the verb arrives at the end of the sentence. If the verb was not planned up until the moment before it was uttered (as suggested by Brown-Schmidt and Konopka, 2015; Iwasaki, 2011; Dell et al., 2008; Brown-Schmidt and Konopka, 2008; Brown-Schmidt and Tanenhaus, 2006; Griffin, 2001; Schriefers et al., 1998), the message that the speaker initially intended to convey with the verb may not be compatible with the sentence that has already been uttered, especially in regard to verbal argument structure (e.g., if there are too many or too few arguments). In the case of long-distance dependencies, there must be a way to maintain those dependencies in memory and produce the appropriate pieces (or empty spaces, in the case of wh-movement) when needed. More generally, it seems that there should not be any point in a dependency at which only one part of the dependency has been generated; for wh-movement, it seems unwise to suggest that the wh-question has been generated first, but the gap which it is meant to fill has not been generated at all (especially given the syntactic constraints on wh-movement). As another example, in Georgian, negative pronominals (in English, these include *no one*, *nowhere*, or *nothing*) can appear anywhere in the sentence when the negative marker *ver* is present, as in the first example below, but when there is no negative marker, the negative pronominal can only appear immediately before the verb, as shown in the second two examples (King, 1996):

(23) a. **versad** šeni Cigni **ver** vnaxe
nowhere your book NEG-VER 1.see.3
'I couldn't see your book anywhere'

b. šeni Cigni **versad** vnaxe
your book nowhere-VER 1.see.3
'I couldn't see your book anywhere'

c. * **versad** šeni Cigni vnaxe

(cf. [versad] šeni Cigni ver vnaxe)

A strictly incremental production model would struggle to explain how such sentences would be produced reliably without error. If the negative pronominal has been generated, but the negative marker has not yet been generated, then the negative pronominal could only appear directly before the verb; sentences like the first one would never arise in the incremental model, because at the point where *versad* is produced, it is not licensed by the negation.

Furthermore, a variety of phenomena suggest that speakers already have a clear intuition about how long their utterance will be, even if they do not yet know what its articulatory content will be. As discussed previously, "downstepping" in prosodic contours is modulated by the length of the utterance (De Looze et al., 2015). In addition, phrasal stress patterns in Bengali are aligned with the right edge of prosodic phrases (Hayes and Lahiri, 1991), and given that proximity to stress can modulate processes like lenition, articulation rate, and reaching of gestural targets (Gordon, 2011), a speaker would need to know how many syllables a phrase has - and whether those syllables are stressable - in order to determine how to pronounce each sound. These phenomena would be especially difficult to account for in a model where the utterance is *only* planned piece by piece.

Utterances are not always planned in their entirety, and they can be planned and produced in segments, but the important aspect of this in the non-lexicalist model is that the preparation of each type of representation may not be happening with the same time course. For example, the preparation of the message representation may not align with syntactic constituency, so that a single component of the message could be mapped to non-adjacent segments of the syntactic structure, as in the case of long-distance dependencies; the preparation of the form may also not align with the message representations or the syntactic representations that have been prepared, in cases where a larger syntactic context is required in order to determine which form corresponds to the syntax that has already been built, as in the case of verb suppletion (i.e., Hiaki verb suppletion that is conditioned by the number of the object of the sentence). And again, because linear order is not solely determined by the syntactic structure, but also by prosodic and phonological constraints, there may be times when a form or linear order cannot be correctly produced without referring to a larger segment of the syntactic structure. The key prediction of the WOW model in regard to incrementality is that in experimental paradigms such as priming, one could expect to see different time courses for semantic, syntactic, and phonological priming for a single element in the utterance depending on what components of the utterance have been planned; except at the phonological level, the time course by which that information becomes available is unlikely to correspond exactly to "word" order.

4.3 Predictions for aphasia

The WOW model asks for a deep re-thinking of the format of linguistic representations and the kinds of processes which operate over them. Consequently, this model has significant implications for how we think about language disorders, given that the aphasia literature is deeply steeped in a lexicalist tradition that takes the grammar-lexicon distinction for granted. Framing language disorders around "syntactic deficits" or "lexical deficits" implicitly makes a particular commitment to how those representations are stored and operated on in the brain - that there are "word" or "lexical" representations that can be selectively impaired, and that there are syntactic rules by which those representations are combined which can also be selectively impaired (Martínez-Ferreiro et al., 2020; Friedmann et al., 2013; Hagoort, 1993; Caramazza and Zurif, 1976; Goodglass and Baker, 1976, a.o.). In this non-lexicalist theory, where the primary operations involved in language processing are active mapping processes between independent representations, a brain injury would affect the tissue which performs the active mapping processes that translate between the different data structures, rather than the specific representations themselves. Reinterpreting clinical data in a non-lexicalist lens via the WOW model - where meaning, syntax, and form are fully independent data structures, and where linear order is not fully determined by the syntax - leads to a variety of predictions.

One of the main principles of the WOW model which differs from standard models of language deficits is that the representations of meaning, syntax, and form are fully independent, and as a consequence of this, the mechanisms within the language system that map between those representations can function autonomously even if the inputs to each mechanism are not fully well-formed or complete. A useful metaphor for this is an assembly line in a factory that makes and decorates cake. The assembly line has three steps: making the batter from raw ingredients and pouring it into molds (message generation), one that bakes the batter and stacks the layers of cake (mapping the message to syntactic structure), and one that decorates the outside of the cake with fondant and frosting (mapping syntactic structure to phonology). If the machine that bakes and stacks the layers of cake is broken, it might under- or over-bake the layers, stack the layers incorrectly, or damage the layers along the way. However, once the cake gets to the frosting machine, the frosting will make it look like a (mostly) beautiful cake even if the structure of the cake is faulty. Because each machine in the assembly line is independent, one machine breaking down does not impact how the other machines (the other mapping processes) perform their function.

In this way, appearances can be deceiving. An utterance that seems to exhibit issues with syntactic structure - incorrect or missing verb inflections, agreement morphology, or complex phrases - may not be lacking in syntax at all, but may instead be the result of an issue in the mapping from syntax to form that happens to disproportionately impact those types of forms. An utterance that seems to be syntactically well-formed - with all of the indicators of “syntax” - may not have been fully generated by the syntax, but the utterance was rescued by a form-mapping mechanism that chose forms that were highly likely within the given phonological context, which happened to be correct. These kinds of scenarios are a natural consequence of the independence of meaning, syntax and form representations. Even if there is an error early in the process of language production, there is nothing stopping the rest of the process from working properly - or at least attempting to, if those mechanisms have any degree of input. By moving away from the “triad”, just knowing that a phrase or phonological word was correctly produced may not be indicative that its meaning and syntax were also correctly generated, only that a form was produced. For that reason, testing theories of aphasia may require more careful thought about what other processes may be at work beyond the one mechanism which is impaired, and how they might hide the real deficits.

Another principle of the WOW model is that the linear order is not fully determined by the syntactic structure, and that the order of the form segments within an utterance is an active calculation based on phonological and prosodic constraints in tandem with the syntactic structure. This area of language processing has been generally overlooked by the aphasia literature, where the linear order of the "lexical items" in a sentence is often implicitly assumed to be completely determined by the syntactic structure (with a few exceptions, e.g., Matchin and Hickok, 2020; Boeckx et al., 2014). By allowing linearization processes to take on a more active role in language production, this model also predicts that there should be selective deficits in these particular linearization operations. Thus, deficits in language production may not only be related to meanings, syntactic structures, forms, and the mapping processes that link those separate representations, but they can also be related to the processes of generating a prosodic structure or calculating and applying the linear order of the phonological elements in the utterance.

The following section is not intended to be an exhaustive account of all of the predictions that the WOW model could make for language disorders. Rather, it illustrates how the key principles of the WOW model can lead to very different conclusions from the standard models of aphasia, or how those existing accounts can be aligned with a non-lexicalist perspective.

4.3.1 Impairment in the message-syntax mapping mechanism

An impairment in this mechanism would result in utterances that might sound fluent and seem to be conceptually well-formed, but involve errors in syntactic structure, as described for paragrammatic aphasia (Matchin et al., 2020, 2023). In this situation, all of the pre-syntactic operations are functioning well so the message itself may be well formed, but its mapping to syntax exhibits some errors. For example, the message may be mapped to the wrong pieces of syntax, or there may be difficulties selecting all of the required pieces of syntax, or different parts of the message may be mapped to incompatible pieces of syntax. However, in a seemingly contradictory way, the utterance which is ultimately produced may appear to be mostly well-formed (at least on the surface), simply because the post-syntactic operations are functioning well. The subsequent mechanisms which map the syntactic structure to a phonological form may use transition probabilities and "default" forms to supply missing pieces that were not provided by the syntactic structure, making it seem as if there are fewer errors in the syntactic structure than there actually were. The language system may be able to produce something that appears phonologically well-formed, even if large pieces of the syntactic structure are missing or incorrect, and even if it does not correspond to the message that the speaker intended.

In a lexicalist perspective, it would be impossible for people with a deficit in the link between meanings and syntactic objects (lemmas or lexical representations) to produce anything at all, given that language production relies on those lexical representations to activate the necessary pieces that drive the subsequent stages of production. However, in the non-lexicalist perspective of the WOW model, it should be entirely possible for an utterance to be produced even if the intended message of the utterance was not mapped to a syntactic structure, as long as the other mechanisms involved in generating a form are functioning. The resulting utterance will seem to be well-formed in many ways, but will lack a coherent meaning or syntactic structure, much like what has been described as "word salad" for people with Wernicke's (fluent) aphasia. In those cases, the speakers were able to generate a message at least enough to trigger the generation of a prosodic spine (in pre-syntactic prosody), but that message failed to map to a syntactic structure. However, because the prosodic spine was generated and sustained by the "linearization" mechanisms in the frontal lobe, the interaction between those mechanisms and spared "representation" mechanisms caused phonological representations to be inserted into that prosodic spine. The resulting utterance will be mostly well-formed, with the caveat that there may be minimal evidence of a complex syntactic structure; whatever indications of syntax that are present in their speech would likely be phonological in nature, with inflected forms that were highly likely within the given phonological context (even if they are not correct in the syntactic context). Several interesting empirical questions arise from this: what kinds of phonological representations are produced in fluent aphasia, and why not others? Do those forms arise because of neural "noise" that causes any random form to be produced, or are there other factors (perhaps inputs from other mechanisms of the language system) which modulate which forms are produced?

4.3.2 Impairment in the syntax-form mapping mechanism

Impairments to this mechanism may result in more serious difficulties with producing speech or providing the phonological form of a syntactic object, as a failure to retrieve a correct or complete phonological graph for a given segment of the syntactic structure. There may also be some issues with phonological operations such as generating inflected forms, infixation, reduplication, or interpolating templates and roots in languages like Semitic, if the mechanism is unable to successfully generate or navigate the different streams within the phonological graph, even in the absence of deficits in mapping meaning to syntax. This profile is consistent with what has been observed for many individuals with Broca's aphasia, arising after damage to both the STG and the pars opercularis (Fridriksson et al., 2015).

Given that the calculation of form is also highly dependent on the calculation of order, and vice versa, it may be difficult at times to disentangle deficits in the syntax-form mapping mechanism from deficits in the post-syntactic prosody mechanism. However, where a deficit in the post-syntactic prosody mechanism would mainly impact forms that exhibit variability or optionality, a deficit in the syntax-form mapping mechanism would result in impaired access to any stored form representation, even in the absence of order or form variation. For example, in English, nouns tend to have a more "static" phonological representation, where the syntactic object [N cat] is associated with only two forms, either [kæt] or [kæts], where the plural contains the entirety of the unmarked form and the plural is always marked on the end; in cases such as these, the syntax is a clear indicator of form and order, and thus may not require the involvement of post-syntactic prosody to assist with the mapping. A deficit in the post-syntactic prosody mechanism may not result in any errors with such forms, whereas a deficit in the syntax-form mapping mechanism will exhibit errors in all forms. This is also consistent with the findings of Fridriksson et al. (2015), which observed that people with injuries to the pars opercularis without injuries to the STG did not have Broca's aphasia.

Furthermore, given that there are few mechanisms which follow this one that could rescue a poorly-formed utterance, there would be very limited opportunities for that utterance to be repaired. Since local phonology and articulation are both mechanisms which take existing segments of form as their inputs, they would also have limited means by which to repair an utterance that is missing a great deal of its form representation or that exhibits many errors.

4.3.3 Impairment in the pre-syntactic prosody mechanism

In contrast to much of the previous literature on prosodic deficits, the WOW model makes a distinction between pre-syntactic and post-syntactic prosody. Because the model makes such a distinction, with pre-syntactic and post-syntactic prosody mechanisms operating over different kinds of data structures, we should expect to see a dissociation between deficits in these two types of prosody.

The existing literature on prosody has focused on emotional or "linguistic" prosody, both of which would be categorized by the WOW model as types of pre-syntactic prosody. These two types of prosody involve different kinds of information which is considered in the message generation stage, but ultimately result in similar consequences for form - a prosodic contour. As this literature discusses, an impairment to pre-syntactic prosody results in difficulties producing natural sentence-level prosodic contours and phenomena like contrastive focus (Walker et al., 2009; Danly and Shapiro, 1982, a.o.). However, given that prosody can serve as a "spine" for the utterance, a deficit in pre-syntactic prosody may result in larger issues with language production caused by an inability to align the linguistic content generated in the temporal lobe with the rhythmic and tonal contour that is generated by the pre-syntactic prosody mechanism.

Conversely, it also should be possible for speech of people with aphasia to have the illusion of a pre-syntactic prosody deficit, but not actually struggle to generate a prosodic contour. As discussed above with the cake metaphor, there are number of things that could disrupt a mechanism of language production that operate later in the process; in this case, a prosodic contour could be correctly generated, but the elements that should fit into that prosodic structure fail to align with that contour.This may be the case for types of aphasia that exhibit "flattened" prosodic contours in combination with other issues of linear order or form, but where they may still be able to produce "content-less" prosodic contours, such as the "huh?" contour used in English and a variety of other languages, which can sometimes stand alone without other articulatory content. This could also be true of people with Wernicke's aphasia, which is associated with abnormal prosody around clause boundaries (Danly et al., 1983).

4.3.4 Impairment in the post-syntactic prosody mechanism

Non-fluent aphasia is the type of aphasia that often arises after lesions to the left IFG. It is also sometimes referred to as "agrammatic" aphasia, because it is characterized by "telegraphic speech" that seems to lack function words and inflectional morphology. Many of the observed deficits in non-fluent aphasia associated with inflectional morphology can be understood not as a deficit in the representations of those morphemes, but instead in the post-syntactic prosodic control mechanisms which support the linear placement and pronunciation of those morphemes. Languages with less flexible morpheme ordering may not involve such a complex process, as the number of plausible linearizations for those morphemes is reduced. Languages like Kalaallisut (West Greenlandic), a polysynthetic language, have a generally fixed morpheme order, with less variable forms; therefore, less cognitive control should be required to linearize the morphemes within a single phonological word. It has been observed that speakers of Kalaallisut with non-fluent aphasia do not exhibit the usual pattern of deficits for functional morphology, and are able to produce the rich inflections of Kalaallisut words with a high degree of accuracy (Nedergaard et al., 2020). While morpheme order is generally fixed in Kalaallisut, word order is not; speakers of Kalaallisut with non-fluent aphasia do tend to produce fewer phonological words in a single utterance, even while the phonological words themselves are well-formed. This post-syntactic prosody mechanism, therefore, can contribute to varying degrees cross-linguistically, depending on the range of different linearization options available to a given structure.

Another commonly observed deficit in agrammatic aphasia is related to verbs, where they produce fewer verbs, and struggle with verb comprehension and action naming (Cho-Reyes and Thompson, 2012). This can also be understood as an issue of linearization, especially involving verbal argument structure and the linear ordering of the elements in the sentence, which can be specified for individual verbs. In addition to the challenges with tense morphology discussed earlier, this can also involve decisions about the order of the objects of the sentence. For example, in the sentence, "John gave some flowers to Susan", the elements can be arranged in a number of ways:

(24) a. John gave some flowers to Susan

b. John gave Susan some flowers

c. Susan was given some flowers by John

d. Flowers were given to Susan by John

Even if the decision between the different constructions can be motivated by different factors (information structure, discourse, etc.) these are all possible ways that the elements in a sentence might be ordered. The influence of prosody and control in producing one of these four sentences - while inhibiting the others - is not trivial (Novick et al., 2010; Hwang and Kaiser, 2014) (cf. Ferreira, 1996).

4.4 Comparison to other models

As a non-lexicalist model, the WOW model provides a clear, specific alternative to lexicalist models which assume that language is organized around "lemma" or "lexical" representations, which store meaning, syntax, and form, and where operations above and below the "word" level are fundamentally separate. In addition, the WOW model is also meaningfully different from other models that have been characterized as "non-lexicalist", both in that meaning, syntax, and form are fully independent representations that are not stored in a single, consolidated lexicon, and in that the "word" level has no significance in the model. The discussion in this section is not meant to be an exhaustive account of all the models and approaches to language production in the literature, but provides a summary of the key distinctions between the WOW model and some popular accounts of language processing.

4.4.1 Lexicalist models

Lexicalist models have received a variety of criticism in recent years (e.g. Bruening, 2018; Jackendoff, 2017; Goldberg, 1995; Halle and Marantz, 1993). The key concerns presented in that paper are that "lemmas" or other "lexical" representations encode a separation between "lexical" and "syntactic" processes, and that it stores meaning, syntax, and form together as a "triad" that cannot be separated. By separating lexical and syntactic processes in these models, the lexicalist account generates a "mechanism ordering paradox", in which there seems to be no possible order for the two mechanisms to work in without encountering a fatal issue. For example, in Levelt and Indefrey (2000), syntactic structure is built incrementally, where the first lemma projects a structure, and each subsequent lemma is inserted into that existing structure; in this case, it is impossible for something which seems to have a single meaning (e.g., idioms, separable verbs in Mandarin, idiomatic collocations in Vietnamese, separable prefixes in German and Dutch, and other similar phenomena) to correspond to non-adjacent linear or syntactic positions, because the lexical items cannot participate in the syntactic structure beyond the point at which they are inserted. Meanwhile, in the "Consensus Model" (Ferreira and Slevc, 2007), there is a distinction between "content" and "function" mechanisms, where lexical processes (including the selection of form) occur in a separate but simultaneous stream from syntactic processes; in this case, it is impossible for a lexical item to have a form that is dependent on its syntactic context (e.g., verb suppletion, especially based on syntactic features that do not always align with meaning), because the form is determined in a separate stream from the mechanism which determines the syntactic relationships between elements. Finally, it would seem impossible for syntactic processes to only follow lexical processes, or there would be many situations where the syntactic requirements of those lexical items (e.g., argument structure) would not be satisfied unless it aligns with the message of the utterance (which is not guaranteed[6]). Because this model does not rely on a lemma or anything similar to a lemma, the model is able to avoid many of those pitfalls.

Crucially, The WOW model differs from neuroanatomical models that posit a distinct brain region or neural mechanism associated with "lexical nodes". In the WOW model, no such distinction is possible; there is no dedicated stage for lexical processes independent of syntactic processes, given that there are no lexical representations. For example, Wilson et al. (2018) proposes that the dorsal lip of the STS is associated with lexical nodes which connect with other components of the lexical items, but that this region is spatially and functionally distinct from "higher level syntax". The study observed that the dorsal lip of the STS responded to both backward speech and scrambled writing, and that the response was seemingly equivalent in both visual (written) and auditory (spoken) modalities. Based on this observation, they concluded that this modality-independent response in the absence of linguistic content or structure suggested the activation of lexical nodes. I favor other explanations; for example, this effect could be the product of modality-independent phonological processing, which is well-known to occur during reading as well as speech. Similarly, Lau et al. (2008) identifies the posterior middle temporal cortex

[6]Argument structure can be idiosyncratic; for example, the difference in argument structure between "eat" and "devour" (where "devour" requires an object, while "eat" does not) cannot be easily accounted for by their semantic/conceptual differences.

as storing lexico-semantic representations, based on its role in generating the N400 effect; these results could also support several alternative accounts, however, where the effect is not exclusive to lexical items, and where it corresponds to semantic, syntactic, or phonological prediction (or "pre-activation") without requiring that those representations be perfectly aligned at the "lexical" level. As another example, the Dual Stream model (Hickok and Poeppel, 2007) assumes separate mechanisms for a "lexical interface", localized to the pMTG, and the "combinatorial network", localized to the aMTG. Again, much of the neuroimaging evidence in support of a "lexical" area could also be compatible with an account where meaning, syntax, and form are independent representations, and where the mapping between meaning and syntax is supported by cortical networks in the MTG, as proposed in the WOW model.

The WOW model also differs from models that assume a "modern lexicalist" approach, which rely on "lexicalized" treelets rather than lemmas (Kempen and Huijbers, 1983; Vosse and Kempen, 2000; Ferreira, 2013; Matchin and Hickok, 2020). The treelets can have a high degree of morphosyntactic complexity, but undergo a process of lexicalization which stores those treelets with a fixed form and meaning, creating a triad. This kind of model assumes that symmetrical mappings (where the meaning-syntax mapping is mirrored by the syntax-form mappings, e.g., for a given piece of syntax that maps to one piece of meaning, it also corresponds to one piece of form) are a fundamental property of the language system. In a non-lexicalist approach, this is not a necessary or central property of the language system. Furthermore, the lexicalized treelets will struggle to capture some kinds of linguistic phenomena; even if the treelets are allowed to be inserted into other treelets, they will encounter issues when there are elements that intertwine, such as seen in Vietnamese idiomatic collocations, where a verbal idiomatic collocation can be interpolated with a nominal idiomatic collocation (Noyer, 1998, ex. 10-11).

The MUC (Memory, Unification, Control) model (Hagoort, 2016, 2013; Baggio and Hagoort, 2011, a.o.) is lexicalist in a similar way, where it assumes that the mental lexicon (memory) is composed of words in a syntactic frame, and as such the syntax is not independent from the lexical elements that are composed via unification processes. These treelets are not morphologically complex, and rely on the assumption that elements stored in memory would include both form and meaning (although meanings can be context-dependent based on the discourse context, to account for homophones like "bank"). By organizing the mental lexicon in this way, the model assumes that such properties are fundamental to the linguistic system. As such, the MUC model would struggle with the same phenomena that the modern lexicalist approaches do, in addition to cases of morphological complexity and places where morphology and syntax interact. Furthermore, as previously discussed, the evidence used to support the claim that the LIFG is the primary locus of combinatorial operations exhibits a number of confounding variables, and those effects could also be attributed instead to the complex mapping or linearization operations that correlate with syntactic or "combinatorial" complexity.

Finally, this model also helps to resolve some of the issues with linearization by allowing other mechanisms such as prosody to influence sentence production at several different stages. As discussed earlier, many of these models do not consider prosody until very late in the language production process, if they consider prosody at all. Especially given the incremental nature of these models - where the fundamental assumption is that these processes increment over "word-sized" things - it would be impossible to account for the role of prosody or phonology in the order of elements in double object constructions, the English comparative, and the Tagalog adjective-noun order.

4.4.2 Other non-lexicalist models

A variety of models have been developed with the aim of rejecting some of the principles of lexicalism, but with different approaches. The key ones discussed here are the Parallel Architecture model (PA; Jackendoff, 1997, 2002, 2007), Construction Grammar (CxG; Goldberg and Suttle, 2010), Elman's lexicon-free account (Elman, 2011), as well as recent neurolinguistic studies which have argued against syntactic representations.

The WOW model shares many of the same premises as the Parallel Architecture model; there is no lexicon-grammar distinction, and meaning, syntax, and form are independent generative components that are linked by interfaces. However, there are several key differences between the two, mainly in the assumptions about the format and content of the lexicon, where the WOW model is more radical in terms of how different representations are associated; there is no point at which things are ever stored as a triad. In PA, lexical items *can* consist of triads or schemas, with pieces of meaning, syntax, and form linked by coindexing. Based on many of the discussions in that line of work, it seems that PA would argue that a lexical item is created for every unique triad of meaning, syntax, and form. As such, regular (rule-based) morphology is also stored as an independent object in the lexicon, while the same syntactic feature realized as an irregular morpheme is listed separately (Jackendoff, 2007, ex. 18). Even if not all lexical items are fully-specified triads (namely, syntactic rules that leave some variables left unspecified), PA makes these triads available, which - in my opinion - is inconsistent with the tenet of the independence of the representations. If form and syntax are meant to be fully independent, irregular morphology is not morphosyntactically irregular, only irregular in its form; having regular and irregular morphology listed separately in the lexicon contradicts that tenet. In addition, it creates a great deal of redundancy in the lexicon, where PA would require a separate lexical entry for every different form of the plural, each specified with the same representations of syntax and meaning. In WOW, there is no need to separately list the meaning and syntax of the plural with every variation in form, since those elements only need to be listed once, and the decision to map between them is context-dependent. In WOW, misalignments of meaning, syntax, and form are not "exceptions to the norm", but a natural and expected consequence of the independent representations and two independent sets of mappings that operate over them.

One issue that PA and CxG share, but WOW resolves, is the representation of idioms and their available non-idiomatic readings. In PA and CxG, idioms must be stored as whole pieces (Jackendoff, 2007, ex. 19), separate from their non-idiomatic counterparts. As a consequence, both PA and CxG must assume that "kick the bucket" (idiom) and "kick the bucket" (compositional phrase) are functionally independent, and the similarities in phrasal structure and form are purely coincidental. In contrast, in the WOW model, the systematicity between the two syntax-form pairings falls out naturally from the fact that they have the same set of syntactic objects, but they happen to have different mappings to meaning depending on the discourse context in which they occur.

Construction Grammar (CxG) and the WOW model also share some properties, namely that there is no grammar-lexicon distinction. However, CxG takes semiotics (pairings between meaning and form) to be a fundamental principle; the alignment between meaning and form is necessary, and whatever else remains must be fully compositional or fully stored. In contrast, WOW allows for asymmetries in the mappings between meaning and syntax and syntax and form, which gives it much more flexibility and less redundancy. Furthermore, CxG assumes that linear order is purely a property of form, with no possibility for prosody or phonological weight to be involved in re-ordering the elements of the utterance; any variations in linear order (e.g., the double object construction vs. the prepositional dative) would be assumed to be a different construction altogether, and any similarities between them is purely coincidental.

On another branch, Elman (2011) argues for a connectionist model of "lexical knowledge without a lexicon". The key principle of this model is that every lexical item stored in long-term memory also has much more nuanced information about frequency and context than standard "dictionary-like" models of the lexicon. In principle, a connectionist model of language production is generally compatible with what the WOW model argues for, where the mapping processes between different data structures can be thought of in connectionist terms. In that sense, the WOW model can also encode much of the information that Elman includes, particularly in its context- and frequency-sensitive mapping processes. However, Elman still orients the discussion of the lexicon around "words", which is incompatible with the non-lexicalist approach that the WOW model pursues.

Another sense of the term "lexicalist" has been used in an area of the literature which debates between "lexicalist" accounts - where structure building is driven by lexical items - and "abstract structural" accounts - where structure building is driven by abstract frames that lexical items are inserted into (Takashima et al., 2020; Rafferty et al., 2023, a.o.). This notion of lexicalist and non-lexicalist approaches is still slightly different from the one taken by the WOW model, mainly because it still assumes that there are such things as lexical items, and that the main point of debate is whether syntactic structure is projected from lexical items, or whether lexical items are just inserted into an independent syntactic structure. The WOW model would assume that neither is correct; in this approach, there is no coherent notion of a "lexical item", and thus syntax cannot be projected from them, nor can syntax be independent of them. Ultimately, even the “abstract structural” account still enforces a division between lexical items and syntax, and establishes lexical items as atomic units that do not participate directly in the syntactic structure.

Conclusion

The assumption that morphosyntactic “words” are at the core of language processing is deeply ingrained in linguistics and adjacent fields. Even if many of those people are aware of the issues with “words” (that they do not constitute a definable natural class, that representations of meaning, syntax, and form are often misaligned, etc.), but they continue to use them out of convenience, orienting models around “word” or “lexical” processing gives the impression that such a representation exists in the mind, and that it contains multiple pieces of stored information that are inextricably linked, and that those pieces of information are not independent at all. By holding onto terms like “lexical access”, or “word-by-word processing”, researchers will continue to fall into the habit of making lexicalist assumptions, regardless of what they believe about syntactic theory.

The WOW model presents a new opportunity for understanding language production in a “world without words”. This model offers an alternative to current models, providing a clear account of the kinds of processes and calculations that are required to produce speech or sign in the absence of a “lexical” representation which links together separate representations of meaning, syntax, and form. This model also gives us a new perspective by which to understand language disorders, which could lead to better clinical interventions for people who live with those disorders.

Chapter 5

Neural responses to mapping complexity during language production

As discussed in earlier chapters, lexicalist models of language production struggle to account for a variety of linguistic phenomena, even in languages like English. In particular, these models struggle to account for "non-compositional" phenomena, where there are misalignments between meaning, syntax, and form, and cases where the form or meaning of a syntactic object is context-dependent. In contrast, according to the WOW model, language production does not just involve combinatorial operations such as syntactic structure building, but also active, context-dependent mapping processes that translate between different structured representations during language production, allowing it to better handle such cases. Crucially, those mappings can vary in complexity depending on the number and arrangement of the representations that are mapped together, as well as the degree to which context is relevant in that mapping. While much prior research has focused on the cognitive and neural processes for accessing "words" and combining them, this chapter looks at a relatively under-investigated question: how do stored complexes get mapped across levels of representation, and how does complexity in those mapping processes modulate neural activity during language production? I used EEG to study neural activity while participants produce a variety of sentences with similar syntactic structures but varying degrees of complexity in the mapping between meaning and syntax, or in the mapping between syntax and form, in order to investigate the extent to which mapping complexity influences neural responses independent of syntactic structure. In particular, this experiment focuses on cases where a meaning is expressed not by a dedicated verb, but rather the use of a verb in a particular syntactic context, and cases where the linear order of phonological elements depends not just on their syntactic configuration, but also on their phonological content, as in the realization of tense and aspect in English. The results provide key insights into the non-compositional aspects of language, providing a broader view of the kinds of processes involved in producing speech or sign.

5.1 Verb meanings and non-compositional phenomena

The first things that most linguists and psycholinguists think of when we talk about non-compositional aspects of language are phrasal idioms like "kick the bucket" or "let the cat out of the bag". However, non-compositional phenomena - where the meaning or form of a single syntactic terminal is not solely determined based on its identity - are pervasive in natural language, in phrases that seem much more commonplace. For example, a number of verbs in English seem to have different meanings depending on their object, as in example (25):

(25) a. *Take the money* (steal X), *take a shower* (perform X task), *take a bus* (travel via X), *take a photo* (create X)

b. *Do a task* (perform X), *do the dishes* (complete X chore), *do the paperwork* (fill out X)

c. *Get the door* (open X), *get the idea* (understand X), *get the bill* (pay for X)

This phenomenon is not just reserved to verbs which are considered semantically 'light', as shown in (26):

(26) a. Pass: *pass a street corner*, *pass the exam*, *pass the law*, *pass a kidney stone*

b. Run: *run a race*, *run a (computer) program*, *run an organization*

c. Kill: *kill the houseplant*, *kill the bottle*, *kill the mood*, *kill the music*

d. Catch: *catch a ball*, *catch a bus*, *catch the flu*, *catch one's breath*

In all of these cases, the object of the verb can be interpreted fully transparently, while the verb itself varies in meaning. To contrast these with phrasal idioms like "kick the bucket", where no element of the phrase is interpreted literally, I will refer to these kinds of phrases as "verbal idioms", because only the verb exhibits this flexibility in meaning. In these phrases, multiple meanings can be associated with a single verb, depending on the syntactic context that the verb appears in.

In a lemma-based view, where the "word" or the "lexical item" is the minimal unit of language processing (Kempen and Huijbers, 1983; Levelt, 1989; Bock, 1995; Levelt et al., 1999; Dell et al., 2008, among others), language production involves retrieving the intended meanings in the utterance, and each meaning activates its unique "lexical" or "lemma" representation, which then subsequently activates its unique form. These cases show, however, that "word" representations should not be uniquely activated by a single "lexical concept". As discussed in Chapter 2.3 and Chapter 3.1.1, describing these as cases of homophony - where there are multiple lemmas for "take", each associated with a different lexical concept - is less than desirable, given that all of those independent lemmas would also need to store the same irregular past tense forms, when lemmas cannot refer to other lemmas to determine their form. The only way for lemma-based models to account for these cases is to treat them as exceptional, subject to different processing mechanisms and constraints than the compositional elements. These verbal idioms would be thought of as cases where each word in the phrase has its own meaning that is being overridden by an idiomatic interpretation; for example, in order to produce the phrase “catch the bus”, the conceptual representations for “use as transport” and “bus” would activate one set of lemmas, (e.g., “ride” and “bus”), which would then be inhibited by another mechanism in order for the “catch” lemma to become active.

Another common account of non-compositional phrases in the lemma tradition is that there is an additional “idiom lemma” which, when activated by the appropriate lexical concept, activates the lemmas of its constituents (in this account, a lexical concept for “using a bus as transport” would activate the idiom lemma “take the bus”, which would then activate the constituent lemmas, “take”, “the”, and “bus” which can be composed into the associated syntactic structure) (e.g., Cutting and Bock, 1997). However, such accounts face several problems. One issue with treating verbal idioms like “take a nap” in the same way as phrasal idioms like “hit the sack” is that the object of a verbal idiom can always be transparently interpreted (“taking a nap” involves “a nap”, while “hitting the sack” does not involve “a sack”). The fact that “take a nap” has a non-trivial relationship with the noun “nap” (much like the compositional phrase “eat the apple” has a non-trivial relationship with “apple”) is treated as a coincidence on the idiom lemma account, which assumes there is no relationship between the lexical concept and the lemmas that the idiom lemma activates. Another difference between verbal idioms and phrasal idioms is that verbal idioms do not require the object to be overt in order to get the same meaning. For example, the phrasal idiom “hit the sack” loses its idiomatic meaning when the object undergoes proform substitution, as in (27a) below, whereas the verbal idiom does not lost its idiomatic meaning when it undergoes *one*-substitution[1], as in (27b):

(27) a. # John needed to hit the sack$_i$, and so he hit ***it***$_i$ when he got home.

b. ✓ John needed to take a nap$_i$, and so he took ***one***$_i$ when he got home.

Lexicalist models have to treat these kinds of instances as exceptions to the standard format of lexical representations. However, these "exceptions" are more frequent and more systematic than one might expect if human language knowledge were really organized that way (recall the previous discussion on suppletion and verbal idioms in Chapter 2.3). Based on these kind of data, an alternative approach is necessary, one which relies less on "words" as the unit of processing.

The WOW model allows much more flexibility in the mapping relationships between meaning, syntax, and form. In contrast to traditional accounts which rely on words as "triads" of meaning, syntax, and form, non-lexicalist accounts allow for more complex mappings and context-dependent mappings to account for cases like the verbs in (25) and (26), as shown in Figure 5.1. Rather than needing to treat these phrases as idioms which lack any compositional meaning, they can instead be treated as cases where multiple meanings can be mapped to a single verb with a transparently interpreted object.

These verbal idioms - cases of many-to-one correspondence between meanings and syntactic objects - have important implications for the way that we think about processing. Instead of "activating" a lexical concept which activates its unique correlates at other levels of representation, the non-lexicalist account requires a more active translation process which takes into account more sources of information, including feedback from subsequent processing stages. Importantly, the identity of the individual conceptual representations within the message may

[1] Proform or *one*-substitution used according to the definiteness of the object

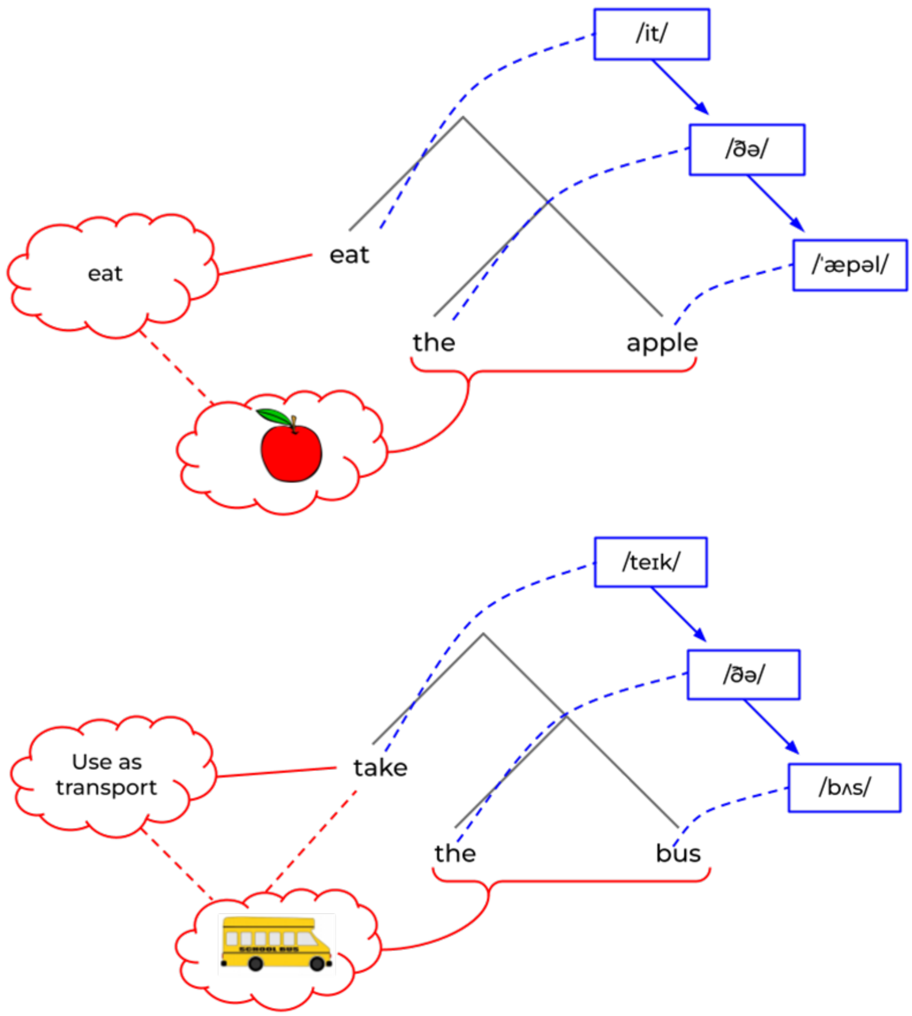

Figure 5.1: A representation of the mappings between a simple transitive phrase like "eat the apple" (above) and a verbal idiom like "take the bus" (below). In both cases, the message is made up of two components, an action and a patient. In the case of the simple transitive phrase, the selection of the verb and the noun object can be done independently, where one component of the message maps to one piece of syntax. In contrast, selecting the target verb in the verbal idiom case must also involve reference to the patient, and cannot be done independently.

not be a sufficient cue - on their own - to complete the translation between meaning and form, and may require some reference to the other conceptual representations in the message as well as their configuration within that message. For example, if one component of the message is the conceptual representation for “use as transportation”, it could correspond to several different syntactic objects, depending on what the intended transportation is. Using a bus for transportation would be associated with the verbs “catch” or “take”, while using a horse for transportation would be more strongly associated with the verb “ride” (whereas “take” or “catch” with “horse” as its object would be more strongly associated with “stealing” or “capturing”). Alternatively, sometimes the mode of transportation can even be productively derived into a verb without overt morphology, like “skateboard” or “Uber”, where the message could be conceptually complex, but it corresponds to a single complex syntactic object. In short, the choice of which verb to use cannot be attributed solely to the particularities of each conceptual representation, but seems to require additional information about the object and the relationship between the agent and the patient, and may also rely on a nuanced set of semantic or conceptual properties of the entities involved in the message.

Thus, the cost of the mapping process between meaning and syntax is dependent on several factors: the amount of information from the (syntactic or message) context that is required to complete the translation, the number of elements involved in the translation, the transparency of the mapping and the number of available syntactic alternatives for a given message component, as well as independent factors such as frequency or salience. These factors could theoretically be quantified for any given phrase, but for the sake of simplicity, I will assume a simple binary of “more complex” for the verbal idioms (given that they require a greater amount of context to complete the mapping and likely have a greater number of syntactic alternatives compared to the simple transitives) and “less complex” for the simple transitives (given that the individual elements in the message can be mapped to pieces of syntax without much reference to the context). The independent factors such as frequency and salience may vary across individual phrases, but I will assume that these factors do not vary significantly between the two groups.

A relatively small number of psycholinguistic studies have investigated the processing of verbal idioms or similar constructions. As mentioned above, the models of idiom processing in the lemma tradition assume that idioms have a “hybrid” representation, where each idiom has its own idiom lemma which activates the constituent lemmas. Sprenger et al. (2006) conducted a series of priming experiments in speech production and found that the literal meanings of the words within an idiom are also active during idiom production, consistent with the hybrid representation account. In a series of structural priming experiments looking at the production of idiomatic and non-idiomatic uses of particle verb constructions, Konopka and Bock (2009) observed that particle verb phrases are sensitive to the same syntactic generalizations regardless of their idiomaticity. Though these studies are both from a lemma tradition, they provide key insights into the kinds of factors that are relevant during the mapping from meaning to form, namely that the literal meaning and idiomatic meaning of a syntactic object should be taken into account, and that all phrases - regardless of idiomaticity - are ultimately subject to the same syntactic generalizations when they are mapped to a syntactic structure (to an extent - not all idioms preserve their meanings when they are syntactically manipulated, e.g., not all idioms can be passivized).

5.1.1 Comprehension of verbal idioms and multi-word expressions

Much of the literature on non-compositional phenomena in language has studied the comprehension of light verbs, idioms and multi-word expressions. These studies may not have a direct implication for our predictions regarding language production, but provide key data points that can guide the present study. The group of phrases we label as "verbal idioms" are a superset of "light verb constructions", where the meaning of the verb is either under-specified or subject to reanalysis, and may also be analyzed as "multi-word expressions", which may be stored as whole pieces of syntactic structure, and thus could be argued to bypass normal compositional processes. These studies provide conflicting evidence about whether the kind of mapping complexity that we are interested in is relevant during language comprehension, leaving an open question about whether this mapping complexity would be expected to influence language production processes. Placing these studies in the context of the non-lexicalist framework may also encourage some reinterpretation of their original conclusions.

Wittenberg et al. (2014) conducted a study which measured the ERP responses when participants listened to light verb constructions ("give a kiss") relative to their non-light counterparts ("give a rose"). The study was conducted in German, where the sentences were verb-final. They observed a sustained negative response for light verb constructions between 500 and 900 ms after the verb onset, and suggested that both light verb constructions and non-light constructions involve the same syntactic structure, but that the effect emerges because the light verb constructions have a less transparent relationship with meaning. These results provide a good starting point for future investigation, given that there are many alternative interpretations of the ERP difference that would need to be examined. Because this was a language comprehension task, if one were to assume that light verbs do not carry much semantic content, the listener may be required to perform another operation in order to get the verb's meaning. Alternatively, if light verbs are analyzed as having a "default" meaning that is overridden by the contextual interpretation, this result could be understood as a "reinterpretation" cost, where the listener initially interprets "give a kiss" as a literal giving action, and is required to reinterpret the utterance in order to get the intended meaning of the light verb construction. An experiment investigating the production of similar sentences would be able to address some of these questions, given that the speaker knows what the intended meaning of the utterance is; any cost which arises would have to be attributed to complexity in the meaning-syntax mapping rather than costs of reanalysis. In addition, assessing a larger set of verbs beyond just light verbs could provide a broader perspective about meaning flexibility and context-dependence in general (sidestepping the debate about whether light verbs carry semantic content or not).

Again on the comprehension side, a number of recent speech perception papers have examined the processing of idioms and multi-word expressions (MWEs). Bhattasali et al. (2018) uses fMRI to show that the appearance of multi-word expressions engages different brain areas from structure building operations (as approximated by a bottom-up parsing algorithm), consistent with an account where commonly-used phrases are subject to different processing constraints. However, these different neural responses may alternatively be related to retrieving the stored representation of the MWE, or based on the high transition probabilities between elements within these phrases, it could be related to the prediction of upcoming content. Cappelle et al. (2010) tests readers' responses to particle verbs like "rise up" (semantically transparent), "heat up" (not semantically transparent), and "fall up" (not attested). They observe that semantic transparency does not have an influence on the MEG signal, but do observe a difference between real particle verb combinations and ones that are not attested. Similarly, Bhattasali et al. (2020) observes that the strength of the association between the different elements in an MWE correlates strongly with fMRI data, suggesting that when the earlier part of an MWE is identified, it can trigger prediction processes to retrieve the rest of the phrase from memory and anticipate upcoming material. However, it is unclear whether neural responses to MWEs in comprehension translate directly into predictions about the neural mechanisms that support production of these expressions, as the operations needed to map from meaning to syntax to form are different from the calculations to map from form to syntax to meaning.

Within the literature on non-compositional phenomena, especially relating to idiom processing, there are a number of studies on comprehension, but very few production studies using neuroimaging methods with high temporal resolution (MEG or EEG). For this reason alone, conducting an EEG investigation during the production of verbal idioms provides an important perspective on the time-course of speech planning for a variety of utterance types, with varying degrees of compositionality and mapping complexity.

5.2 Context-Dependence in the Realization of English Tense and Aspect

It is often assumed - both in lexicalist models of language processing as well as many theoretical approaches - that syntactic constituency encodes linear order, or that it is the only cue required in order to determine the linear order of elements in the sentence; the syntax tree or dependency tree is printed flat on the page, and the phonological elements can fall off the tree and land (more or less) in the right order (e.g., Bock et al., 2002; Ferreira and Slevc, 2007). However, in the mapping from syntactic structures to form, there are several ways that this translation is not immediately straightforward; the syntactic configuration is not always sufficient to complete that transformation, but also requires information about the content of the syntactic configuration.

For example, although tense and aspect have a very straightforward syntactic structure, in English they can be realized both analytically - as multiple free phonological elements - and synthetically - as bound morphemes that attach to a verb. As a result, the position of a single tense morpheme and the way that it is realized phonologically can be significantly different across different sentences:

(28) a. Future Simple: *She **will** eat the apple* (Analytic, without agreement)

b. Preset Simple: *She eat-**s** the apple* (Synthetic, with agreement)

c. Present Progressive: *She **is** eat-**ing** the apple* (Analytic and synthetic)

Importantly, despite their differences in form, I will assume that the syntactic structure across these different phrases is approximately the same, but with different features in the tense and aspect positions (+FUT, +PRES, or +PROG), as shown in Figure 5.2.

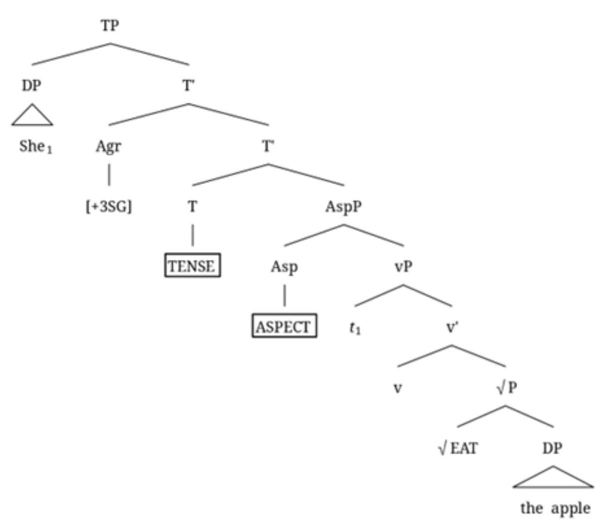

Figure 5.2: The basic tree structure underlying all verb phrases

Thus, because the syntactic structure has the same configuration across all phrases, there is an additional level of complexity in mapping between syntax and form, namely in that there is a calculation which determines the linear order of the elements in the verb phrase, as well as their form, taking into account both syntactic context as well as identity.

In some syntactic theories, the realization of different tense/aspect configurations is determined based on syntactic operations, namely a syntactic operation of head movement (Emonds, 1978; Travis, 1984; Pollock, 1989; Chomsky, 1986, 1991). For example, because the present tense cannot stand on its own, it lowers in the syntax to attach to the main verb; meanwhile, the future tense is pronounced as "will", and because "will" can stand on its own, it is not required to move to the lower position. In this type of theory, the syntax must transparently reflect how tense and aspect are realized in the phonology. In other accounts, however, this kind of head movement (T-to-v lowering) has been analyzed as post-syntactic, as in, it operates over a structure which is the output of syntax, and may be sensitive to individual phonological and/or prosodic properties of the elements that it operates on (Harizanov and Gribanova, 2019; Platzack, 2013; Embick and Noyer, 1999). For example, the realization of the comparative morpheme in English, *-er*, is more likely to arise as the periphrastic *more* when the adjective it modifies is longer than one syllable (e.g., *smart-er*, *more intelligent*, cf. # *more smart*, # *intelligent-er*), suggesting that the choice of whether the comparative morpheme is realized synthetically or analytically is not based purely on its syntactic configuration, but rather the syntactic configuration in tandem with phonological and prosodic information. Although I prefer the post-syntactic account, the distinction between syntactic and post-syntactic head movement may not matter for the present discussion, as long as one assumes that the producer would need to perform several calculations on-line in order to calculate the form of tense and aspect in either account (whether the calculation is over syntactic objects alone, or syntactic objects in combination with phonological and prosodic representations, may not necessarily generate divergent predictions).

In psycholinguistic theory, the standard description leaves something to be desired. Lemma-based models of language production reduce the internal structure of the verb phrase to a set of features on the verb, often merging tense and aspect into a single feature. Bound verbal morphology, such as "-ed" in the past tense, "-s" in the present tense, and "-ing" in the progressive aspect, arise because of the presence of certain tense/aspect features on the lemma; in contrast, auxiliaries like "will" in the future tense and "be" in the progressive aspect are not bound to the main verb, so they must constitute a separate lemma. There does not seem to be another way for tense to be realized analytically except through the presence of an additional lemma. Thus, the lemma model has to treat the English future tense differently from the past and present tenses in terms of the number of lemmas that are needed, even though - in most syntactic analyses - the different verb tenses occupy the same syntactic position and are just realized differently in the phonological form.

A non-lexicalist production model has much more flexibility in handling these types of phenomena, where the translation between syntax and form can be sensitive to the syntactic context. Given that the syntactic configuration alone does not fully determine linear order (assuming that all sentences have the same verbal spine), there are additional calculations sensitive to the identity of the syntactic elements and their possible phonological realizations that need to be performed in order to determine the form of the utterance. The question can be reframed around the types of information that is present in the syntactic structure that the language system can use to generate a phonological form with all of the elements in the correct order. Rather than presenting it in terms of syntactic features that have to attach directly to a verb lemma, the WOW model can assume the same basic structure for all verb phrases and allow context-dependent mapping to form, while integrating processes such as agreement, morphological attachment, and linearization.

Complexity in this mapping process can be approximated by the amount of context required in order to calculate the form, the transparency of the mapping and the number of alternative phonological realizations for a single syntactic object, and independent factors such as frequency and saliency. These factors can theoretically be quantified, but I will assume a three-way contrast between the future tense, the present simple tense, and the present progressive tense/aspect. The future tense can be interpreted as being the least complex, given that it is always realized as "will", it does not show any (overt) agreement with the subject, and it does not need to undergo any movement, suggesting that it requires minimal context to determine its form and linear position. Therefore, if the tense feature is identified as +FUT at the beginning of the translation between syntax and form, a simple linearization routine for that tense feature based on the syntactic configuration alone can be selected. In contrast, the form and position of the simple present tense needs to be computed on the basis of the other elements that are around it, especially agreement morphology; it can arise either as *-s* or a null morpheme depending on the features of the subject, having moved to the lower position and attached to the main verb. Finally, in the present progressive, because the progressive aspect is realized as the auxiliary verb *be* and an *-ing* suffix, the present tense can arise as *am*, *are*, or *is* depending on the features of the subject, because the tense and agreement features will attach to the auxiliary verb rather than the main verb. The form and position of each element in the present progressive - tense, aspect, and the verb root - is conditioned by the identity of the other elements in the phrase. Independent factors such as frequency and saliency may vary to some degree across the different tenses/aspects, but given that these morphemes are all fully productive and occur in a wide variety of contexts, I will assume that these factors do not vary significantly between the three groups.

A similar process has been studied in Clahsen et al. (2018), an ERP investigation of language production which focused on the formation of the comparative in English. The authors suggest that, because the two possible realizations of the comparative (*more Adj* vs A*dj-er*) involve different mechanisms (with or without affixation), the processes of realizing the comparative morpheme analytically or synthetically should incur different neural responses. The study found a frontal bilaterally distributed negative waveform for synthetic realization of the comparative morpheme (*-er*) with the analytic realization (*more*), which emerged around 300 ms after a silent production cue[2]. One way that this is different from the realization of tense and aspect in English is that the comparative can only determine its form by referring to the adjective that it modifies; it is *only* context-sensitive, whereas tense has the possibility of not referring to context depending on the tense feature (e.g., the future tense). In both cases, however, the element that is modified (the adjective in the comparative case, or the verb in the tense case) has a context-dependent realization, where its phonological form may be modified depending on the syntactic and phonological elements in its context. Thus, in terms of the realization of tense and aspect in English, we would expect synthetic realizations of tense to incur different neural responses than the analytic realizations of tense, where future tense should have a different effect from the present simple; in addition, the present progressive should show effects of both synthetic and analytic realizations, where aspect is realized both synthetically and analytically, and tense is realized synthetically on the auxiliary verb. This experiment also provides a clue for the time course of this process, where an effect may emerge as early as 300 ms post stimulus onset, with a potential delay based on the complexity of the task.

[2]Other conditions were included to control for the effect of syllable count, given that *-er* arises more often on shorter adjectives, while *more* arises more often with longer ones. The effect of analytic vs. synthetic realization was still significant.

Another factor which may modulate the complexity of the mapping between meaning and form is frequency, including the frequency of the different syntactic features, as well as the frequency of the occurrence of different kinds of realization strategies - are abstract syntactic features like tense more likely to lower to attach to another syntactic object, or are they more likely to arise independently? How often are morphemes dependent on the syntactic object they modify or attach to in order to determine their form (e.g., the English comparative, irregular verbs in the English past tense, etc.)? Given that the answers to these questions cannot be gathered straightforwardly from English corpora, I consider these outside the scope of the current chapter, but questions ultimately worth pursuing.

5.3 Experiment

For a wide variety of linguistic phenomena, language production involves not only combinatorial operations such as syntactic structure building, but also active, context-dependent mapping processes between stored representations of meaning, syntax, and form. The two sets of mappings vary in complexity depending on the number and arrangement of the representations that are mapped together, as well as the degree to which context is relevant in that mapping. For a single mapping process, there might be multiple alternative syntactic structures that are considered for a single message component, or multiple form segments that are considered for a single syntactic object. The calculations which determine which syntactic objects or forms are generated cannot be easily characterized as just passive spreading activation between fixed triads of meaning, syntax, and form, as is typically assumed for the lemma model; instead, this calculation involves active consideration of multiple alternative candidates based on the context - in terms of meaning, syntax, and form - as well as feedback from subsequent processing stages.

The WOW model emphasizes the role of mapping processes alongside the combinatorial processes, which raises many new questions about how those mapping processes are engaged as a speaker or signer prepares an utterance. How does the complexity of the mapping processes modulate neural activity? How is the mapping process affected by factors such as the number of actions or entities involved in the conceptual representation, the size and complexity of the syntactic representation, and the transparency of the relationship between those two representations? What happens when the information used to transform the data structures at each stage are ambiguous or unreliable? New neuroimaging experiments are required to answer these questions, and to better understand how those mapping processes may be implemented during language production.

This experiment introduces and tests a starting hypothesis about how these mapping processes in production translate into neural processing cost measurable with non-invasive recording techniques, providing insights into questions that are challenging or awkward to answer in a lexicalist framework. I hypothesize that in the computation of both meaning-to-syntax mappings and syntax-to-form mappings, greater mapping complexity is associated with greater neural processing cost, and that the neural responses to these two types of complexity should be temporally and spatially distinct. Although there are different options for formalizing mapping complexity, here I focus on relatively obvious complexity differences: cases where there is reduced systematicity in how similar representations on one side get mapped to the other, and cases where context needs to play a greater role in the mapping process.

5.3.1 Design

To investigate the role of mapping complexity - independent of syntactic complexity - in modulating neural activity during language production, this study records neural activity during the pre-articulation stages of sentence production, for sentences that vary in the complexity of the meaning-syntax mapping and the syntax-form mapping, but which have the same basic syntactic structure. Within this experiment, there are essentially two sub-experiments, investigating the different mapping processes.

Experiment 1 compares the production of sentences that are more or less transparent in the mapping between meaning and syntax, such as "he will wash the dishes" (where all elements of the sentence can be interpreted transparently) compared to "she will strike the pose" (which has a non-default, context-dependent meaning for "strike"). In the default case, producing a phrase like "wash the dishes" would involve a speaker thinking of a washing event, where the patient of that event is dishes; the speaker would just need to access a verb and a noun that correspond to each of those conceptual/semantic representations. In the verbal idiom case, we could not say that producing a phrase like "strike the pose" involves a speaker thinking of a striking event, where the patient of that event is a pose; the speaker could not have accessed that syntactic representation through the default method of linking those event and patient representations to a corresponding noun and verb. Instead, it must involve a context-dependent mapping that considers the meaning associated with the verb in the context of the object.

Experiment 2 investigates complexity in the mapping between syntax and form, particularly in the way that verb tense and aspect is realized. In English, the future tense involves the addition of an auxiliary verb *will* which precedes the main verb ("will walk"), the present simple tense involves the addition of a suffix that exhibits agreement with the subject ("walks"), and the present progressive involves an inflected *be* auxiliary that agrees with the subject as well as a suffix on the main verb ("is walking"). As a consequence, the computations involved in the phonological realization of tense and aspect are modulated by the amount of context required to perform the translation - whether the form of the tense morphology can be determined independently, like in the future tense, or whether it needs to refer to the features of the subject and the form of the verb, like in the case of the present simple and present progressive. The present simple and the present progressive both involve affixation and agreement, but the present progressive involves an affix on both the auxiliary verb and the main verb which may incur an additional cost compared to the present simple.

As illustrated in Figure 5.3, the full experiment (Experiments 1 and 2 combined) consists of four conditions, with one set of verbal idioms in the future simple tense, and three sets of simple transitive phrases, one each in the future simple tense, present simple tense, and present progressive tense/aspect. In each trial, the participant first hears a question, they see an image, and they use that image to formulate an answer to the question with the appropriate inflections. The questions provide the participants with the target tense/aspect and person/number for their response (e.g., "what is this person doing today?", "what will this person do tomorrow?", "what do these people do every day?"). Asking participants to use the phrase in response to a question ensures that participants actually engage the correct conceptual representations and syntactic operations when they produce the target phrase, and do not bypass those representations by just relying on a memorized auditory representation.

To review, experiments 1 and 2 involve two sets of comparisons. The first investigates complexity in the meaning-syntax mapping by comparing the production of verbal idioms and

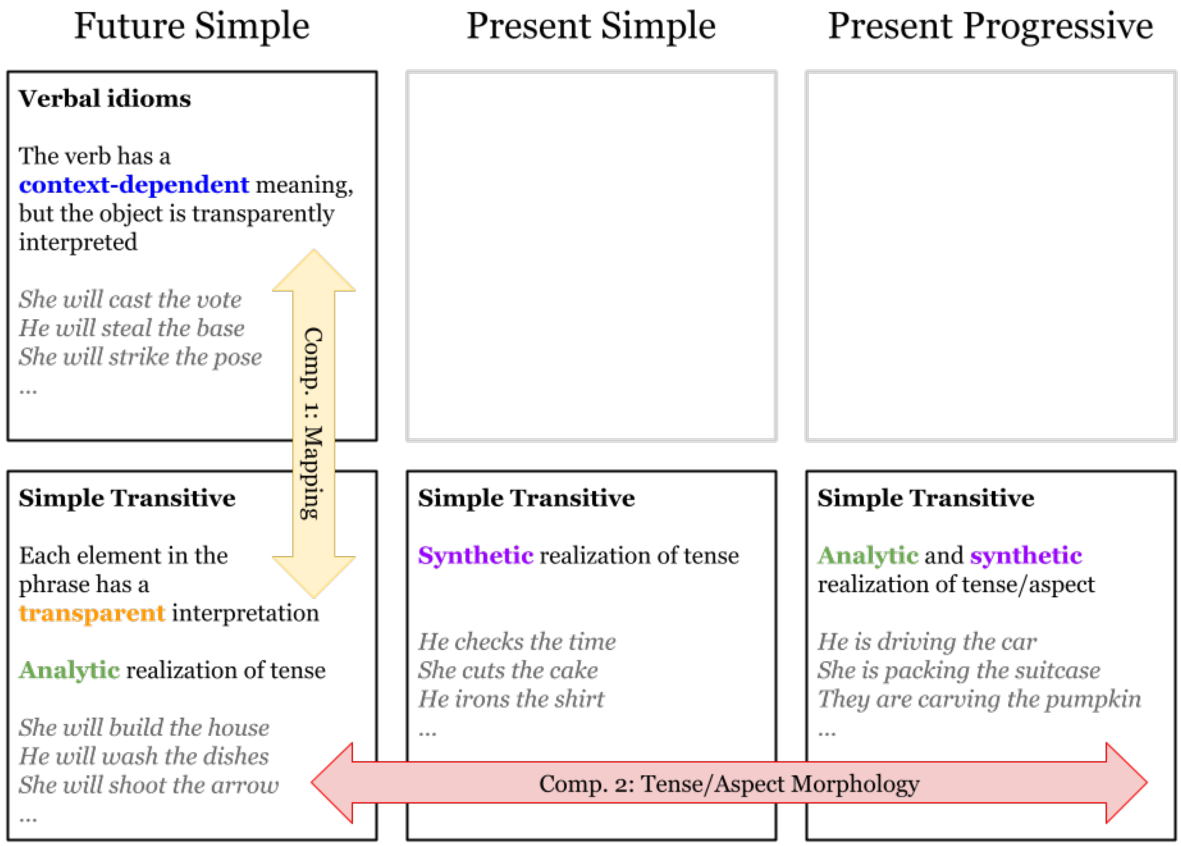

Figure 5.3: Diagram illustrating the four conditions and the two comparisons involved in this experiment, testing the effects of different kinds of mapping complexity

simple transitive phrases in the future simple tense. The future tense is realized analytically, with the auxiliary verb "will" appearing before the verb without any overt markers of agreement, so any differences in neural responses to those conditions can only be attributed to the role of context in the mapping between meaning and syntactic structure. The second comparison investigates the processes of phonetic realization and linearization of tense and aspect morphology by comparing the production of simple transitive phrases in the future simple (realized analytically with no overt markers of agreement), the present simple (realized synthetically with agreement based on person and number), and the present progressive (realized both synthetically and analytically, with the auxiliary verb "be" inflected for person and number). The differences and similarities between these three groups can be attributed to the role of linearization and agreement processes in the mapping between syntax and form.

5.3.2 Stimuli

120 stock photographs were selected to use as image stimuli, 90 corresponding to the Simple Transitive condition, and 30 corresponding to the Verbal Idiom condition. The images for each phrase include one or more people, or an implied person (i.e., showing hands but no face). The images were selected based on how recognizable they are for the target phrase, with limited alternative phrases. The images are in full color with limited text, with people that are diverse in perceived age, gender, and race. Each image was cropped square to reduce the need for eye movements during the trials.

The target Simple Transitive phrases which we aimed to elicit with the selected images are all fully compositional, where the meaning of the verb and the object are completely transparent. The target Verbal Idioms which we aimed to elicit are all phrases where the object's meaning is transparently interpreted, but the verb has something other than its "default" meaning. This was assessed by putting the verb in a sentence context with a non-word as its object (e.g., "he strikes the prindle" gives the verb an interpretation like hitting, in contrast to the verbal idiom "strike the pose"). Some verbs were repeated across different phrases in this condition, but none of the repeated verbs had the same meanings across phrases (e.g., "catch the bus", "catch the wave", and "catch the flu" each use the same verb, but with different associated meanings). All of the phrases use transitive verbs that can take a definite noun as its object without the addition of any prepositions (no phrasal verbs or particle verbs), and all of the object nouns were able to be used as definite nouns, without any reflexives or possessive pronouns (e.g., phrases like “brush their teeth” were not used). Furthermore, all of the phrases were judged to be acceptable with the given set of tenses and aspects (present simple, present progressive, and future simple). The phrases were selected for imageability and plausibility, and many of them were highly frequent.

One question was created for each combination of subject type (singular or plural, and by age, according to the figure(s) in the image) and the three target tense conditions. In addition to the tense morphology on the verb, the questions also included a modifier to further indicate the temporal nature of the action. The future tense questions included “tomorrow” (“what will this person do tomorrow?”, “what will these people do tomorrow?”, and “what will this kid do tomorrow?”), the present simple tense questions included “every day” (“what does this person do every day?”, “what do these people do every day?” and “what does this kid do every day?”) and the present progressive questions included “today” (“what is this person doing today?”, “what are these people doing today?”, and “what is this kid doing today?”). These modifiers help the questions to sound more natural, and also give the participants an additional cue for which tense to use.

The stimuli were presented in one of three lists, with a total of 120 items. All lists include the Verbal Idioms with a question in the future simple tense (N=30), while the Simple Transitive phrases are equally distributed between the three target tenses: future simple (N=30), present simple (N=30), and present progressive (N=30). The Simple Transitive phrases were presented with a different target tense in each list. The order of the stimuli was randomized across lists.

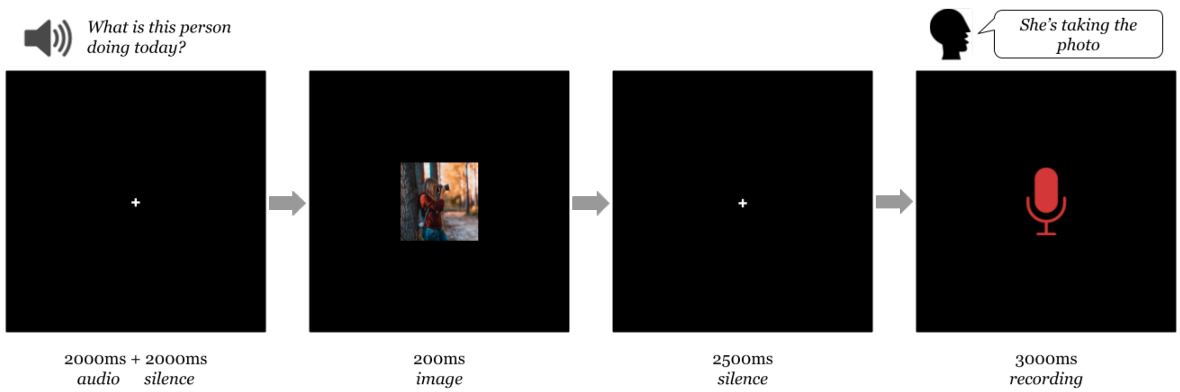

Figure 5.4: Illustration of a single trial in the experiment

5.3.3 Procedure

Before the study begins, the participants are familiarized with the images and target phrases in a picture book that they can go through at their own pace while the EEG equipment is set up. The matching phrases are presented alongside the images as text, without tense or subject agreement, in order to limit the availability of phonological or rote memory to aid in the task.

As shown in Figure 5.4, each trial of the experiment begins with a fixation cross (white on a black background) appearing in the center of the screen while a question is presented auditorily, with a duration of approximately 2 seconds, followed by a 1 second delay to allow all auditory processing to conclude. Then, one of the familiarized images appears in the center of the screen for 200 ms, long enough for the image to be consciously recognizable but not long enough for the participant to plan a saccade. The images are all 300 pixels square. The fixation cross then returns for 3300 ms, during which the participants prepare their response to the question using the phrase associated with the image which was presented, and inflecting it for person and tense, for a total of 3500 ms of preparation time. Finally, a red microphone icon appears on the screen (100 pixels square), cueing participants to articulate the response that they had prepared.

Participants receive four practice trials at the beginning of the experiment, with familiarized images and phrases that are not included in the main experiment. The experiment consists of 120 trials, with four short breaks evenly distributed throughout. Participants are asked to avoid any unnecessary eye or body movements during the trials while the fixation cross is on the screen, to reduce any movement artifacts. The duration of the experiment is approximately 40 minutes, so an experimental session including EEG setup lasts approximately 90 minutes.

The analysis focuses on the neural activity during the 1000 ms after the image is presented, in the pre-articulation period when the sentence is being planned, for correct trials only. A trial is considered "correct" if the participant produces the correct phrase with the appropriate tense as indicated by the question prompt, with no speech errors. Trials are also marked as "correct" if the phrase which the participants produced is sufficiently similar to the target ("They will fold clothes" is marked correct when the target was "They will fold the laundry"), or a plausible description for the image that has the same syntactic structure and is in the same category as the target sentence (as a Verbal Idiom or a Simple Transitive phrase; e.g., "He is drinking the coffee" is marked correct when the target was "He is sipping the tea"). Trials are excluded if the participant produces the phrase before they are cued, or if they produce the phrase too long after the cue; the movements associated with articulation would create artifacts if they begin speaking during the target period, and a delay in the articulation could indicate that the participant failed to recall the correct phrase or plan the sentence in time, suggesting that the neural activity occurring immediately post stimulus presentation may not reflect sentence planning. Given the relatively open-ended nature of the task, participants were expected to have relatively lower accuracy rates than for a single-word production task or a language comprehension task.

5.3.4 Participants

All 28 participants were native speakers of English, many of whom were undergraduate students at the University of Maryland. Participants were all right-handed with normal or corrected-to-normal vision. They signed an informed consent form at the beginning of the experiment, and were paid or received course credit as compensation. 10 participants were excluded due to low accuracy in the production task (below 60%) or due to high rates of uncorrectable artifacts in the EEG data. 18 participants were included in the final data analysis (14 female-identifying, mean age 19.5, range: 18-24).

5.3.5 EEG data acquisition

Twenty-nine tin electrodes (O1, O2, P7, P3, Pz, P4, P8, TP7, CP3, CPz, CP4, TP8, T7, C3, Cz, C4, T8, FT7, FC3, FCz, FC4, FT8, F7, F3, Fz, F4, F8, FP1, FP2) were held in place on the scalp by an elastic cap (Electro-Cap International, Inc., Eaton, OH). Bipolar electrodes were placed above and below the left eye and at the outer canthus of the right and left eyes to monitor vertical and horizontal eye movements. Additional electrodes were placed over the left and right mastoids. Scalp electrodes were referenced online to the left mastoid and re-referenced off-line to the average of left and right mastoids. The ground electrode was positioned on the scalp in front of Fz. Impedances were maintained at less than 10 kΩ for all scalp and ocular electrode sites and less than 5 kΩ for mastoid sites. The EEG signal was amplified by a NeuroScan SynAmps® Model 5083 (NeuroScan, Inc., Charlotte, NC) with a bandpass of 0.05–100 Hz and was continuously sampled at 500 Hz by an analog-to-digital converter.

5.3.6 Predictions

Prior neuroimaging studies give some rough constraints about the time course of planning stages in production tasks. Using intracranial electrophysiology, Sahin et al. (2009) identified a stage of "lexical" processing occurring before 200 ms post stimulus onset, and a stage of grammatical composition occurring before 350 ms. Strijkers et al. (2010) observed cognate and frequency effects in ERPs of Spanish-Catalan bilinguals as early as 150 ms, and Strijkers and Costa (2011) conducted a meta-analysis of similar studies on the time course of production, suggesting that the process of "lexical access" occurs within 200 ms. In addition, as discussed above, Clahsen et al. (2018) observed a difference between neural responses when producing synthetic and analytic morphology starting around 300 ms.

The results of these previous studies only allow us to determine an approximate time course for the mapping processes that we are interested in. Given that I am taking a non-lexicalist approach, there is no stage of "lexical access" or "lexical processing" in my model, but rather a stage of conceptual preparation followed by a mapping to syntactic structure, followed by a mapping to phonological form. Furthermore, these studies tested participants on single-word production tasks or simple phrases, while we are planning to elicit complete sentences. Based on those previous findings, I assume that conceptual preparation should be completed by around 200 ms, morphosyntactic composition should be completed by around 300-350 ms, and that everything after 350 ms is related to phonological processes and linearization. These time windows are consistent with the earlier literature which tested single words and short phrases, but longer sentences may cause these stages to occur later and last longer.

For the meaning-syntax mapping, the WOW model suggests that the Verbal Idiom condition has a more complex mapping than the Simple Transitive condition, and consequently should have a greater neural response. This would not be a divergent prediction from a lemma-based model, which would suggest that the Verbal Idioms should incur a greater neural response than the Simple Transitives because they would engage an additional "idiom lemma". However, both of these predictions would diverge from a view where light verbs and the verbs in the Verbal Idiom condition have less semantic content, which may predict that the Verbal Idioms would be easier to retrieve than the Simple Transitives simply because there is less semantic content to retrieve.

In the syntax-form mapping, the predictions between the WOW model and lemma-based models do diverge significantly. The WOW model suggests that context-dependence adds complexity to the mapping process, so the future tense should be the least complex given that each syntactic element is more or less independent, while the present simple is more complex, given that the verb has to be inflected for tense and person/number, and the present progressive is the most complex, where the auxiliary verb has to be inflected for tense and person/number and *-ing* is attached to the main verb. In contrast, lemma models would predict that the future tense and the progressive aspect should be more costly, because they involve an additional lemma (*will* or *be*). Moreover, this effect should be present from an early stage, given that those lemmas have to be retrieved via their lexical concept (rather than as a product of the syntactic configuration).

In the WOW model, greater complexity in the mapping between meaning and syntax should incur greater neural activity early in the production process, between 200 ms and 300 ms (where the Simple Transitive condition is less complex than the Verbal Idiom condition, given that the mapping between meaning and syntax is more transparent in the Simple Transitive condition). This activity is predicted to be localized to the middle temporal lobe, as discussed in Chapter

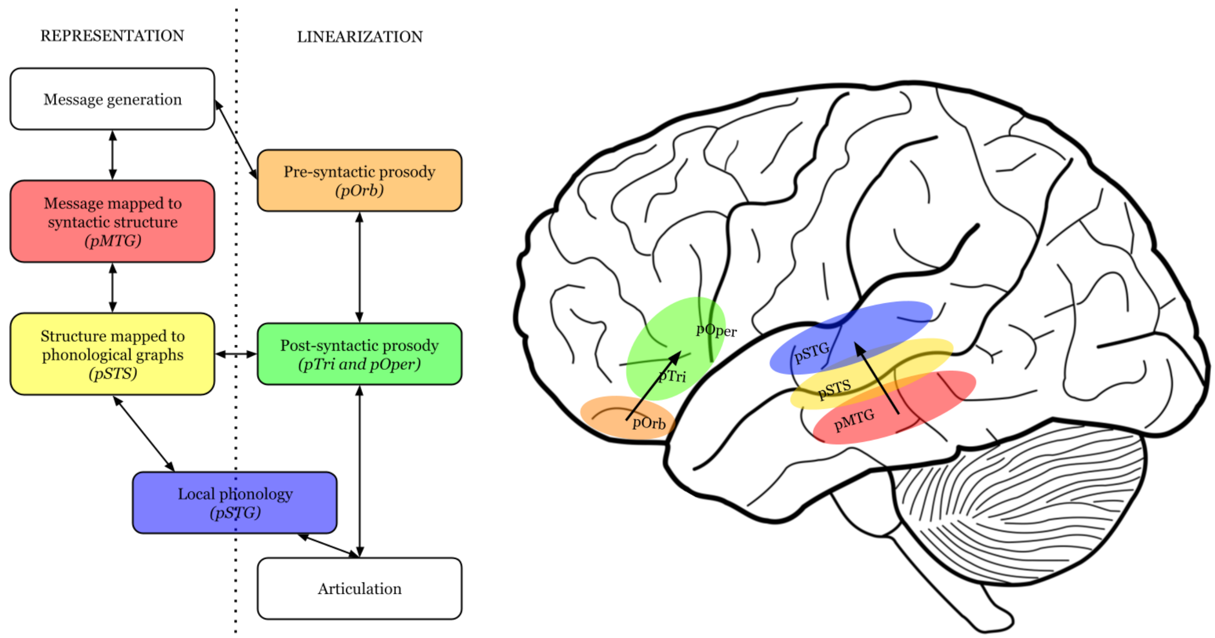

Figure 5.5: The mechanisms of the WOW model and their associated brain areas.

4, shown again below in Figure 5.5. For the second set of mappings, greater complexity in the mapping between syntax and form should incur greater neural activity later in the production process (perhaps with different effects for synthetic and analytic morphology). This activity is predicted to be localized to superior temporal lobe structures and the frontal lobe, corresponding to the combined effects of calculating the phonological form (pSTS and pSTG), as well as linear control mechanisms that resolve competition between different order alternatives based on phonological and prosodic constraints (linearization, localized to the pTri and pOper). In other words, greater phonological uncertainty for a given syntactic object will incur greater neural activity as the phonological form of that syntactic object is being calculated. The exact timing of this activity may be more varied compared to that of the meaning-syntax mapping, given that phonological planning may occur according to the order that the elements appear in.

A mock plot of the predicted results is shown in Figure 5.6. Two kinds of statistical analyses will be conducted, the first comparing the EEG signal in the 200 ms-300 ms post stimulus onset

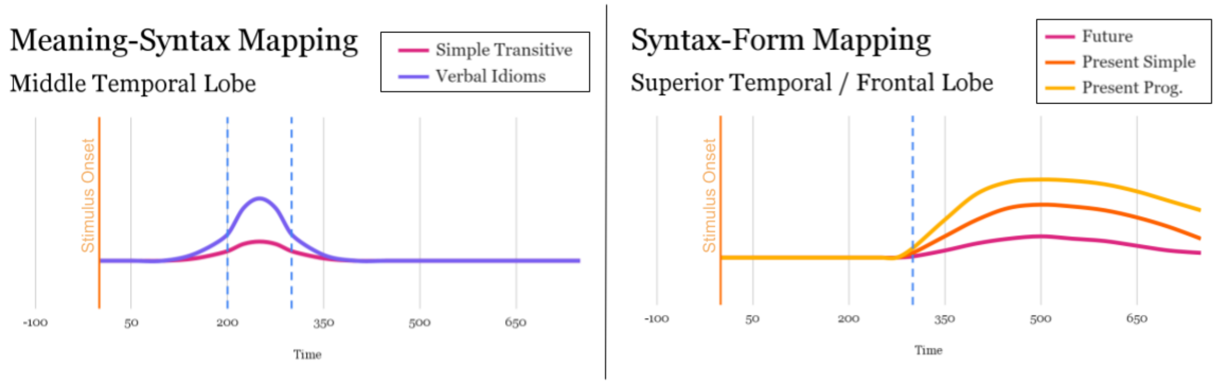

Figure 5.6: Predicted EEG signals for the conditions in Experiment 1 (left) and Experiment 2 (right)

time window across the different conditions of the meaning-syntax mapping, with the prediction that the condition with more complex mappings (a greater role of context in selecting the verb) should have a larger response than the condition with more transparent mappings. The second statistical analysis compares the EEG signal after 300 ms across different tense/aspect conditions, with the prediction that greater complexity (a greater role of context in determining form) should correspond to a larger response compared to forms with less complexity.

5.4 Results

For each participant, all trials where they did not produce the target phrase were removed from the analysis (22.8% of trials were excluded per participant on average). Accuracy on the three Simple Transitive conditions were approximately the same across participants (∼82.5%), with lower accuracy on the Verbal Idiom condition (63.3%). Each participant's data were preprocessed individually to remove bad channels, and an independent components analysis (ICA) was performed to remove artifactual components such as blinks, muscle artifacts, and other noise. For two participants, a high-pass filter (1 Hz) was applied to the continuous EEG remove drifts.

Any remaining artifacts were rejected automatically in EEGLAB (11.6% of trials rejected per participant on average). After removing incorrect responses and cleaning the data, an average of 69% of the 120 trials were kept for each participant (range: 57%-90%) and included in the statistical analysis.

The ERP was time-locked to the moment of image presentation, from -100 ms to 1000 ms post stimulus onset. A low-pass filter (40 Hz) was applied to all datasets before calculating the grand average for each condition. It should be noted that the amplitude of the ERP responses overall was large (peaking at around -15 μV), but that this is consistent with what has been observed in other studies that involved complex scene presentation (e.g., Lowe et al., 2018; Mudrik et al., 2010; Ganis and Kutas, 2003).

5.4.1 Experiment 1: Meaning-syntax mapping complexity

The first comparison investigates the role of mapping complexity between meaning and syntax, where the Verbal Idiom (future tense) condition is argued to involve more complexity than the Simple Transitive (future tense) condition. The ERP waveforms for those two conditions are presented in Figure 5.7. As predicted, there is a difference between the two conditions that arises before 300 ms post stimulus onset, particularly in the 220-280 ms time window. This effect seems to be broadly distributed, as seen in Figure 5.8, which shows a scalp map of the average difference between the two conditions in that time window.

The exploratory statistical analyses based on the visual inspection above confirm those observations. A one-way ANOVA of the ERP response across all scalp electrodes in the 220-280 ms time window showed a near-significant effect of condition ($F(1, 17) = 3.292$, $p = 0.087$). This

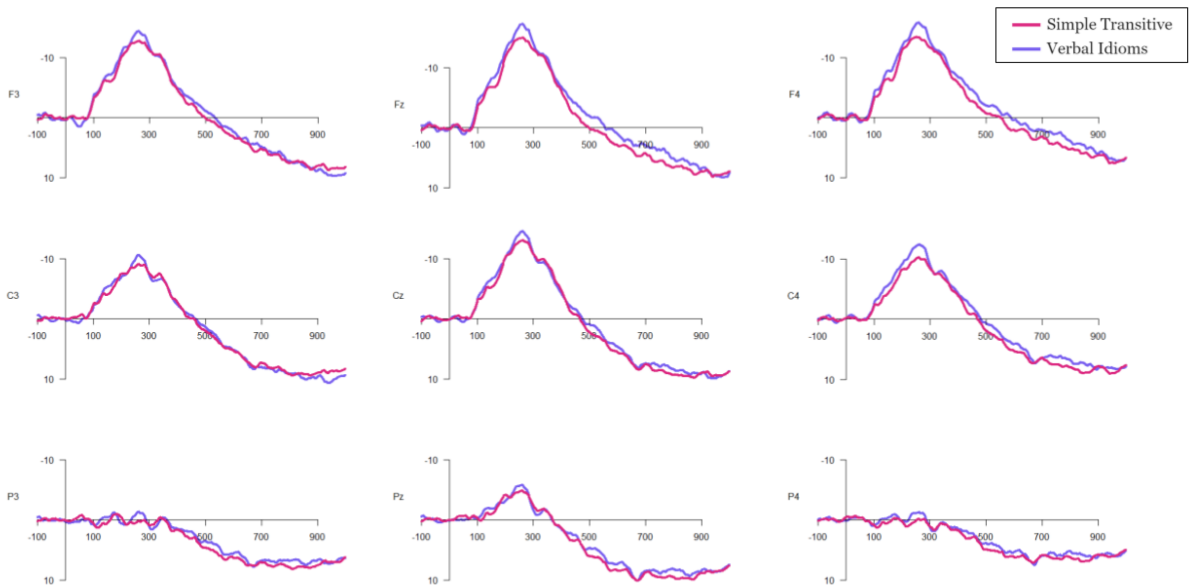

Figure 5.7: ERPs at electrodes F3, Fz and F4 (top), C3, Cz and C4 (middle), and P3, Pz and P4 (bottom) for the Verbal Idioms and the Simple Transitives presented in the future tense (purple = Verbal Idiom/Future, pink = Simple Transitive/Future)

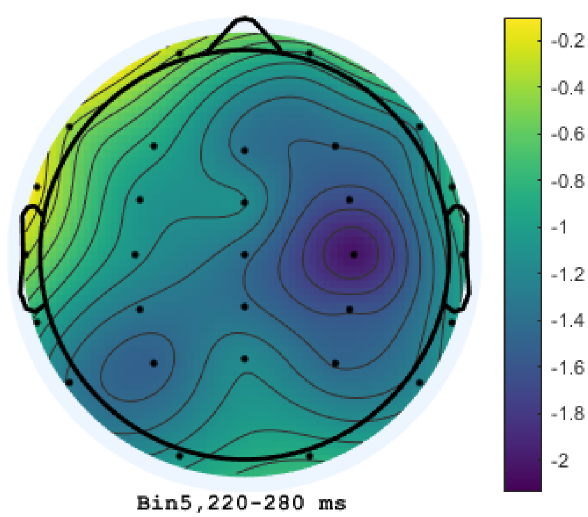

Figure 5.8: Scalp map of the average difference between the Verbal Idioms and the Simple Transitives between 220 ms and 280 ms post stimulus onset

effect may become statistically significant with a larger number of participants.

This time window is consistent with the predictions discussed earlier in Section 5.3.6, where the mapping between the message and syntax occurs after the processes of conceptual preparation and message generation (completed within the first 200 ms post stimulus presentation) and before the syntactic structure is mapped to form (beginning around 300 ms). This time window was estimated based on studies using “single-word” production tasks, which concluded that “lexical access” should occur within the first 200 ms. In a non-lexicalist framework, this could correspond to the mapping processes to syntax or form, depending on whether the experiment assessed effects of semantic, syntactic, or phonological properties. Given that these tasks involve less linguistic content than a sentence production task, with less conceptual preparation required, it is not surprising that the effect in the present experiment occurs slightly later than in those previous studies.

If we take a lexicalist perspective, and also assume that the single-word timeline (Strijkers and Costa, 2011) is applicable here, this would suggest that the component lemmas for *take the stairs* were retrieved by 200 ms, but there would be no obvious explanation for the effect observed between 220-280 ms (or the lack of an effect before then). One possible analysis would be that the idiom lemma for *take the stairs* is activated later, where an initial set of lemmas are retrieved that have a transparent relationship with the message in both conditions, and then in the Verbal Idiom condition, an idiom lemma could perform a type of “lemma correction” to substitute those retrieved lemmas for ones that are part of an existing idiom. This would be a very different account from what has been discussed previously, where the idiom lemma is argued to be retrieved first and activate its component lemmas, and also seems like it would make idioms much more costly to produce (perhaps prohibitively so). It is much more parsimonious -

in both lexicalist and non-lexicalist accounts - to assume that the conceptual preparation process takes longer for phrases than for object names, and that the visual analysis of complex scenes takes longer than for single objects, thus delaying some of the associated neural responses. The single-word timeline should be adjusted to account for that.

The items in the Verbal Idiom condition and the Simple Transitive condition are different phrases, and as a cross-item comparison, there is a possibility that the effect arises solely from those conceptual, semantic, or phonological differences between the different target phrases. In addition, the Verbal Idiom condition involves verbs that do not have a transparent meaning, so the interpretation of the image and the process of identifying the events involved in the scene may have been less straightforward. For example, the scene paired with the target phrase "wash the dishes" shows a man at a sink washing dishes; from that image, the participant can identify a man, a dish, and a washing event. The scene paired with the target phrase "take the stairs" shows a man climbing the stairs, and from that image, the participant can identify a man, a set of stairs, and a climbing event, but not a "taking" event. In the more transparent case, the washing event in the image can be quickly associated with the "washing" verb, whereas in the more context-dependent case, the climbing event in the image cannot be quickly associated with the "taking" verb. That being said, the non-transparent relationship between the verb and the visual elements in the scene should be parallel to the non-transparent relationship between the verb and the meaning elements in the message. The participant needs to perform additional calculations in order to link the climbing event to the non-transparent target phrase compared to the washing event and the transparent target phrase, which is the goal of this experiment. Ultimately, the complexity of the visual task may not be separable from the complexity of the meaning-syntax mapping. If these two kinds of complexity do incur different neural responses, they are most likely co-occurring.

5.4.2 Experiment 2: Syntax-form mapping complexity

The second comparison investigates the role of mapping complexity across different tenses of the Simple Transitive condition. Because the same phrase was presented in all three tenses across different subjects, this comparison has better controls for semantic variation between items. The three ERPs are presented in Figure 5.9. According to the earlier predictions, this effect should arise after 300 ms post stimulus onset, where either the future tense condition should be less complex than the present simple condition which is less complex than the present progressive condition (assuming that the future tense is the least context-sensitive), or where the future tense condition should be more complex than the present simple or present progressive conditions (because the future tense condition involves an additional phonological word, while the others do not[3]). Visual inspection identifies two time windows where the ERP waveforms diverge, one where each prediction seems to be correct; the first is between 300 and 350 ms, where the future tense elicits greater negativity than the present progressive or present simple tenses, and the second starts at 500 ms post stimulus onset and persists to the end of the epoch, where the present simple tense elicits greater positivity than the future or present progressive conditions. The effect in the 300-350 ms time window seems to be more anterior than the effect in the 500+ ms time window, as seen in the scalp maps of the difference between tense conditions in Figure 5.10 and Figure 5.11.

The exploratory statistical analyses based on that visual inspection provide some support

[3]It is worth noting here that *is* does not usually constitute its own phonological word when it is phrase-medial and unstressed (Selkirk, 2014)

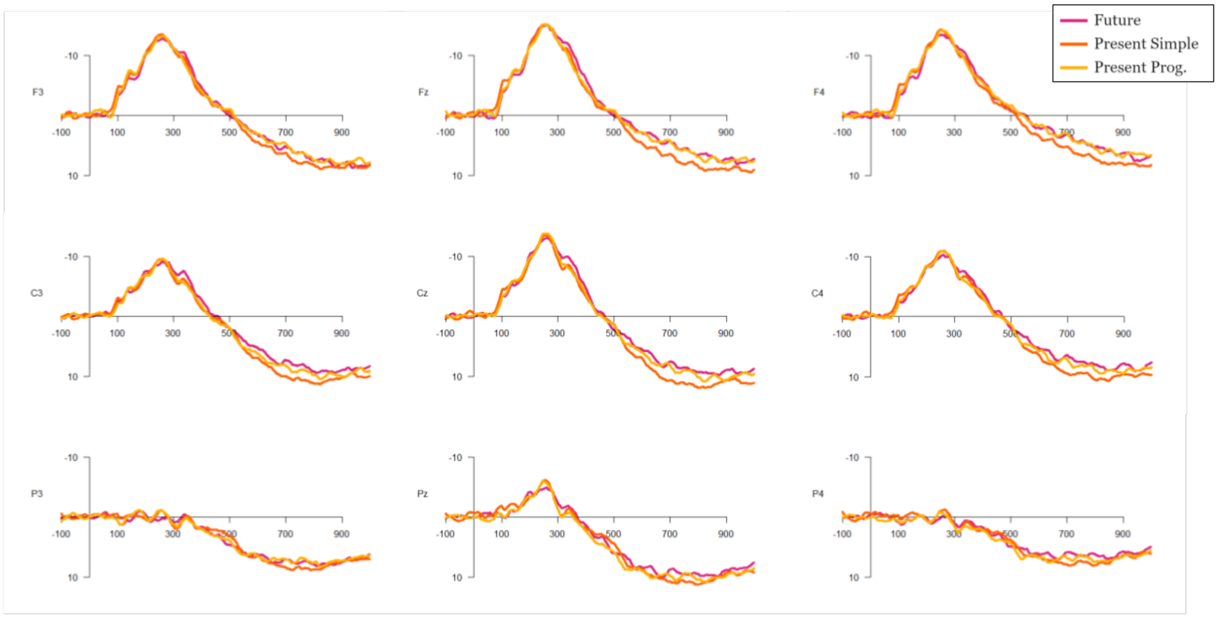

Figure 5.9: ERPs at electrodes F3, Fz and F4 (top), C3, Cz and C4 (middle), and P3, Pz and P4 (bottom) for the Simple Transitives presented in the future tense, present simple tense, and present progressive tense/aspect (pink = Simple Transitive/Future, orange = Simple Transitive/present simple, yellow = Simple Transitive/present progressive).

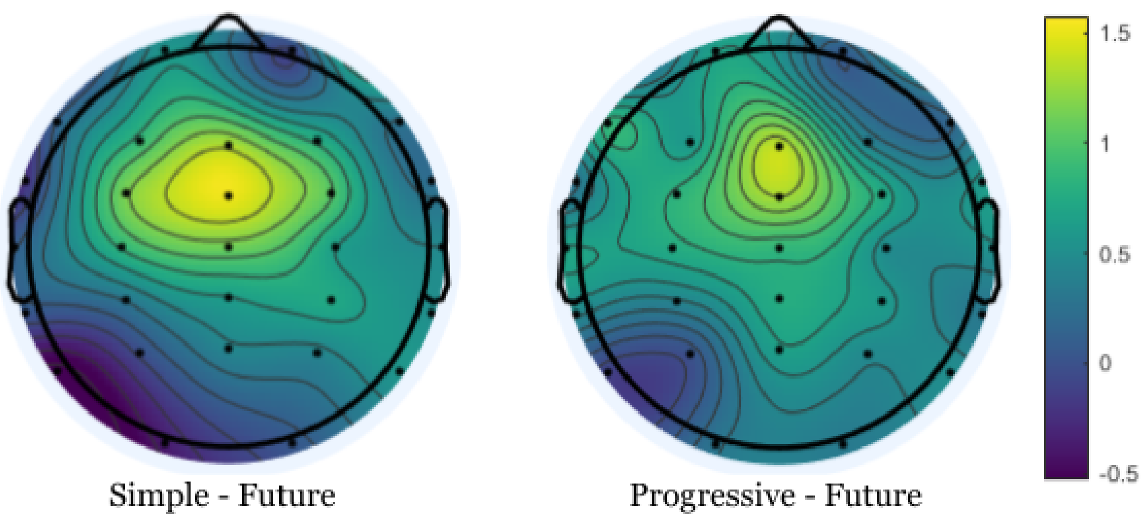

Figure 5.10: Scalp map of the average differences across tense conditions between 350 ms and 400 ms. **Left:** difference between the Future and the Present Simple, both in the Simple Transitive condition; **Right:** a greater difference is observed between the Future and the Present Progressive.

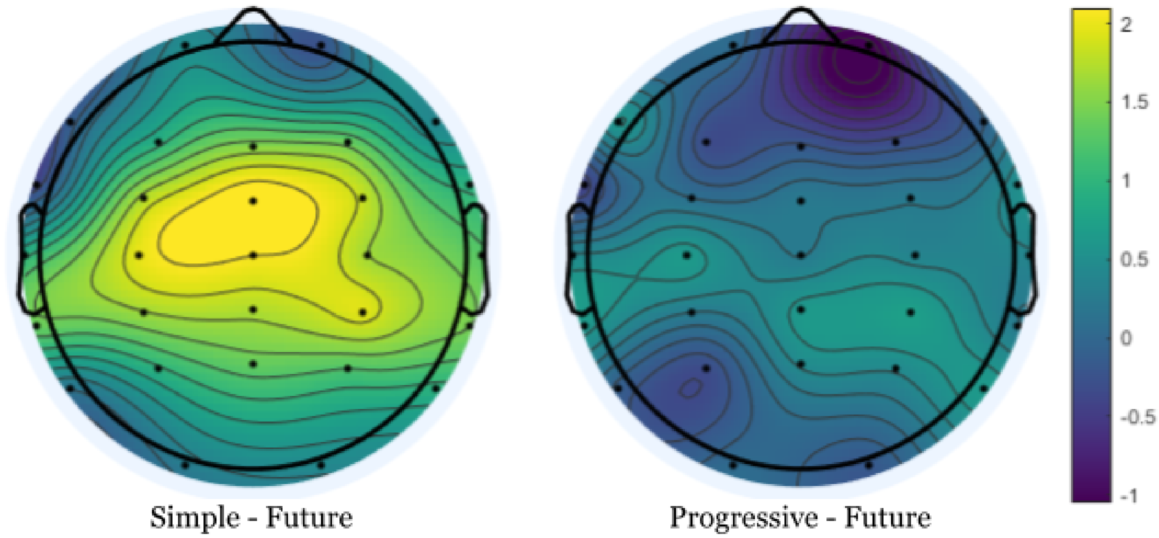

Figure 5.11: Scalp map of the average differences across tense conditions between 700 ms and 800 ms. **Left:** difference between the Future and the Present Simple, both in the Simple Transitive condition; **Right:** no significant difference was observed between the Future and the Present Progressive

for those observations. A one-way ANOVA of the ERP response across all scalp electrodes for the future tense and the present progressive tense in the 350-400 ms time window showed a trend, but was not significant ($F(1,17) = 2.325$, $p = 0.146$), while a one-way ANOVA of the ERP response across all scalp electrodes for the future tense and the present simple tense in the 700-800 ms time window showed a near-significant effect of condition ($F(1,17) = 3.564$, $p = 0.076$). These two effects may become significant with a larger number of participants.

Based on these results, it appears that syntax-form mapping complexity elicits a cost shortly after meaning-syntax mapping complexity does, where the neural response is sensitive to context-dependent calculations of form and linear order. This is consistent with the time window discussed earlier in Section 5.3.6, particularly Clahsen et al. (2018), which observed a difference between the analytic and synthetic realizations of the comparative starting around 300 ms. In addition, if we assume that the "grammatical composition" process discussed by Sahin et al. (2009)

(argued to conclude before 350 ms) corresponds to the phonological realization of syntactic structures (the syntax-form mapping), rather than the building of syntactic structures themselves, then the present results are also consistent with those findings.

The later effect observed after 500 ms seems to be associated with phonological processing. In particular, it seems to be sensitive to a combination of the number of phonological words that are being produced and sustained in working memory, and the number of syllables in the utterance. The future tense involves the addition of the phonological word "will" (sometimes contracted with the subject, "he'll wash"), and the present progressive tense involves the addition of one or two syllables ("he is washing / he's washing"[4]), whereas the present simple only involves a single phonological word, usually with the same number of syllables ("he dries / washes", where verbs that end in sibilants have a vowel inserted before the /z/ suffix creating an additional syllable). Participants were instructed to form their responses to the questions in the way that felt most natural to them, allowing them to use contractions. With contractions, all three tense conditions involve the same number of phonological words, and only the present progressive condition or present simple verbs ending in sibilants involved an additional syllable. There was variation both across participants and across trials for whether they used contractions or not. Assuming that the rate of contraction was uniform across conditions, the observed results for the three tense conditions seems most consistent with a combination of the number of phonological words and the number of syllables associated with that tense condition, where the effect is driven by the trials which did not involve contractions.

The later effect does not seem to be driven solely by the number of phonological words or solely by the number of syllables, as another measure of phonological length, based on which

[4]Where "is" appearing sentence-medially is not considered an independent phonological word

tense conditions pattern together in the ERP response. Because the future tense is the only condition that involves an additional phonological word, if the number of phonological words alone was the main factor driving the effect, this would predict that the future tense would exhibit more negativity, while the present progressive and present simple tense conditions would pattern together; this is not observed. In addition, the future tense and the present simple tense often involve the same number of syllables (the future tense adds 0 or 1 syllables, with or without contractions: "he'll / he will wash the dishes"; the present simple tense adds 0 or 1 syllables, depending on the final consonant of the verb: "he dries / washes the dishes"), while the present progressive is usually longer (the present progressive adds 1 or 2 syllables: "he's / he is washing the dishes"). If the number of syllables alone was the main factor driving the later effect, we would expect to see a significant difference between the present progressive and the other conditions, while the future and present simple would exhibit a much smaller difference, or a three-way separation where the present progressive is most negative, the future simple is less negative, and the present simple is the least negative. Correlations between the amplitude of the ERP in this time window and the number of phonological words or syllables in the utterance may exist at the level of individual trials, however, even if it is not identifiable in the average waveforms; this should be investigated in future analyses.

5.5 Discussion

Language production has often been thought of in terms of retrieving, combining, and articulating "lemmas". As discussed earlier, this view of language production is problematic in two main ways: firstly, in the assumption that units of meaning, syntax, and form are stored in a single atomic representation, and secondly, in the assumption that accessing one meaning representation automatically and deterministically leads to the retrieval of the syntactic and form representations. The non-lexicalist framework described in this dissertation centers the mapping processes between different representations in language production, which can involve different levels of complexity depending on the amount and type of information that is used to translate between those different data structures.

The experiment described in this chapter tests some of the consequences of this approach: that mapping complexity should also modulate neural responses independent of the representations that are accessed. I use a novel experimental paradigm which elicits the production of simple sentences ("he is washing the dishes") while EEG signals are being recorded. By manipulating the complexity of the mapping between meaning and syntax in the target sentence, as well as the complexity of the mapping between syntax and form in the tense and aspect of the target utterance, the neural responses to the different conditions provide an index of the complexity of the two sets of mappings. As this is the first experiment design of its kind, there were not *a priori* time windows and regions of interest, so the data analysis was very exploratory in nature. This experiment observed an effect of meaning-syntax mapping complexity around the 220-280 ms time window post stimulus onset, as well as an effect of syntax-form mapping complexity around the 300-350 ms time window. In addition, a later time window starting after 500 ms exhibited an effect which appeared to correspond to the number of syllables and phonological words in the utterance, perhaps the result of holding those elements in phonological working memory prior to articulation. These results provide some evidence that this novel experimental paradigm is sufficient to elicit neural responses to mapping complexity, and shows promise for future EEG and MEG language production studies. Given the exploratory nature of the experiment and analysis, future work will need to validate this initial set of results with additional stimuli and with different linguistic phenomena that incur different kinds of complexity in either of the two mappings. The rest of this section will discuss the likely interpretation if these initial findings are reliable and replicable.

These results provide another perspective on non-compositional phenomena, expanding on the results from Wittenberg et al. (2014) and the studies on multi-word expressions discussed earlier. A key takeaway from these results is that the differences between non-compositional phrases and fully compositional phrases cannot be attributed to reanalysis or revising an underspecified meaning, because the speaker would know the target meaning of their utterance and would not be required to revise that meaning based on other disambiguating material, like they could in a comprehension task. This study also provides some insights into the Bhattasali et al. (2018) results, which suggested that multi-word expressions may be retrieved from memory as whole pieces, not necessarily engaging syntactic processes.

These results also provide another perspective on the processes involved in linearization and phonological realization. As discussed above, Clahsen et al. (2018) observed a difference in the neural responses for the synthetic realization of the comparative (*smart-er*) and the analytic realization of the comparative (*more intelligent*). Similarly, Experiment 2 exhibited a difference in the neural responses for the synthetic realization of tense (*walk-s*, *is walk-ing*) and the analytic realization of tense (*will walk*). In earlier discussions, the difference in Experiment 2 was predicted based on how much information is needed from the context in order to determine the phonological form of the tense morpheme, where the future tense involves the least amount of context (it does not need to refer to the verb to determine its form), whereas the present simple and present progressive tense/aspects both attach to the verb, modifying its form. This kind of framing is less helpful for the realization of the comparative, given that in both cases, the comparative morpheme needs to be sensitive to the form of the adjective it modifies, which would lead to the prediction that both realizations of the comparative should exhibit similar costs. However, by reframing the discussion around the realization of the adjective, where the realization of the adjective is dependent on whether the comparative attaches or not, the results start to make more sense. In the case of *smart-er*, the phonological graph of *smart* must be modified to take on the comparative suffix, whereas in *more intelligent*, neither phonological graph needs to be modified. Thus, the effect observed in Clahsen et al (2018) and Experiment 2 seems to be crucially related to modifying the phonological graphs based on the syntactic and phonological context, either adding the tense morphology or the comparative morpheme. This generates some interesting predictions for polysynthetic languages, which would presumably need to involve many instances where phonological graphs are modified and combined.

Several of the predictions made by the WOW model in the context of this experiment would also be consistent with those of a lemma- or word-based model of language production (for example, the effect in the 220-280 ms time window argued to be associated with meaning-syntax mapping complexity could also be caused by the retrieval of an additional "idiom lemma"). Thus, this experiment alone cannot be seen as evidence to "disprove lexicalism". Such evidence already exists, as detailed in Chapter 2 and Chapter 3. One of the predictions of a lemma-based model that this experiment does challenge is in regard to the neural response based on the number of phonological words in the utterance. According to a lemma- or word-based model, the production of a wordform can only occur if the appropriate lexical concept was retrieved to activate the corresponding lemma. This would suggest that the future tense condition and the present progressive condition should both incur a greater neural response compared to the present simple tense condition (because they involve two wordforms, rather than one inflected one), even at early stages in language production, because an additional lemma would need to be activated compared to the simple tense condition. According to this view, rather than exhibiting an effect of the number of phonological words only in the late time window (after 500 ms), that effect should also appear in earlier time windows. Because it does not, appearing only after the time window which appears to reflect meaning-syntax mapping complexity and after the time window which appears to reflect syntax-form mapping complexity, this suggests that the language production system is not sensitive to the number of phonological words in the utterance until later in the language production process - and not during the earlier stages of conceptual/semantic processing or syntactic processing.

Another account of the contrast in Experiment 1 could suggest that the observed difference in neural responses arises because, in the Verbal Idiom condition, the producer only needs to retrieve one stored complex piece of syntax from memory to produce that phrase, whereas in the Simple Transitive condition, they must retrieve multiple pieces of syntax from memory and combine them on-line. This kind of account would suggest that the direction of the difference is different, where the greater negativity observed for the Verbal Idiom condition is actually the result of a smaller processing cost compared to the Simple Transitive condition, rather than a greater processing cost caused by greater mapping complexity. Without specifically controlling for this, however, it is not possible to draw any conclusions based on the direction of the effect, given that the electrical fields measured in EEG are affected by the direction and location of the neural activity, multiple electrical fields can overlap in space, and source localization in EEG is an inverse computation problem that has many possible solutions. In addition, based on the discussion about MWEs, the likelihood that a complex piece of syntax is stored in memory as a whole is conditioned by frequency, which would likely affect the neural responses for both the Verbal Idioms as well as the Simple Transitives, since both sets of items include highly frequent phrases. To summarize, this experiment does not provide any evidence to disambiguate between a mapping complexity account and a “storage-based” account.

Conclusion

In contrast to standard lemma-based models, the key mechanisms of the WOW model in this non-lexicalist framework are mapping processes that link separate representations of meaning, syntax, and form. In this novel experimental paradigm, EEG signals are recorded during elicited sentence production. This study finds an effect of meaning-syntax mapping complexity around the 220-280 ms time window, and an effect of syntax-form mapping complexity around the 300-350 ms time window. The results of this study provide important insights into the role of mapping complexity in language production, allowing us to capture a broader view of what processes are involved in producing speech or sign. The neural signatures which index the complexity of these mappings identified in this study should be replicated and tested in other languages in order to confirm the results and develop cross-linguistic validity.

Chapter 6

Modeling inflection deficits in agrammatic aphasia with a recurrent neural network

The WOW model characterizes language processing as a series of transformations between data structures. Consequently, this kind of approach is very compatible with a connectionist architecture, where the mappings between independent symbolic data structures are performed in a distributed, continuous manner (Smolensky, 2012; Smolensky and Legendre, 2006). The work in this chapter constitutes a first step toward a connectionist implementation of the language network in a non-lexicalist framework, shedding light on the kinds of factors that may be relevant in conditioning language processing, as well as how those processes can be impacted in acquired language disorders.

One of the key ways that lexicalist theories of language are different from the WOW model is in the distinction between the grammar and the lexicon, and the kinds of phenomena that are considered "syntactic" or not. For example, lexicalist theories have often framed regular verb inflections as a rule-based operation that is stored in the grammar, whereas irregular verb inflections must be stored in the lexicon, and so processing or producing these forms must involve two separate mechanisms, a "dual-route" model. This distinction has been supported by clinical data, which includes several case studies that seem to show a double dissociation between regular and irregular verbs in people with opposite lesion patterns. The type of aphasia associated with deficits to the regular rule-based inflections has been labeled as “agrammatic” aphasia (also referred to as “nonfluent” or “Broca’s” aphasia; (Ullman et al., 1997; Caramazza et al., 1981, a.o.). There are several properties of agrammatic speech which lend itself to such an analysis; people with agrammatic aphasia tend to produce speech which lacks inflectional morphology and function words, some of the overt indicators of syntactic operations, as well as particular deficits for verbs (Cho-Reyes and Thompson, 2012). Based on this, it does seem that the pattern of deficits is sensitive to syntactic properties. However, in a non-lexicalist framework, regular and irregular verbs have the same syntactic configurations, but critically only differ in terms of how predictable their inflected form is based on phonological patterns. Furthermore, given the independence of meaning, syntax, and form, the surface forms that are produced may not necessarily reflect a deficit in the syntax itself, but in some other process which disproportionately impacts the phonological realization of certain syntactic features, or the manipulation of phonological objects.

This is part of a decades-long debate in cognitive science about how people represent regular and irregular inflectional morphology (Clahsen et al., 1992; Goebel and Indefrey, 2000; Hahn and Nakisa, 2000; Hare et al., 1995; Laaha et al., 2006; Lachter and Bever, 1988; MacWhinney and Leinbach, 1991; Marcus et al., 1995; McClelland and Patterson, 2002; Plunkett and Marchman, 1991; Plunkett and Juola, 1999; Pinker and Prince, 1988a; Pinker, 1998; Pinker and Ullman, 2002; Rumelhart and McClelland, 1986). Whereas proponents of the dual-route model have argued that separate mechanisms are responsible for the regular and irregular inflections, proponents of the “single-route” model have argued that the phonological relationship between inflected and uninflected forms can be captured within a single pattern associator. Recently, state-of-the-art single-route neural models have been able to achieve high degrees of accuracy when learning the English past tense (Kirov and Cotterell, 2018), addressing many of the initial issues that previous single-route models faced (with some limitations; see McCurdy et al., 2020). However, the dissociation between regular and irregular inflections in people with different types of aphasia has been a key sticking point for the single-route approach.

Whereas previous work on the single-route model has cast it in opposition to the traditional idea of grammatical rules in linguistics, I pursue a different interpretation in our discussion of these results, based on recent advances in linguistic theory. I situate the single-route model as performing a particular function that any traditional model of grammar would need to carry out: converting the underlying abstract hierarchical syntactic structure into a linear order and making calculations over phonological space to determine its form. In the WOW model, this is a calculation which translates between syntactic structures and phonological forms, while in Distributed Morphology, this is the process of "Vocabulary Insertion" (Halle and Marantz, 1993). Part of the goal of this chapter is to reconcile the Past Tense debate with more modern, non-lexicalist syntactic theory, and to come to a better understanding of how verb inflections are represented in the brain and how they are impacted by brain injury. I argue that agrammatic aphasia should be approached as a deficit not in syntactic operations, but in transformations over phonological objects.

In this chapter, I begin by reviewing the evidence from aphasia, discussing empirical data as well as previous models. I then give an overview of the single-route model from Kirov and Cotterell (2018) which has been used to learn the past tense inflection, which I adopt in my simulations. The next section describes the results of simulations in English, German, and Dutch, showing that manipulating certain factors in the model's training data can modulate how well regular and irregular verb forms are preserved after lesioning, and that a primary factor which conditions form preservation after lesioning is the relative frequency of individual verb forms in the language. This kind of model can better account for the cross-linguistic patterns of deficits in agrammatic aphasia, giving a simpler explanation of the variation between speakers of different languages. These findings contribute to the ongoing past tense debate in cognitive science, providing an alternate single-route interpretation for one of the key pieces of evidence in support of the dual-route model, while also providing some support for the WOW model's interpretation of the clinical data. I conclude by discussing the relationship between the past tense debate and linguistic theory in light of substantial recent advances in both fields.

6.1 Morphological deficits in aphasia

As an extension of contemporary models of memory circuits, Ullman (2004) developed the declarative/procedural model to connect language to more domain-general cognitive processes. This theory suggests that the grammar is a subdivision of procedural memory, like other motor and cognitive skills, while the lexicon is a subdivision of declarative memory, which stores arbitrary relations, facts, and events. Consequently, this type of dual-route model predicts that regular and irregular forms should be served by different neural substrates associated with the different types of memory - procedural memory corresponding to the basal ganglia and the frontal cortex, and declarative memory corresponding to the hippocampus and temporo-parietal regions. Data from language disorders such as aphasia can provide exactly the kind of evidence needed to test this hypothesis. Aphasia is an acquired language disorder that occurs as result of a brain injury - such as a stroke, tumor, or head injury - that impacts areas of the brain associated with language.

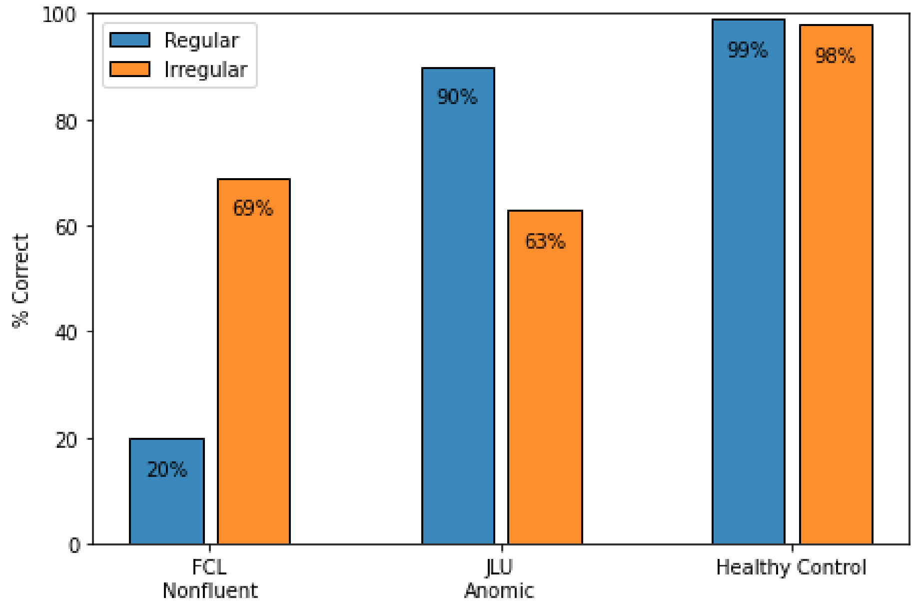

Figure 6.1: Patients FCL and JLU are used as evidence of a dissociation between regular and irregular inflections. FCL exhibits left anterior perisylvian lesions and a greater deficit for regular inflections, while JLU exhibits left temporo-parietal lesions and a greater deficit for irregular inflections (Ullman et al., 1997).

There are many varieties of aphasia, but the two varieties relevant to the current discussion are fluent aphasia (also known as "Wernicke's aphasia") and agrammatic aphasia (also referred to as "nonfluent" or "Broca's" aphasia). The agrammatic variety of aphasia is characterized by "telegraphic speech", which tends to include few function words and morphology, so it has often been analyzed as a syntactic deficit. Meanwhile, the fluent variety of aphasia often involves "empty speech", syntactically well-formed sentences that lack meaning or message, so it has often been analyzed as a "lexical" deficit. Though a brain injury is not guaranteed to be isolated to one area of the brain, analysis of lesion patterns and behavioral deficits can provide convincing evidence for functional dissociations.

Ullman et al. (1997) identified two patients, FCL and JLU, with opposite lesion patterns, and tested their ability to produce the past tense form of regular and irregular verbs. Their results are shown in Figure 6.1. FCL, who had agrammatic aphasia after an injury to the left anterior perisylvian region, produced the regular past tense with only 20% accuracy, while producing the irregular past tense with 69% accuracy. In contrast, JLU, who had fluent aphasia after injury to the left temporo-parietal area, produced regular verbs with 90% accuracy, and produced irregular verbs with 63% accuracy. These two patients are the key evidence for the proposed double dissociation between regular and irregular inflections. They show opposite patterns of accuracy rates for regulars and irregulars, as would be expected under the dual-route model. Based on these data points alone, there does not seem to be an obvious explanation that would support the single-route account.

There are several reasons why these data should be reexamined. First, whereas the rhetoric surrounding the original past tense debate characterized dual-route models as more consistent with linguistic theory, this has changed due to advances in linguistic theory. The Pinker and Prince (1988b) model was built in a lexicalist framework, which assumes that syntactic operations do not extend below the word level, while also treating all patterns that are descriptively rule-based (such as regular past tense morphology) as part of the syntax. This assumption has been challenged in the past 40 years, especially with the development of theoretical frameworks such as Distributed Morphology and the non-lexicalist approach adopted in the WOW model. These kinds of theories argue that words and phrases are built using the same syntactic processes, and functional morphemes such as the English past tense do not have a "built-in" phonological form (Embick, 2015). In Distributed Morphology, their form is instead supplied in the post-syntactic operation of Vocabulary Insertion; in the WOW model, the form is provided by a mapping from sets of syntactic units to sets of phonological units. According to these views, from a syntactic perspective, the verbs "ran" and "walked" do not differ in terms of their hierar-

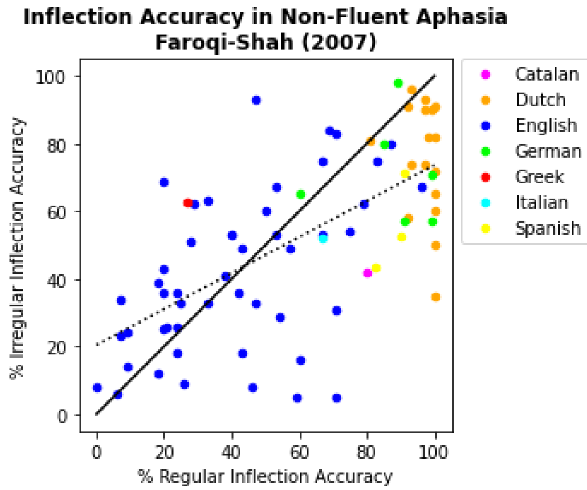

Figure 6.2: Plot summarizing the inflection accuracy of participants in the studies cited in the Faroqi-Shah (2007) analysis of agrammatic aphasia (n=66). The solid line corresponds to equal accuracy on regular and irregular inflections. The dotted line corresponds to the line of best fit.

chical structure—both involve a verb and the past tense morpheme combined in a particular way. The only difference is how that past tense morpheme conditions the form of the output, and how complex and frequent those transformations may be.

Second, studies of agrammatic aphasia cross-linguistically suggest that the pattern of deficits even in one variety of aphasia seems to be more nuanced, and not entirely consistent with the predictions of the dual-route model. Faroqi-Shah (2007)'s meta-analysis of agrammatic aphasia across seven different languages compared participants' performance on the regular and irregular past tense inflection. The analysis included 25 studies published between 1980 and 2006. The resulting dataset contains 66 different participants, and includes speakers of Catalan, Dutch, English, German, Greek, Italian, and Spanish. Figure 6.2 summarizes the participants' performance (line of best fit: $y = 0.53x + 20$, $R^2 = 0.429$).

According to the dual-route model, agrammatic aphasia should be an impairment in the grammar, and thus a selective deficit for regular verbs. If this were the case, Figure 6.2 would show many more participants - from all language groups - performing at or near ceiling for the irregular verbs, with varying degrees of accuracy on regular verbs (i.e., the line of best fit would be at the top of the plot). This would mean more points near the top of the plot or above the solid equivalence line, and few below the line. Instead, this shows a large number of German- and Dutch-speaking participants that have high accuracy on regular verbs but varying accuracy on irregular verbs, as well as a number of English-speaking participants with better accuracy on regular verbs than irregular verbs. Across all individuals tested, performance on the two groups of inflections is somewhat correlated, with an r^2 value approaching 0.5 (i.e., roughly half of the variation in accuracy on irregular inflections can be accounted for by variation in the accuracy on regular inflections). The "dissociation", consequently, is less straightforward than Pinker and Ullman (2002) suggested, especially cross-linguistically; these data instead suggest that the patterns of deficits observed across languages could be better understood as consequences of the unique properties of the past tense forms in each language, such as frequency distributions and phonological patterns, rather than a universal distinction between "regular" and "irregular" forms.

These data present a puzzle. It is not only challenging for the dual-route model, because it does not present an obvious selective deficit for regular inflections, but also for the associated analyses of agrammatic aphasia. The WOW model suggests that the brain area implicated in agrammatic aphasia, the IFG, is involved in post-syntactic prosodic and control operations, instead of rule-based syntactic operations. This model suggests that the pattern of deficits observed for verb inflections in agrammatic aphasia may not be caused by a syntactic deficit, but rather a phonological one, where one set of verbs can be disproportionately impacted over the another based on phonological differences.

If the deficit is phonological in nature, rather than a consequence of a dissociation between the grammar and the lexicon, patterns of deficits in aphasia could be the result of an interaction between irregularity, the complexity of the phonological transformation, and frequency, or a number of other factors. For example, whereas regular verb inflection in English can make word-final consonant clusters more complex (as in *walk* $\sim$ *walked*) or add a syllable (*load* $\sim$ *loaded*), as discussed by Bird et al. (2003). Irregular verbs can also vary in the complexity of the transformation: *sing* only changes one vowel sound, whereas *be* and *go* undergo suppletion. Furthermore, irregular verbs tend to have higher frequencies than regular verbs, partly due to factors involved in language change (Bybee, 1995; Lieberman et al., 2007). Because every language involves different kinds of phonological transformations with varying complexity, and with varying frequency distributions, agrammatic aphasia may result in different patterns of deficits cross-linguistically based on these properties, even if the lesion pattern is similar across patients. This kind of explanation could account for the differences between language groups in Faroqi-Shah (2007), where the dual-route model cannot.

Recent work revisiting the original past tense debate using state-of-the-art neural models, such as the Encoder-Decoder (ED) network used by Kirov and Cotterell (2018), has addressed many of the initial critiques that were directed at the Rumelhart and McClelland (1986) model. Specifically, they showed that a model trained on English inflections exhibits high accuracy rates, produces human-like novel forms, and mirrors some child-like learning patterns. It can also achieve high accuracy rates when trained on multiple inflection classes such as the gerund, past participle, and third-person singular forms. The success of this neural model suggests that a single-route model is sufficient for both regular and irregular inflections, but it does not explain how the aphasia data discussed above might arise.

Agrammatic aphasia has been modeled in connectionist networks before by Penke and Westermann (2006), which simulated the pattern of deficits for German-speakers. This model was trained to classify German verbs into different inflection types based on their form, and then lesioned to simulate agrammatic aphasia. However, their model was not fully homogeneous, allowing for a direct connection between the input and output layers as well as connections through a hidden layer. Thus, it is not clear if it provides conclusive evidence for a single-route account, if regular and irregular verbs are ultimately able to utilize different pathways through the network. Furthermore, though the input to the model was a phonological representation of the German verb stem, the output of the model was a classification that determined how the participle might be formed given the verb stem. The model was not required to produce a phonological form as its output. As mentioned previously, German people with agrammatic aphasia tend to exhibit greater deficits for irregular inflections over regular inflections, so the goal of the Penke and Westermann (2006) model was to achieve results that exhibited that pattern; it is not clear whether their model would be able to generalize to other languages. Though the model was able to replicate the German results, it is not clear that it provides conclusive evidence for a single route account. Given that there have been substantial advancements in the architecture of single-route models, it is possible to develop a homogeneous network which does not allow direct mappings between input and output layers, and which produces a phonological form as its output rather than just a classification.

Given this background, if agrammatic aphasia is not a deficit in rule-based syntactic operations, but in the phonological transformations involved in inflection, then it should be possible to simulate the effects of agrammatic aphasia in a single-route neural model that performs that kind of phonological transformation. In this chapter, I demonstrate that a state-of-the-art single-route model can be used to account for the pattern of deficits in regular and irregular past tense forms observed in agrammatic aphasia. In Simulation 1, as a proof of concept, I show that training an Encoder-Decoder network on datasets with different frequency distributions leads to greater effects of lesioning in one set of verbs. In Simulation 2, I show that a model trained on an English-like dataset using perplexity-based sampling can lead to a more English-like pattern of deficits. As predicted by the single-route approach, the dissociation can be captured without appealing to separate lexical and grammatical mechanisms. In Simulations 3 and 4, I test two other languages that were included in the Faroqi-Shah (2007) meta-analysis, German and Dutch, showing that the model can also capture cross-linguistic patterns in agrammatic aphasia.

Despite the differences between recurrent neural network models and the actual neural architecture of the IFG, this kind of simulation provides several key benefits in understanding the properties of agrammatic aphasia, compared to clinical observations. First, the training data for the model can be completely controlled and manipulated in a number of ways, allowing me to see how factors such as frequency can influence form preservation after lesioning. In contrast, it is not possible to fully control for language experience in human subjects or to experimentally manipulate the properties of their language, and it is extremely difficult to collect language performance data from participants before their stroke to perform baseline testing. In addition, this kind of model can be lesioned with precision, in such a way that no other linguistic processes are impacted (lesions in the human brain are rarely isolated to a single brain area) and that no other processes are available to "rescue" the target function (in humans, other brain areas may be able to compensate for the lesioned brain areas). Thus, based on what is included - or not included - in the training data, we can draw conclusions about which linguistic representations or distinctions are necessary to account for the clinical observations, and the factors that condition whether a form is more or less likely to be preserved after the network is damaged. The model architecture used in these simulations is the same as the one used by Kirov and Cotterell (2018), which has been shown to learn the English past tense with a high degree of accuracy, though any single-route model architecture that can successfully learn the past tense should in theory be sufficient for the current purposes; the key manipulation in these simulations is the composition of the training datasets, so the specifics of the model architecture may not be consequential in evaluating these hypotheses.

6.2 The Model

The model used in the present study is a homogeneous, single-route, single-mechanism model that - when lesioned - can simulate the pattern of deficits observed in agrammatic aphasia in English, German, and Dutch. In doing so, this model indicates that the dissociation that has been a key piece of evidence for the Dual-Route model does not, in principle, require multiple mechanisms to capture this particular pattern of deficits.

This study involves four simulations. Simulation 1 trains models on English datasets with different proportions of regular and irregular verbs. Simulation 2 looks at the effect of using perplexity—essentially, the model's degree of surprise at a particular data point—to constrain learning. Simulation 3 trains models on German datasets, and Simulation 4 trains models on Dutch datasets, both with different proportions of regulars and irregulars.

6.2.1 Architecture

Our models use the Encoder-Decoder network architecture specified by Kirov and Cotterell (2018), which includes the architecture described by Bahdanau et al. (2014) and hyperparameters set by Kann and Schütze (2016). This network operates over strings of characters which represent phonemes. The encoder includes a bidirectional LSTM with two layers. Each character has an embedding size of 500 units. The dropout value between layers is 0.3. The encoder and decoder have 100 hidden units each. The Adadelta training procedure (Zeiler, 2012) was used, with a learning rate of 1.0 and a minibatch size of 64. The decoder uses a beam search (k=12). The model was implemented using OpenNMT-py 3.0 (Klein et al., 2020).

6.2.2 Training Data

The English, German, and Dutch datasets were all developed from the CELEX database (Baayen et al., 1996). The English base dataset includes 3,882 English verbs, 146 of which are irregular (3.8% of the dataset). The German base dataset includes 3,341 German verbs, 953 of which are irregular (28.5% of the dataset). The Dutch base dataset includes 2,850 Dutch verbs, 871 of which are irregular (30.6% of the dataset). All of the items had a frequency greater than 0 per million in their respective corpora. The inputs to the model - the present tense form of the verb - and the target outputs - the past tense or past participle form - were formatted as a sequence of characters representing the phonemes of those forms, so each phoneme was presented to the model individually in order.

For each language, 20% of the dataset was held out for validation. The testing dataset included 80 regular verbs and 80 irregular verbs which were all seen in training. The model was only tested on verbs that appeared in training, rather than a held-out portion of the dataset, since I am more interested in the model's performance after lesioning than how accurately it can generalize to new forms.

For Simulations 1 and 2, three datasets were developed from the English dataset. The

Regular-Bias dataset has each verb in the corpus appear exactly once, so that only 3.8% of the verbs in that training dataset are irregular. The **Proportional** dataset had each verb appear proportional to its COBUILD frequency, so the dataset was composed of about 30% irregular verbs. In the **Irregular-Bias** dataset, a subset of high-frequency regular verbs were selected from the base dataset, and these appear only once; the remaining 141 irregular verbs are repeated proportional to the inflected forms' COBUILD frequency, in such a way that the resulting dataset is composed of 5% regular verbs and 95% irregular verbs.

For simulations 3 and 4, two datasets were developed each for German and Dutch. The first was the **Regular-Bias** dataset, with each verb in the corpus appearing exactly once, so 28.5% of the German Regular-Bias dataset was irregular, and 30.6% of the Dutch Regular-Bias dataset was irregular. The second was the **Proportional** dataset, where each verb appears proportional to its corpus frequency (the German frequencies were gathered from the Mannheim German Reference Corpus (Lüngen, 2017), and the Dutch frequencies were gathered from the Instituut voor Nederlandse Lexicologie (INL) corpus (Kruyt and Dutilh, 2012), accessed from CELEX). The German Proportional dataset had 38.5% irregular verbs, and the Dutch Proportional dataset had 40.3% irregular verbs.

Within each language group, all of the models were trained for the same number of train steps, equivalent to 100 epochs of training on the largest dataset for that language, so that each one was exposed to the same amount of training data despite the differences in the sizes of the training datasets.

6.2.3 Lesioning

After training, each model was "lesioned" by randomly resetting connection weights to 0. The proportion of weights that were reset ranged from 5% to 100% by increments of 5%. After lesioning, the models' accuracy on the test dataset was measured. This was repeated 5 times for each model in order to observe the effect with different random seeds.

During a stroke or brain injury, lesions are not randomly distributed across a brain area, but instead localized to a set of adjacent neurons. In the neural network, there is no sense of "adjacency"; all of the neurons in one layer are connected to all of the neurons in the neighboring layers. I assume that randomly lesioning the neural network provides a reasonable approximation of the effect of a brain injury, despite the differences between neural architectures and neural networks.

6.3 Simulation 1: English

In Simulation 1, one model was trained for each of the three datasets, reaching an accuracy around 93% prior to lesioning. Each was lesioned randomly 5 times at 5% intervals. The earlier discussion leads to several predictions. Firstly, if any of the single-route models can simulate a dissociation similar to what was observed for FCL (Ullman et al., 1997), this would provide a proof in principle that the dissociation can be captured without requiring multiple distinct mechanisms. Secondly, if models that observe different proportions of regulars and irregulars in their training data exhibit different behavior after lesioning, this would suggest that the frequency distribution of regulars and irregulars impacts how those transformations are encoded, and thus, influences how well those forms are preserved after lesioning. This would also suggest that the cross-linguistic differences observed by Faroqi-Shah (2007) might depend on such factors.

As shown in Figure 6.3, irregular verbs make up a relatively small percentage of the English dataset by count, but because they have a much greater frequency than regular verbs on average, irregular verbs make up nearly 30% of the dataset when the verbs are repeated proportionally to their corpus frequency.

Figure 6.4 shows the performance of each of the models. These plots compare accuracy on regular and irregular inflections. Each point represents a different instance of the model, with different proportions of connections lesioned, and with different random seeds. The black diagonal line represents equal performance on both sets of verbs, so points above the line represent models performing better on irregular verbs, while points below the line represent models performing better on regular verbs.

For the model that was trained on the Regular-Bias dataset, the performance was consistently better on regular inflections than irregular inflections. The model trained on the Proportional dataset shows better performance on irregular inflections than the Regular-Bias model, but most points still fall below the line of equivalence. For the model that was trained on the

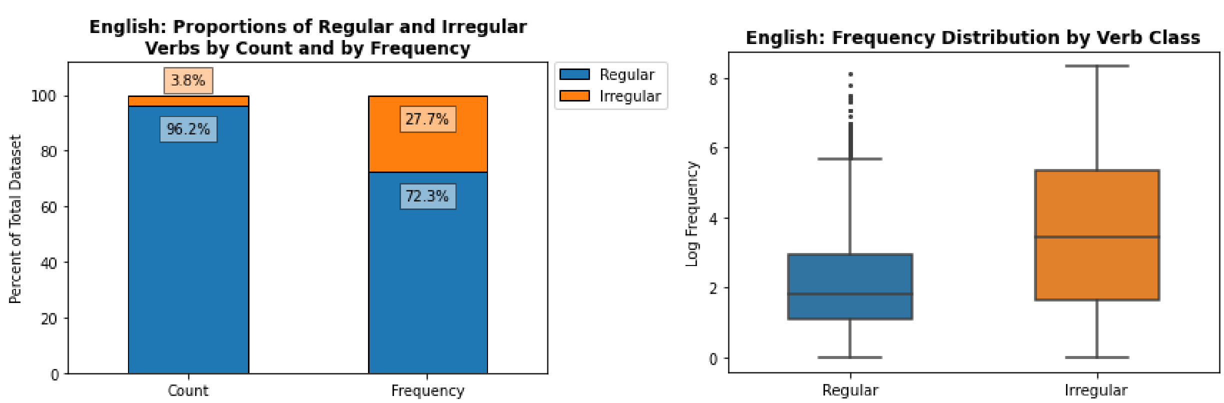

Figure 6.3: The proportions of English verb classes in the corpus by type frequency and token frequency (left), and the log frequency distribution of regular and irregular verbs (right)

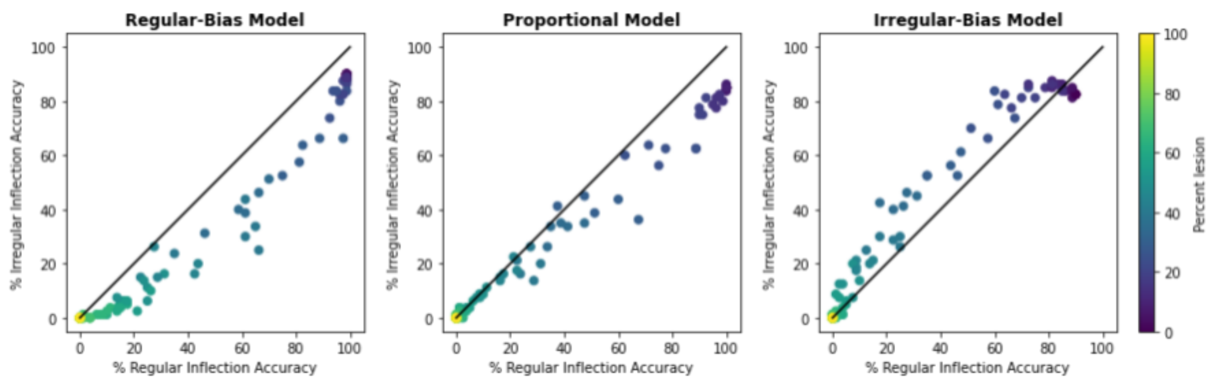

Figure 6.4: Results from Simulation 1. **Left:** Plot summarizing the performance of the model trained on the Regular-Bias dataset, where less than 5% of the dataset is made up of irregular verbs. **Center:** Plot summarizing the performance of the model trained on the Proportional dataset, where 26% of the dataset is made up of irregular verbs. **Right:** Plot summarizing the performance of the Irregular-Bias model, where 95% of the dataset is made up of irregular verbs.

Irregular-Bias dataset, the model performs better at irregular verbs than regular verbs at nearly every point. This simulation demonstrates that it is possible to achieve a range of outcomes with a model trained on regular and irregular verb inflections without appealing to a grammatical or lexical distinction.

This simulation shows that the frequency distribution of the training dataset does play a role in the pattern of deficits observed after lesioning the model, suggesting that this factor could be partially responsible for the cross-linguistic differences observed by Faroqi-Shah (2007). Although the lesioning did not quantitatively replicate the proportions of regular and irregular verbs produced correctly by FCL, it did show that a single-route model can produce qualitative asymmetries. The fact that the asymmetries in the model do not quantitatively resemble FCL may be due in part to the fact that FCL is an outlier in the clinical data, and because the models do not generate much variety within the groups of lesioned models in terms of accuracy; the points on the plot are relatively closely grouped together. We return to this point in the discussion.

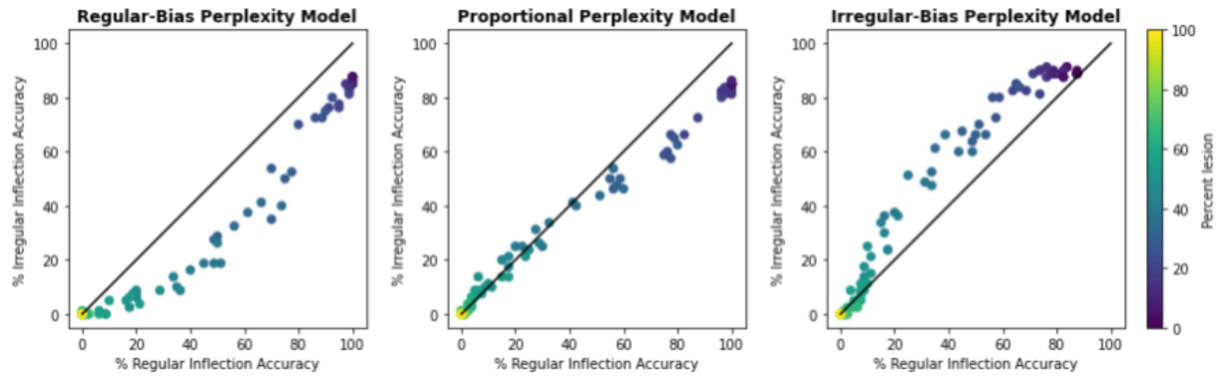

Figure 6.5: Results from Simulation 2. **Left:** Plot summarizing the performance of the model trained on the Regular-Bias dataset, where less than 5% of the dataset is made up of irregular verbs. **Center:** Plot summarizing the performance of the model trained on the Proportional dataset, where 26% of the dataset is made up of irregular verbs. **Right:** Plot summarizing the performance of the Irregular-Bias model, where 95% of the dataset is made up of irregular verbs.

6.4 Simulation 2: English

In Simulation 2, a second set of models were created that used perplexity-based sampling (Fernandez and Downey, 2018). One model was trained on each of the three datasets.

Perplexity-based sampling involves calculating a "perplexity score" for each item in the training dataset, which measures how well the model was able to predict the output form given its current state. The model then selects the items with the highest score for the next training step. This simulates the attention, surprisal, and reanalysis that a learner may experience when presented with a form that was not predicted (Clark, 2013), used here to simulate a slightly more cognitively plausible learning method. If models trained using perplexity-based sampling better capture the empirical data than models trained with random sampling, this would suggest that factors such as surprisal, attention, or reanalysis can play a role in shaping neural representations and encoding.

The results for Simulation 2 are shown in Figure 6.5. reaching an accuracy around 93%

prior to lesioning. Each was lesioned randomly 5 times at 5% intervals. The model trained on the Regular-Bias dataset with perplexity-based sampling exhibited similar performance to the Regular-Bias model in Simulation 1, where most of the points fell below the solid line. The model trained on the English-like Proportional dataset with perplexity-based sampling falls close to the equivalence line. The difference between this model and the equivalent one in Simulation 1 is slight, but closer inspection shows that the model trained with perplexity-based sampling has higher accuracy on both types of inflections compared to the model trained with random sampling when matched on the proportion of lesions, but with a greater impact on irregular verbs, as shown in Figure 6.6. This suggests that perplexity sampling may not only help the model's baseline performance to improve, but also impact how well those forms are preserved after lesioning. Again, though this would not be able to capture patient FCL necessarily, the English speaking group shown in Faroqi-Shah (2007) was roughly centered around the equivalence line, especially at higher rates of lesioning. Lastly, the model trained on the Irregular-Bias dataset with perplexity-based sampling was once again almost completely above the equivalence line, performing better at irregulars than regulars at every point. The pattern observed for the Proportional model in Simulation 2 suggests that perplexity-based sampling does shift it closer to the pattern observed for the Irregular-Bias model, to a degree. In addition, the difference in performance between these three models in Simulation 2 indicates that the frequency distribution of regulars and irregulars has an impact for these models even in the context of perplexity-based sampling.

Overall, both simulations show that a single-route model can capture dissociations between regulars and irregulars, and that this relationship depends on specific characteristics of the training data. Based on the complexity and the frequency of the transformation, a network may encode

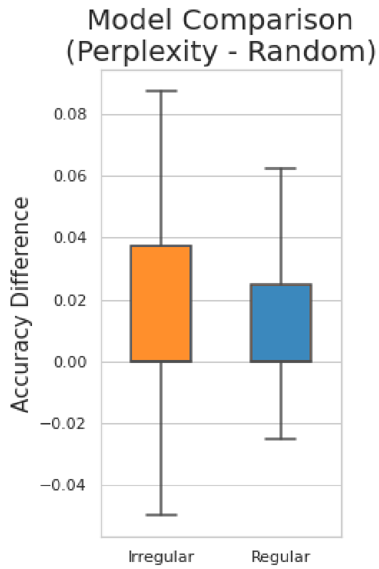

Figure 6.6: Perplexity-based sampling had a positive impact on both groups of verbs, but seems to impact irregulars more than regulars

the two verb classes differently, and this encoding can then affect performance both before and after lesioning. For example, a more complex transformation may lead to lower accuracy rates in an non-lesioned network, but if it has a more distributed representation it may be more resilient to lesions. The model trained on a dataset biased toward irregulars illustrates this effect using a dataset based on English. If the model were trained on inflections in another language, with different frequency distributions and different degrees of complexity, we might observe a different pattern of results. Furthermore, the degree of "perplexity" that the learner experiences as they encounter new forms may vary across languages, depending on the patterns of sub-regularity. In this way, patterns that vary across language groups can arise not because of representational differences, but because of differences in the complexity and frequency of the transformations.

6.5 Simulation 3: German

All languages are likely to have proportions of regular and irregular verbs somewhere between the two extremes of the English Regular-Bias and Irregular-Bias datasets, and so their pattern of deficits may be predicted - in part - based on that. The next two simulations test two additional languages, German and Dutch, to see how frequency distributions, proportions of verb types, and transformation complexity can condition form preservation after lesioning. Although the three languages are typologically similar, as Germanic languages, the distribution of the different groups of speakers in the clinical data are quite different, as seen in Figure 6.2, where the Dutch speakers cluster on the far right of the plot (perfect or near-perfect accuracy on regulars with varying accuracy on irregulars), and where the German speakers occupy a middle position between the English speakers and the Dutch speakers. This model's ability to capture patient data in multiple languages provides further support for the current account of agrammatic/nonfluent aphasia which this model presents.

As shown in the table below, the German past participle can be put into two groups, the regular "weak" verbs, and the irregular "strong" verbs. The regular past participle can be formed from the infinitive form by removing the *-en* suffix, leaving behind the verb stem, adding the prefix *ge-* to the beginning of the verb stem, and adding the suffix *-t* to the end ("ge-kauf-t"). Many verbs in German have prefixes, some of those prefixes are separable[1], and some of which are inseparable. For the separable verbs, the *ge-* prefix from the past tense is inserted between the stem prefix and the verb ("ein-ge-kauf-t"). The inseparable prefixes, on the other hand, are able to replace the *ge-* prefix altogether; these are still considered regular despite not having the *ge-* prefix ("be-such-t"). Additionally, there is another set of regular verbs which end in *-ieren* that omit the *ge-* prefix.

The irregular "strong" verbs can be irregular in several ways. Many of these verbs' partici-

[1] For example, the prefix ein- in the verb "einkaufen" (to shop) can appear separately from the verb, such as when it is tensed; "die Kunden **kaufen** im Laden **ein**", "the customers shop in the store".

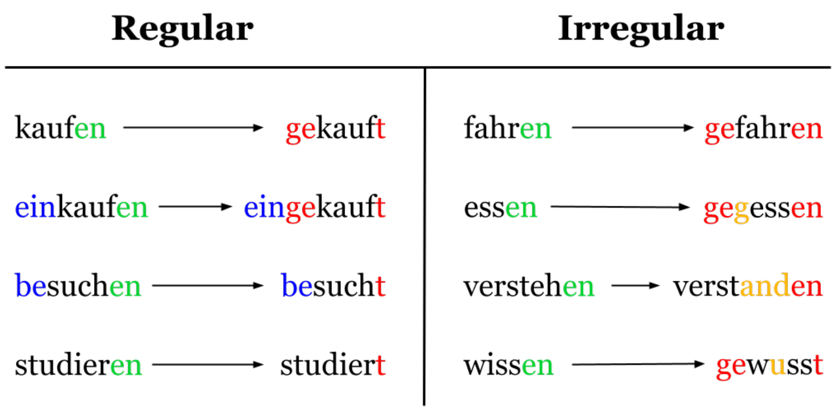

Figure 6.7: Verb classes in German (green = infinitive, blue = prefix, red = past participle, orange = stem changes

ples end in *-en* rather than *-t*, and may have stem changes (the stem *versteh* changes to *verstand* in *verstanden*). Mixed verbs have the regular *ge-* prefix and *-t* suffix, but exhibit stem changes (the stem *wiss* changes to *wuss* in *gewusst*). Many irregular verbs have unique patterns in generating the past participle, but some patterns of subregularity can be identified across verbs.

In the base dataset of German verbs, there are 3,341 items, 953 of which are irregular, which constitutes about 28.5% of the dataset[2]. However, because irregular verbs tend to be more frequent than regular verbs, irregular verbs constitute about 38.5% of the summed frequencies of the verbs in the dataset. These proportions are much greater than in English, which had about 5% irregulars by count, and 26% irregulars by frequency. Based on the previous simulations, with phonological properties being held equal, considering the proportion of irregulars alone would lead to the prediction that both the Regular-Bias and Proportional models in German should have results similar to the Proportional or Irregular-Bias models in English. However, given that the

[2]Many verbs were not included in the training dataset because their corpus frequency was equal to 0 per million words. The dataset including these verbs would have 9,185 items, 1,928 of which are irregular, constituting 21% of the total dataset. Excluding verbs with 0 frequencies did remove more regular verbs than irregular verbs from the training dataset, but this only impacts the Regular-Bias dataset where each verb appears once; those verbs would not have appeared in the Proportional dataset even if they had been included because their frequency was equal to 0.

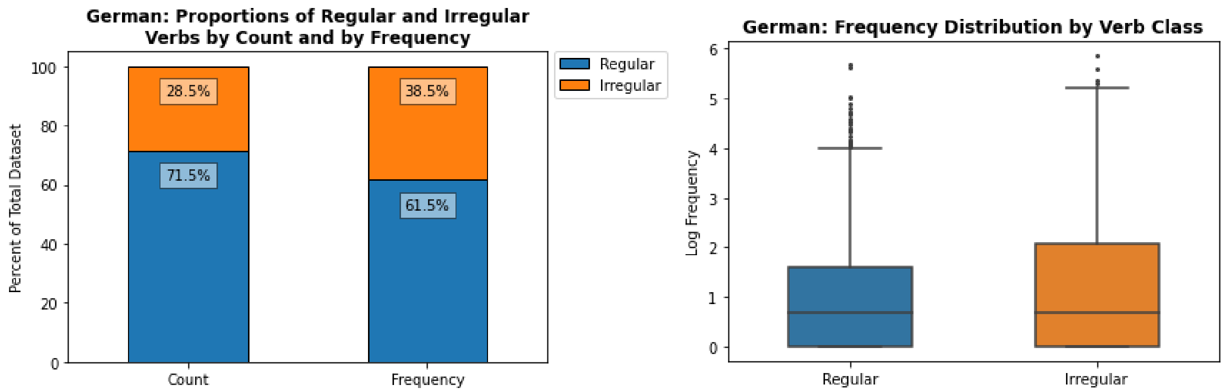

Figure 6.8: The proportions of German verb classes by type frequency and token frequency (left), and the log frequency distribution of regular and irregular verbs (right)

average frequency of irregular verbs in German is proportionally lower than irregular verbs in English (5% of English verbs make up 26% of the total frequency; 28.5% of the German verbs make up 38.5% of the total frequency), this may lead to the prediction that the Proportional model in German should perform more like the Regular-Bias model in English, if the frequencies of the individual verbs (the "token frequency") is more influential in form preservation than the frequency of the whole verb class (the "type frequency").

Five models were trained on each dataset using perplexity-based sampling, in order to assess if the randomness in the model training could generate a wider spread of accuracy as observed in the clinical data. The models reached an average of 96.6% accuracy for all verbs prior to lesioning (the Regular-Bias models had an average of 97.8% accuracy on irregular verbs and 95.2% accuracy on regular verbs, while the Proportional models had an average of 97.8% accuracy on irregular verbs and 95.7% accuracy on regular verbs). Each was lesioned randomly 5 times at 5% intervals. The results of the models are shown in Figure 6.9. Each point represents a different instance of one of the five models, with different proportions of connections lesioned, and with different random seeds. The black diagonal line represents equal performance on both

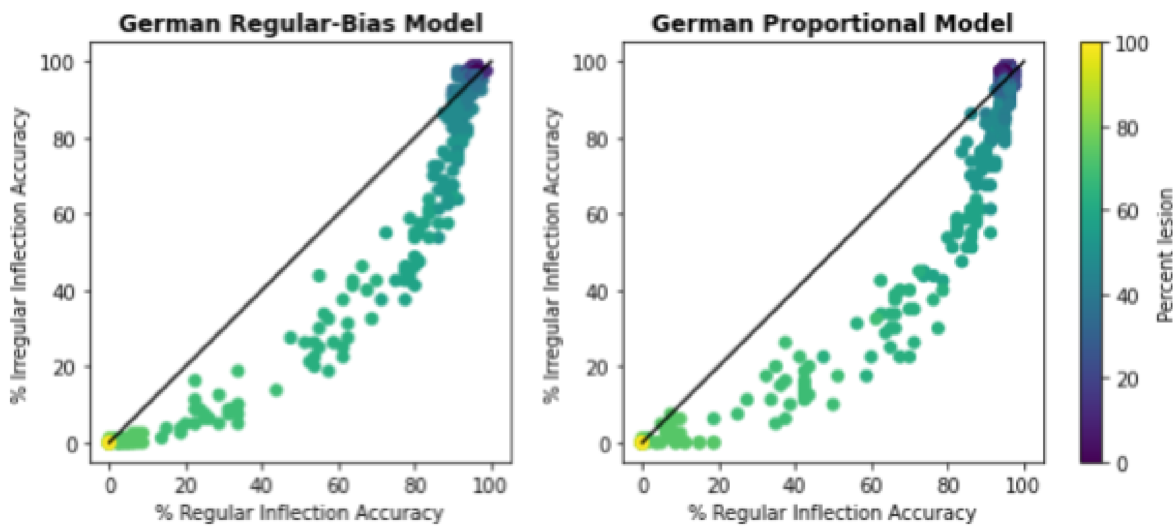

Figure 6.9: Results from Simulation 3. **Left:** Plot summarizing the performance of the model trained on the Regular-Bias dataset, where 28.5% of the dataset is made up of irregular verbs. **Right:** Plot summarizing the performance of the model trained on the Proportional dataset, where 38.5% of the dataset is made up of irregular verbs.

sets of verbs.

In contrast to the English models, but consistent with the German clinical data, both of the models performed better on regular verbs than irregular verbs. The difference between the two models is slight, despite the effect of frequency; however, the primary difference between the Regular-Bias and Proportional models is that the Proportional model exhibited a very small improvement for regular verbs compared to the Regular-Bias model, where regular verbs were preserved longer, at a higher lesioning percentage. The two models performed at approximately the same accuracy levels for irregular verbs. Despite the Proportional dataset involving more irregular verbs than regulars, the model may be more sensitive to the individual frequencies of the verbs in the dataset rather than the proportions of each verb class in the dataset; the frequency effect may be impacting a small set of verbs that are outliers in their frequency distributions, rather than the entire class.

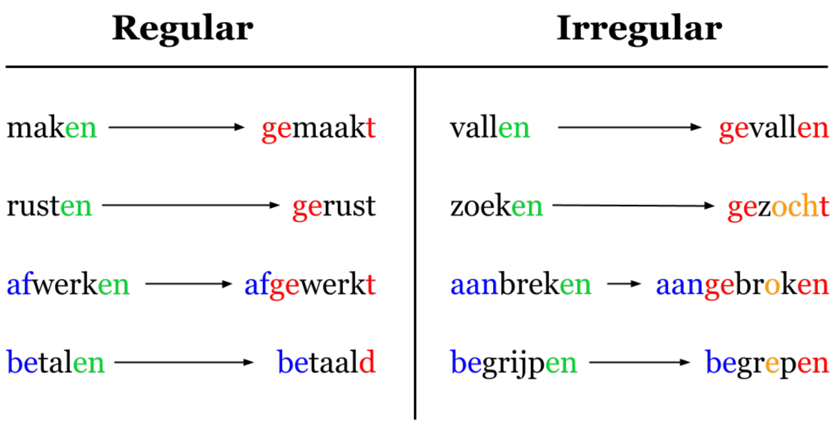

Figure 6.10: Verb classes in Dutch (green = infinitive, blue = prefix, red = past participle, orange = stem changes

6.6 Simulation 4: Dutch

Similar to German, Dutch verbs can also be grouped into regular "weak" verbs and irregular "strong" verbs. The regular past participle is formed from the infinitive form by removing the *-en* suffix, and adding the prefix *ge-* and the suffix -/t/ (orthographically, *-t* or *-d*), where the voicing of the final consonant of the stem assimilates with that of the suffix when a voiceless alternative is available. The suffix is not added to verb stems that end in /t/ or /d/ ("ge-rust"). Again, similarly to German, there are separable and inseparable prefixes, where the separable prefixes appear before the *ge-* prefix, while the inseparable prefixes are able to replace the *ge-* prefix.

Irregular verbs can be irregular in a variety of ways, the most common way being that the suffix of the past participle is *-en* rather than *-t* or *-d* (*gevallen*). Other irregular verbs undergo changes to the vowels and/or consonants in the stem (the stem *zoek* changes to *zoch* in *gezocht*), or undergo suppletion (the past participle of *zijn* is *geweest*).

The proportion of Dutch verbs which are irregular is even higher than German, with about

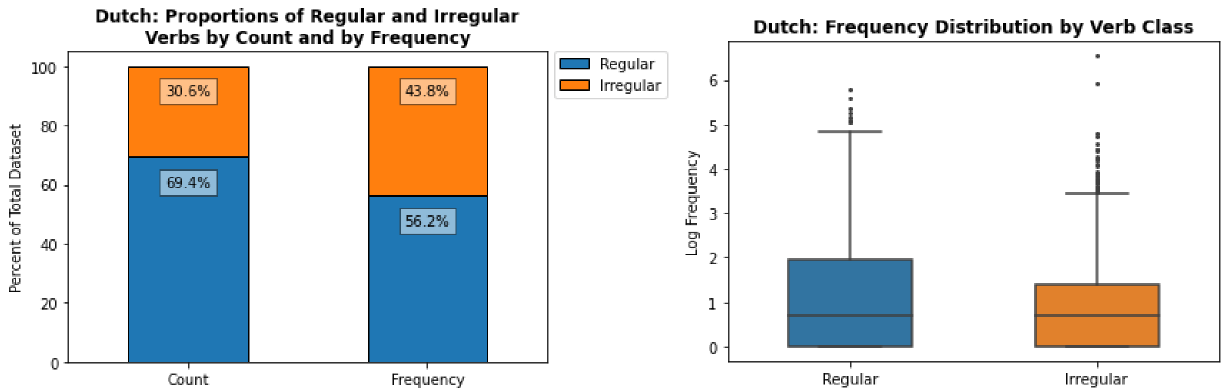

Figure 6.11: The proportions of Dutch verb classes by type frequency and token frequency (left), and the log frequency distribution of regular and irregular verbs (right)

30.6% of the dataset being irregular[3]. Again, irregular verbs are more frequent than regular verbs, so irregular verbs constitute nearly 44% of the summed frequency of the verbs. Based on the proportion of irregular verbs alone, this would lead to the prediction that the Dutch models should have results similar to the Irregular-Bias models in English, with irregular verbs preserving longer in the Dutch models than in the German models. However, because the average frequency of the irregular verbs in Dutch is proportionally lower than both English and German, this may lead to the prediction that the Proportional model in Dutch should perform more like the Regular-Bias model in English, with irregular verbs not preserving as long as in the German or English models, assuming that token frequency is more influential in form preservation compared to type frequency.

Five models were trained on each dataset, reaching an average of 95.9% accuracy for all verbs prior to lesioning (the Regular-Bias models had an average of 97.5% accuracy on irregular verbs and 94.5% accuracy on regular verbs, while the Proportional models had an average of 97.5% accuracy on irregular verbs and 94% accuracy on regular verbs). Each was lesioned

[3]Like in the German dataset, verbs were excluded from the Dutch training dataset if their corpus frequency was 0 per million. The dataset including these verbs would have 10,878 items, 2076 of which are irregular, constituting 19.1% of the total dataset.

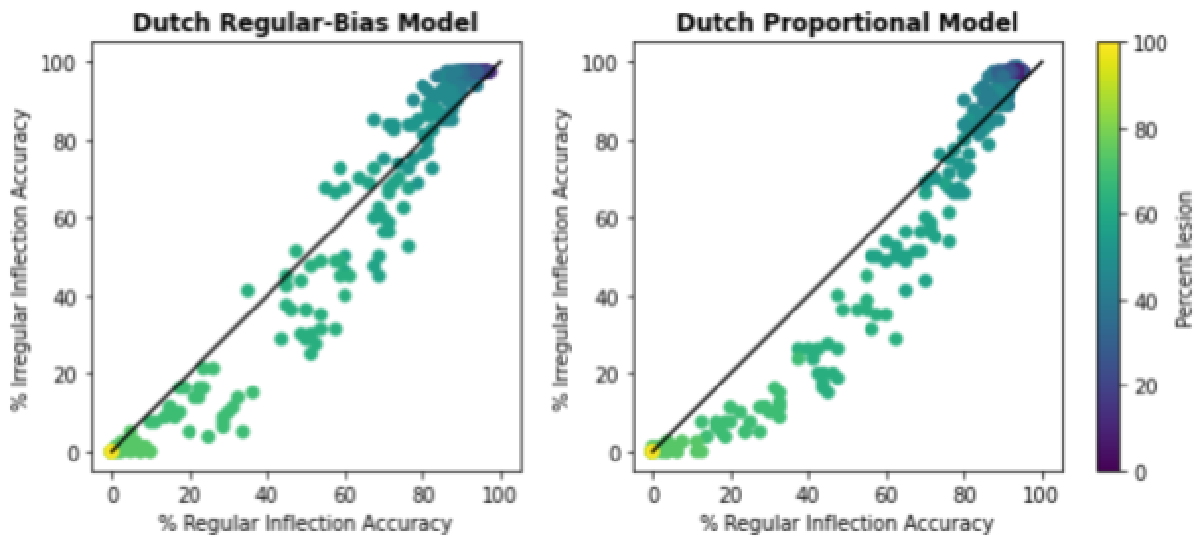

Figure 6.12: Results from Simulation 4. **Left:** Plot summarizing the performance of the model trained on the Regular-Bias dataset, where 30.6% of the dataset is made up of irregular verbs. **Right:** Plot summarizing the performance of the model trained on the Proportional dataset, where 43.8% of the dataset is made up of irregular verbs.

randomly 5 times at 5% intervals. The results of the models are shown in the figure below. Each point represents a different instance of one of the five models, with different proportions of connections lesioned, and with different random seeds. The black diagonal line represents equal performance on both sets of verbs.

Similarly to the German data, the Proportional models pull the accuracy further below the equivalence line compared to the Regular-Bias models; however, that effect in the Dutch models arises mainly because the Proportional models do not preserve the irregular verbs as long, rather than being caused by better performance or preservation on the regular verbs. Dutch irregular verbs have a lower average frequency than regular verbs, but there are several high-frequency outliers which cause the irregular verbs as a whole to constitute a greater proportion of the frequency-biased dataset. This evidence again suggests that the factor which conditions form preservation is individual form frequency, rather than the proportions of verb classes in the dataset.

Although some of the clinical data points are within the range of the models' performance, these models are not a perfect simulation of the Dutch clinical data. This could be due to the model not fully learning the regular past tense in Dutch (baseline accuracy on regular verbs was about 94.3%, whereas several of the Dutch patients have perfect or near-perfect accuracy on regular verbs). If the Proportional model were to retain the curved shape of its distribution but increase its baseline accuracy on regular verbs, it would more accurately align with the clinical data points.

6.7 Discussion

These results demonstrate that the individual frequency of verbs likely plays a key role in modulating how well those forms are preserved in aphasia. This result provides an alternate account for the apparent dissociation between regular and irregular inflections observed by Ullman et al. (1997), which was previously a sticking point for the single-route model in the past tense debate. Simulations 1 and 2 demonstrate that when the model is trained on a dataset with representation of regular and irregular verbs proportional to their frequency, the verb class with higher average frequencies preserves better after lesioning. Simulations 3 and 4 similarly demonstrate that when the model is trained on languages with varying proportions of regulars and irregulars, and with varying transformation patterns, the verbs are preserved according to their individual frequencies, rather than that of the entire verb class[4]. The models' performances were consistent with the clinical data for the speakers of those languages. Thus, rather than assuming that the pat-

[4]Even though the test batteries used in the clinical data often matched the sets of regular and irregular verbs on frequency, the patients may be sensitive to patterns of sub-regularity and their relative frequency, causing them to perform differently on the two sets of verbs.

tern of deficits in agrammatic aphasia arises because of syntactic or lexical features of the verbs, it can be argued instead that the process involved in the IFG is a phonological transformation, conditioned by the individual frequencies of the verbs in that language.

Faroqi-Shah (2007) placed Ullman et al. (1997)'s findings into a broader context of agrammatic aphasia across different languages, showing that FCL's performance may be an outlier when compared to other individuals with agrammatic aphasia, consistent with the argument in Plaut (1995), that dissociations do not require modularity of those mechanisms. The broader set of cross-linguistic clinical data described by Faroqi-Shah (2007) does not clearly support the dual-route model as proposed by Pinker and Ullman (2002). The models trained on frequency-biased training data are able to mimic the patterns observed for speakers of those languages with agrammatic aphasia, suggesting that the patterns of deficits in agrammatic aphasia - which seemed to be contradictory in a dual-route account - can be better understood as a result of differences in the individual frequencies of regular and irregular verbs in those languages, and not because of differences in the underlying representations of those verbs.

One route for future work would be to test other languages included in the Faroqi-Shah (2007) meta-analysis, such as Spanish, Italian, Greek, and Catalan. Those languages were not included in the present chapter due to the small sample sizes available in the published clinical data (Spanish, n=3; Italian, n=1; Greek, n=1; Catalan, n=1). Generalizing based on single case studies can be problematic, given that they may be outliers (as in the case of FCL), and basing the models' success on how well they can capture those single data points may not lead to accurate conclusions. More clinical data from a wider variety of languages is necessary to perform these kinds of comparisons. Although English, German, and Dutch are all Germanic languages, the inflection deficits in each language group do exhibit relatively different distributions, so the models' accuracy on all three languages should not be attributed to the similarities between them.

Faroqi-Shah's meta-analysis also demonstrated that there is a significant amount of variation between individuals with aphasia, even when the injury impacts very similar brain areas. This could emerge due to individual differences in how each form is encoded neurally, or due to differences in which groups of neurons are impacted and whether they are able to be reorganized through neural and synaptic regeneration processes or by leveraging alternate pathways through the brain (Campbell et al., 2019). The models used in this chapter also observed that variation between models' starting states had an impact on their performance after lesioning. Simulations 3 and 4 trained and lesioned 5 models each, which exhibited a wider range in performance compared the single models in simulations 1 and 2. This suggests that randomness in the model's baseline state does impact its outcomes after lesioning. Other forms of variation could be introduced to these models by using a different lesioning method, such as adding Gaussian noise to all weights between layers, unit ablation—severing all outgoing connections from some units—or adding Gaussian noise to the activations of units. Each strategy can have a different effect on the performance of the model, as discussed in Guest et al. (2020).

There are many differences between our model and the actual IFG, both in terms of the neural architectures and learning methods. This means that it cannot be an exact model of what might be happening at the neural level after a brain injury such as a stroke. Designing a model to more closely reflect the architecture of the IFG might yield different results or further insights in future work. However, even neural networks that are designed to have similar functions or mechanisms as the human brain can sometimes behave very differently (Rajalingham et al., 2018). This approach is simpler, but is nevertheless sufficiently similar to the hypothesized function of the IFG to allow me to test predictions of what should happen when the mechanism is damaged in a controlled way.

Interestingly, some findings have suggested that lesions to the frontal lobe alone are not sufficient to cause agrammatic aphasia, where patients with damage to the IFG do not always have agrammatic speech; of the participants that have damage to the IFG, only the ones with co-occurring damage to the superior temporal gyrus (STG) had agrammatic speech (Fridriksson et al, 2015). In tandem with the current results, this point may provide evidence in support of the WOW model's account, where the STG performs phonological operations post-syntactically, and where the connectivity between temporal lobe and frontal lobe structures is a key component of fluent speech production.

Future work should simulate the effects of fluent aphasia, or the accuracy data for the patient JLU from Ullman et al. (1997). Fluent aphasia is not as well-understood as agrammatic/nonfluent aphasia, and there are many other competing theories that may not be as easy to test with this kind of neural network. For example, because fluent aphasia often involves "empty speech" (syntactically well-formed sentences that lack meaning or message), it has sometimes been concluded that fluent aphasia involves difficulties with discourse coherence (Linnik, 2016), or a failure to properly inhibit incorrect lexical items that are retrieved (Prather et al., 1997). These may be difficult to implement as a neural model, though additional investigation in this area could be fruitful. In addition, future work could involve additional connectionist mechanisms in the WOW framework, with different kinds of representations as inputs and outputs, though the exact format of those representations would have to be determined.

Once again, these findings should not be interpreted as a challenge to the view that words and sentences have hierarchical structure that is relevant during on-line language comprehension and production. Rather, I interpret this model as performing a calculation over phonological space. It represents a post-syntactic operation such as “Vocabulary Insertion”, as characterized by Distributed Morphology, or “mapping from sets of syntactic units to sets of phonological units”, as described by the WOW model. This model is an opportunity to integrate connectionist theories of neural representation with current symbolic theories of language and language processing. From the perspective of the WOW model, which is mainly interested in transformations between independent data structures, the connectionist approach is an entirely reasonable way to represent these kinds of processes.

Chapter 7

Conclusion

This dissertation has explored the consequences of lexicalism in psycho- and neuro-linguistics, and detailed a non-lexicalist, non-semiotic framework for linguistic exploration. Chapter 2 outlined the arguments that have been used against lexicalist approaches in linguistic theory, using cross-linguistic data to show that structure above and below the word level is all a part of the same system, that the language system cannot be organized around "triads" of meaning, syntax, and form, and things both larger and smaller than a "word" must be listed in our knowledge of language. Chapter 3 describes the ways that lexicalist assumptions have been implemented in psycho- and neuro-linguistic models, using language production as a case study. Because of the lexicalist influence on these models, where "lemmas" or "lexical items" are stored as triads of meaning, syntax, and form, and where syntax and morphology are both representationally and functionally divided, they struggle to account for a wide array of linguistic phenomena. The WOW model described in Chapter 4 is a positive proposal for what language production can look like without relying on lemmas or lexical items to drive language production, emphasizing independent mapping processes between meaning and syntax, and syntax and form, as well as the role of prosody in linearizing the elements of the utterance. The model's predictions regarding mapping complexity are supported in Chapter 5, which used a novel experimental paradigm to measure neural responses in EEG during language production. The preliminary results observed in this experiment showed greater neural responses in two different time windows for cases where either the meaning-syntax or the syntax-form mappings are less transparent and require more contextual information to complete the translation. Finally, the WOW model's predictions for aphasia and observed dissociations between regular and irregular verb forms is supported by the results of the recurrent neural networks in Chapter 6. Because the single-mechanism model was able to simulate the patterns in the observed clinical data after lesioning, this supports a view where regular and irregular verbs are only different in terms of their phonological patterns and individual frequency distributions, and not because they have different syntactic properties underlyingly.

This dissertation is just a starting point for creating a world in which "words" are not the focus of psycho- or neuro-linguistic inquiry. There are many interesting questions raised by the WOW model and the non-lexicalist framework that have yet to be explored, such as testing the model's predictions about the localization of language functions in the brain, the role of pre-syntactic and post-syntactic prosody in generating fluent speech, and providing a more detailed implementational account of the translations between complex data structures as it occurs over neural architectures. Future work could also consider language comprehension in a non-lexicalist framework, provide details on the limits and constraints on the mappings between different representations, and develop a connectionist model of language production, working toward a neural implementation of the computations involved in language.

Another aspect of the move toward a non-semiotic theory of language is that the learning problem becomes much more complex; the learning process would not be just about learning a pairing between a "word form" and a "word meaning" and then figuring out how they can be combined, but about building complex representations of meanings and forms and linking them to a fully abstract, hierarchical structure that they have no direct evidence for. There are a variety of predictions about child language acquisition in a non-lexicalist framework that could be tested. For example, do children begin with the assumption that form-meaning pairings are the standard? If so, at what point do they begin to posit a non-semiotic system, and what information leads them to make that shift? Alternatively, perhaps the advantage that human brains have over non-human brains in learning natural language is that the neural architecture is such that a non-semiotic theory of language is the starting point; for example, perhaps the only neural pathway that can link conceptual representations and phonological forms passes through multiple sections of cortex that foster more complex, abstract calculations that intermediate between meaning and form? In all, applying a non-lexicalist framework to the learning problem generates much more interesting questions about human cognition and learning than word-based models.

One question that this work also leads to is about the role of lexicalism in the development of linguistic theory; what would linguistic theory look like if lexicalism had not been the starting assumption at the outset, if we had always been in a "world without words"? If we had started with the assumption that meaning, syntax, and form are fully independent and do not always align, rather than assuming that our knowledge of language is organized around direct links between meanings and forms, what kinds of linguistic theories would exist? Or perhaps lexicalism was a necessary initial hypothesis, a simplifying assumption, and only by testing that hypothesis could we have discovered that there is a better alternative?

Bibliography

Acheson, D. J., Hamidi, M., Binder, J. R., and Postle, B. R. (2011). A common neural substrate for language production and verbal working memory. *Journal of Cognitive Neuroscience*, 23(6):1358–1367.

Adams, M. E. (2014). *The comparative grammaticality of the English comparative*. Stanford University.

Aikhenvald, A. Y. et al. (2003). Evidentiality in typological perspective. *Typological Studies in Language*, 54:1–32.

Anttila, A. (2016). Phonological effects on syntactic variation. *Annual Review of Linguistics*, 2(1):115–137.

Anttila, A., Adams, M., and Speriosu, M. (2010). The role of prosody in the English dative alternation. *Language and Cognitive Processes*, 25(7-9):946–981.

Aronoff, M. (1976). *Word formation in generative grammar*. Linguistic Inquiry Monographs. MIT Press.

Baayen, R. H., Piepenbrock, R., and Gulikers, L. (1996). The CELEX Lexical Database.

Baggio, G. and Hagoort, P. (2011). The balance between memory and unification in semantics: A dynamic account of the N400. *Language and Cognitive Processes*, 26(9):1338–1367.

Bahdanau, D., Cho, K., and Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Baker, M. (1985). The mirror principle and morphosyntactic explanation. *Linguistic Inquiry*, 16(3):373–415.

Belyk, M., Brown, S., Lim, J., and Kotz, S. A. (2017). Convergence of semantics and emotional expression within the IFG pars orbitalis. *Neuroimage*, 156:240–248.

Berge, A. (2016). Eskimo-aleut. In *Oxford Research Encyclopedia of Linguistics*. Oxford University Press.

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

Bhattasali, S., Fabre, M., Luh, W.-M., Al Saied, H., Constant, M., Pallier, C., Brennan, J. R., Spreng, R. N., and Hale, J. (2019). Localising memory retrieval and syntactic composition: an fMRI study of naturalistic language comprehension. *Language, Cognition and Neuroscience*, 34(4):491–510.

Bhattasali, S., Hale, J., Pallier, C., Brennan, J., Luh, W.-M., and Spreng, R. N. (2018). Differentiating phrase structure parsing and memory retrieval in the brain. *Society for Computation in Linguistics*, 1(1).

Bhattasali, S., Popa-Fabre, M., Pallier, C., and Hale, J. (2020). Modeling conventionalization and predictability within MWEs at the brain level. *Society for Computation in Linguistics*, 3(1).

Binder, J. R., Desai, R. H., Graves, W. W., and Conant, L. L. (2009). Where is the semantic system? a critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral cortex*, 19(12):2767–2796.

Bird, H., Ralph, M. A. L., Seidenberg, M. S., McClelland, J. L., and Patterson, K. (2003). Deficits in phonology and past-tense morphology: What's the connection? *Journal of Memory and Language*, 48(3):502–526.

Blonder, L. X., Bowers, D., and Heilman, K. M. (1991). The role of the right hemisphere in emotional communication. *Brain*, 114(3):1115–1127.

Bock, K. (1995). Sentence production: From mind to mouth. In Miller, J. L. and Eimas, P. D., editors, *Speech, Language, and Communication (Second Edition)*, Handbook of Perception and Cognition, pages 181–216. Academic Press, San Diego, 2nd edition.

Bock, K. and Ferreira, V. S. (2014). Syntactically speaking. *The Oxford handbook of language production*, pages 21–46.

Bock, K., Levelt, W., and Gernsbacher, M. A. (2002). Language production: Grammatical encoding. *Psycholinguistics: Critical Concepts in Psychology*, 5:405–452.

Boeckx, C., Martinez-Alvarez, A., and Leivada, E. (2014). The functional neuroanatomy of serial order in language. *Journal of Neurolinguistics*, 32:1–15.

Boemio, A., Fromm, S., Braun, A., and Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature neuroscience*, 8(3):389–395.

Bresnan, J. (1978). chapter a realistic transformational grammar. *Linguistic theory and psychological reality*, pages 1–59.

Briggs, J. L., Johns, A., and Cook, C. (2015). Dictionary of Utkuhiksalingmiut Inuktitut postbase suffixes. *Iqaluit: Nunavut Arctic College*.

Brown-Schmidt, S. and Konopka, A. E. (2008). Little houses and casas pequeñas: Message formulation and syntactic form in unscripted speech with speakers of english and spanish. *Cognition*, 109(2):274–280.

Brown-Schmidt, S. and Konopka, A. E. (2015). Processes of incremental message planning during conversation. *Psychonomic Bulletin & Review*, 22(3):833–843.

Brown-Schmidt, S. and Tanenhaus, M. K. (2006). Watching the eyes when talking about size: An investigation of message formulation and utterance planning. *Journal of Memory and Language*, 54(4):592–609.

Bruening, B. (2018). The lexicalist hypothesis: Both wrong and superfluous. *Language*, 94(1):1–42.

Buchsbaum, B. R., Hickok, G., and Humphries, C. (2001). Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cognitive science*, 25(5):663–678.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, 10(5):425–455.

Călinescu, L., Ramchand, G., and Baggio, G. (2023). How (not) to look for meaning composition in the brain: A reassessment of current experimental paradigms. *Frontiers in Language Sciences*, 2:1096110.

Campbell, B. C., De Silva, D. A., Macleod, M. R., Coutts, S. B., Schwamm, L. H., Davis, S. M., and Donnan, G. A. (2019). Ischaemic stroke. *Nature Reviews Disease Primers*, 5(1):1–22.

Cappelle, B., Shtyrov, Y., and Pulvermüller, F. (2010). Heating up or cooling up the brain? MEG evidence that phrasal verbs are lexical units. *Brain and Language*, 115(3):189–201.

Caramazza, A., Berndt, R. S., Basili, A. G., and Koller, J. J. (1981). Syntactic Processing Deficits in Aphasia. *Cortex*, 17(3):333–347.

Caramazza, A. and Zurif, E. B. (1976). Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *Brain and language*, 3(4):572–582.

Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., and Knight, R. T. (2010). Categorical speech representation in human superior temporal gyrus. *Nature neuroscience*, 13(11):1428–1432.

Cho-Reyes, S. and Thompson, C. K. (2012). Verb and sentence production and comprehension in aphasia: Northwestern assessment of verbs and sentences (NAVS). *Aphasiology*, 26(10):1250–1277.

Chomsky, N. (1970). *Readings in English Transformational Grammar*, chapter Remarks on nominalization, pages 184–221. Ginn and Company.

Chomsky, N. (1982). *Some concepts and consequences of the theory of government and binding*, volume 6. MIT press.

Chomsky, N. (1986). *Barriers*. Number 13 in Linguistic inquiry monographs. MIT Press.

Chomsky, N. (1991). Some notes on economy of derivation and representation. In Freidin, R., editor, *Principles and parameters in comparative grammar*, number 20 in Current studies in linguistics series. MIT Press, Cambridge, Mass. Meeting Name: Princeton Workshop on Comparative Grammar.

Chomsky, N. (1993). *Lectures on government and binding: The Pisa lectures*. Number 9 in Studies in generative grammar. Walter de Gruyter.

Chomsky, N. (2007). Approaching UG from below. In *Interfaces + recursion = language?: Chomsky's Minimalism and the View from Syntax-Semantics*, pages 1–30. De Gruyter Mouton.

Chomsky, N. and Lasnik, H. (1993). The theory of principles and parameters. *Syntax*, 1:506–569.

Clahsen, H., Paulmann, S., Budd, M.-J., and Barry, C. (2018). Morphological encoding beyond slots and fillers: An erp study of comparative formation in english. *Plos one*, 13(7):e0199897.

Clahsen, H., Rothweiler, M., Woest, A., and Marcus, G. F. (1992). Regular and irregular inflection in the acquisition of German noun plurals. *Cognition*, 45:225–255.

Clark, A. (2013). Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences*, 36(3):181–204.

Cook, C. and Johns, A. (2009). Determining the semantics of Inuktitut postbases. *Variations on Polysynthesis*, pages 149–170.

Cutting, J. C. and Bock, K. (1997). That's the way the cookie bounces: Syntactic and semantic components of experimentally elicited idiom blends. *Memory & Cognition*, 25(1):57–71.

Danly, M., Cooper, W. E., and Shapiro, B. (1983). Fundamental frequency, language processing, and linguistic structure in Wernicke's aphasia. *Brain and Language*, 19(1):1–24.

Danly, M. and Shapiro, B. (1982). Speech prosody in Broca's aphasia. *Brain and language*, 16(2):171–190.

Darby, D. G. (1993). Sensory aprosodia: A clinical clue to lesions of the inferior division of the right middle cerebral artery? *Neurology*, 43(3 Part 1):567–567.

De Looze, C., Yanushevskaya, I., Murphy, A., O'Connor, E., and Gobl, C. (2015). Pitch declination and reset as a function of utterance duration in conversational speech data. In *Interspeech 2015*, pages 3071–3075. ISCA.

Dell, G. S., Oppenheim, G. M., and Kittredge, A. K. (2008). Saying the right word at the right time: Syntagmatic and paradigmatic interference in sentence production. *Language and Cognitive Processes*, 23(4):583–608.

Di Sciullo, A.-M. and Williams, E. (1987). *On the definition of word*, volume 14. Springer.

Eberhard, K. M., Cutting, J. C., and Bock, K. (2005). Making syntax of sense: number agreement in sentence production. *Psychological review*, 112(3):531.

Elfner, E. (2011). The interaction of linearization and prosody: Evidence from pronoun postposing in irish. In Carnie, A., editor, *Formal Approaches to Celtic Linguistics*, pages 17–40. Cambridge Scholars Publishing, Newcastle upon Tyne, UK.

Elman, J. L. (2011). Lexical knowledge without a lexicon? *The mental lexicon*, 6(1):1–33.

Embick, D. (2015). *The Morpheme*. De Gruyter Mouton.

Embick, D. and Noyer, R. (1999). Locality in post-syntactic operations. *MIT working papers in linguistics*.

Embick, D. and Noyer, R. (2001). Movement operations after syntax. *Linguistic Inquiry*, 32(4):555–595.

Emonds, J. (1978). The verbal complex v-v in french. *Linguistic inquiry*, pages 151–175.

Fabb, N. A. J. (1984). *Syntactic affixation*. PhD thesis, Massachusetts Institute of Technology.

Faroqi-Shah, Y. (2007). Are regular and irregular verbs dissociated in non-fluent aphasia?: A meta-analysis. *Brain Research Bulletin*, 74(1-3):1–13.

Fedorenko, E., Blank, I. A., Siegelman, M., and Mineroff, Z. (2020). Lack of selectivity for syntax relative to word meanings throughout the language network. *Cognition*, 203:104348.

Fernandez, J. and Downey, D. (2018). Sampling informative training data for rnn language models. In *Proceedings of ACL 2018, Student Research Workshop*, pages 9–13.

Ferreira, F. (2013). Syntax in language production: An approach using tree-adjoining grammars. In Wheeldon, L., editor, *Aspects of language production*, pages 303–342. Psychology Press.

Ferreira, V. S. (1996). Is it better to give than to donate? syntactic flexibility in language production. *Journal of memory and language*, 35(5):724–755.

Ferreira, V. S. and Slevc, L. R. (2007). Grammatical encoding. In Gaskell, M. G. and Altmann, G. T. M., editors, *The Oxford Handbook of Psycholinguistics*, pages 453–469. Oxford University Press, Oxford, UK.

Fridriksson, J., Fillmore, P., Guo, D., and Rorden, C. (2015). Chronic broca's aphasia is caused by damage to Broca's and Wernicke's areas. *Cerebral Cortex*, 25(12):4689–4696.

Friedmann, N., Biran, M., and Dotan, D. (2013). Lexical retrieval and its breakdown in aphasia and developmental language impairment. *The Cambridge handbook of biolinguistics*, pages 350–374.

Ganis, G. and Kutas, M. (2003). An electrophysiological study of scene effects on object identification. *Cognitive Brain Research*, 16(2):123–144.

Gaston, P., Brodbeck, C., Phillips, C., and Lau, E. (2023). Auditory word comprehension is less incremental in isolated words. *Neurobiology of Language*, 4(1):29–52.

Geranmayeh, F., Wise, R. J., Mehta, A., and Leech, R. (2014). Overlapping networks engaged during spoken language production and its cognitive control. *Journal of Neuroscience*, 34(26):8728–8740.

Goebel, R. and Indefrey, P. (2000). *A Recurrent Network with Short-term Memory Capacity Learning the German -s plural*, pages 177–200. Oxford University Press, Oxford.

Goldberg, A. and Suttle, L. (2010). Construction grammar. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(4):468–477.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. University of Chicago Press.

Goodglass, H. and Baker, E. (1976). Semantic field, naming, and auditory comprehension in aphasia. *Brain and language*, 3(3):359–374.

Gordon, M. (2011). Stress: Phonotactic and phonetic evidence. *The Blackwell companion to phonology*, 2:924–948.

Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition*, 82(1):B1–B14.

Guenther, F. H. and Hickok, G. (2015). Role of the auditory system in speech production. *Handbook of Clinical Neurology*, 129:161–175.

Guest, O., Caso, A., and Cooper, R. P. (2020). On simulating neural damage in connectionist networks. *Computational brain & behavior*, 3(3):289–321.

Hagoort, P. (1993). Impairments of lexical-semantic processing in aphasia: Evidence from the processing of lexical ambiguities. *Brain and Language*, 45(2):189–232.

Hagoort, P. (2013). MUC (memory, unification, control) and beyond. *Frontiers in Psychology*, 4:45726.

Hagoort, P. (2016). MUC (Memory, Unification, Control): A model on the neurobiology of language beyond single word processing. In *Neurobiology of Language*, pages 339–347. Elsevier.

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

Halle, M. and Marantz, A. (1993). Distributed morphology and the pieces of inflection. In Hale, K. L., Keyser, S. J., and Bromberger, S., editors, *The View from Building 20*, pages 111–176. The MIT Press.

Hare, M., Elman, J. L., and Daugherty, K. G. (1995). Default generalization in connectionist networks. *Language and Cognitive Processes*, 10:601–630.

Harizanov, B. and Gribanova, V. (2019). Whither head movement? *Natural Language & Linguistic Theory*, 37(2):461–522.

Harley, H. (2008). On the causative construction. In Miyagawa, S. and Saito, M., editors, *Oxford Handbook of Japanese linguistics*, pages 20–53. Oxford University Press, Oxford, UK.

Harley, H. (2014). On the identity of roots. *Theoretical Linguistics*, 40(3-4):225–276.

Haspelmath, M. (2017). The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia Linguistica*, 51(s1000):31–80.

Hayes, B. and Lahiri, A. (1991). Bengali intonational phonology. *Natural language & linguistic theory*, 9:47–96.

Heim, I. and Kratzer, A. (1998). *Semantics in generative grammar*. Blackwell Oxford.

Hein, G. and Knight, R. T. (2008). Superior temporal sulcus—it's my area: or is it? *Journal of cognitive neuroscience*, 20(12):2125–2136.

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, 13(2):135–145.

Hickok, G. and Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in cognitive sciences*, 4(4):131–138.

Hickok, G. and Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5):393–402.

Hickok, G., Venezia, J., and Teghipco, A. (2023). Beyond broca: neural architecture and evolution of a dual motor speech coordination system. *Brain*, 146(5):1775–1790.

Hoosain, R. (1992). Psychological reality of the word in chinese. In *Advances in psychology*, volume 90, pages 111–130. Elsevier.

Hwang, H. and Kaiser, E. (2014). Having a syntactic choice is not always better: the effects of syntactic flexibility on korean production. *Language, Cognition and Neuroscience*, 29(9):1115–1131.

Idsardi, W. (2022). Underspecification in time. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, 67(4):670–682.

Idsardi, W. and Raimy, E. (2013). Three types of linearization and the temporal aspects of speech. *Challenges to linearization*, 1:31–56.

Iwasaki, N. (2011). Incremental sentence production: Observations from elicited speech errors in Japanese. *Processing and producing head-final structures*, pages 131–151.

Jackendoff, R. (1975). Morphological and semantic regularities in the lexicon. *Language*, pages 639–671.

Jackendoff, R. (1997). *The architecture of the language faculty*. Number 28 in Linguistic Inquiry Monographs. MIT Press.

Jackendoff, R. (2002). English particle constructions, the lexicon, and the autonomy of syntax. In Nicole, D., editor, *Verb-particle explorations*, pages 67–94. Mouton de Gruyter, Berlin, Germany.

Jackendoff, R. (2007). A parallel architecture perspective on language processing. *Brain research*, 1146:2–22.

Jackendoff, R. (2017). In defense of theory. *Cognitive Science*, 41:185–212.

Joshi, A. K. and Schabes, Y. (1997). Tree-adjoining grammars. In *Handbook of Formal Languages*, pages 69–123. Springer.

Kann, K. and Schütze, H. (2016). Med: The lmu system for the sigmorphon 2016 shared task on morphological reinflection. In *Proceedings of the 14th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 62–70.

Kaplan, L. (2003). Inuit snow terms: how many and what does it mean? In *Building Capacity in Arctic Societies: Dynamics and Shifting Perspectives. Proceedings from the 2nd IPSSAS Seminar*, volume 2, Iqaluit, Nunavut, Canada.

Kempen, G. and Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive science*, 11(2):201–258.

Kempen, G. and Huijbers, P. (1983). The lexicalization process in sentence production and naming: Indirect election of words. *Cognition*, 14(2):185–209.

King, T. H. (1996). Licensing negative pronominals in Georgian. In *Annual Meeting of the Berkeley Linguistics Society*, volume 22, pages 232–243.

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

Klein, G., Hernandez, F., Nguyen, V., and Senellart, J. (2020). The opennmt neural machine translation toolkit: 2020 edition. In *Proceedings of the 14th Conference of the Association for Machine Translation in the Americas (AMTA 2020)*, pages 102–109.

Konopka, A. E. and Bock, K. (2009). Lexical or syntactic control of sentence formulation? structural generalizations from idiom production. *Cognitive Psychology*, 58(1):68–101.

Kruyt, J. and Dutilh, M. (2012). A 38 million words dutch text corpus and its users. *Lexikos*, 7(1).

Kuiper, K., Van Egmond, M.-E., Kempen, G., and Sprenger, S. (2007). Slipping on superlemmas: Multi-word lexical items in speech production. *The Mental Lexicon*, 2(3):313–357.

Laaha, S., Ravid, D., Korecky-Kröll, K., Laaha, G., and Dressler, W. U. (2006). Early noun plurals in German: regularity, productivity, or default? *Journal of Child Language*, 33(2):271–302.

Lachter, J. and Bever, T. G. (1988). The relation between linguistic structure and associative theories of language learning - a constructive critique of some connectionist learning models. *Cognition*, 28:195–247.

Lapointe, S. (1980). *A theory of grammatical agreement*. PhD thesis, University of Massachusetts, Amherst.

Lau, E. F., Phillips, C., and Poeppel, D. (2008). A cortical network for semantics:(de) constructing the N400. *Nature Reviews Neuroscience*, 9(12):920–933.

Levelt, W. J. (1989). *Speaking: From intention to articulation*. ACL-MIT Press series in natural-language processing. MIT Press, Cambridge, MA.

Levelt, W. J. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, 42(1-3):1–22.

Levelt, W. J. and Indefrey, P. (2000). The speaking mind/brain: Where do spoken words come from? In *Image, Language, Brain: Papers from the First Mind Articulation Project Symposium*, pages 77–94. MIT Press.

Levelt, W. J., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(1):1–38.

Liebenthal, E., Desai, R. H., Humphries, C., Sabri, M., and Desai, A. (2014). The functional organization of the left STS: a large scale meta-analysis of PET and fMRI studies of healthy adults. *Frontiers in Neuroscience*, 8:289.

Lieberman, E., Michel, J.-B., Jackson, J., Tang, T., and Nowak, M. A. (2007). Quantifying the evolutionary dynamics of language. *Nature*, 449(7163):713–716.

Linnik, A. (2016). *Coherence and structure in aphasic and non-aphasic spoken discourse*. PhD thesis, Universität Potsdam.

Lowe, M. X., Rajsic, J., Ferber, S., and Walther, D. B. (2018). Discriminating scene categories from brain activity within 100 milliseconds. *Cortex*, 106:275–287.

Lüngen, H. (2017). Dereko–das deutsche Referenzkorpus: Schriftkorpora der deutschen Gegenwartssprache am Institut für deutsche Sprache in Mannheim. *Zeitschrift für germanistische Linguistik*, 45(1):161–170.

MacWhinney, B. and Leinbach, J. (1991). Implementations are not conceptualizations: Revising the verb learning model. *Cognition*, 49:235–290.

Marcus, G. F., Brinkmann, U., Clahsen, H., Wiese, R., and Pinker, S. (1995). German inflection: The exception that proves the rule. *Cognitive Psychology*, 29(3):189–256.

Margulies, D. S. and Petrides, M. (2013). Distinct parietal and temporal connectivity profiles of ventrolateral frontal areas involved in language production. *Journal of Neuroscience*, 33(42):16846–16852.

Marr, D. (1982/2010). *Vision: A computational investigation into the human representation and processing of visual information*. W.H. Freeman and Company.

Martin, L. (1986). “Eskimo words for snow“: A case study in the genesis and decay of an anthropological example. *American Anthropologist*, 88(2):418–423.

Martin, R. C. and Slevc, L. R. (2014). Language production and working memory. *The Oxford handbook of language production*, pages 437–450.

Martínez-Ferreiro, S., Bastiaanse, R., and Boye, K. (2020). Functional and usage-based approaches to aphasia: the grammatical-lexical distinction and the role of frequency. *Aphasiology*, 34(8):927–942.

Matar, S., Dirani, J., Marantz, A., and Pylkkänen, L. (2021). Left posterior temporal cortex is sensitive to syntax within conceptually matched arabic expressions. *Scientific reports*, 11(1):7181.

Matchin, W. (2023). Lexico-semantics obscures lexical syntax: commentary on Fedorenko et al.(2020). *Frontiers in Language Sciences*, 2:1217837.

Matchin, W., Basilakos, A., Stark, B. C., den Ouden, D.-B., Fridriksson, J., and Hickok, G. (2020). Agrammatism and paragrammatism: a cortical double dissociation revealed by lesion-symptom mapping. *Neurobiology of Language*, 1(2):208–225.

Matchin, W., den Ouden, D.-B., Basilakos, A., Stark, B. C., Fridriksson, J., and Hickok, G. (2023). Grammatical parallelism in aphasia: a lesion-symptom mapping study. *Neurobiology of Language*, 4(4):550–574.

Matchin, W. and Hickok, G. (2020). The cortical organization of syntax. *Cerebral Cortex*, 30(3):1481–1498.

Mayer, J., Wildgruber, D., Riecker, A., Dogil, G., Ackermann, H., and Grodd, W. (2002). Prosody production and perception: converging evidence from fMRI studies. In *Speech Prosody*.

McClelland, J. L. and Patterson, K. (2002). Rules or connections in past-tense inflections: What does the evidence rule out? *Trends in Cognitive Sciences*, 6(11):465–472.

McCurdy, K., Goldwater, S., and Lopez, A. (2020). Inflecting when there’s no majority: Limitations of encoder-decoder neural networks as cognitive models for german plurals. *arXiv preprint arXiv:2005.08826*.

Meekings, S. and Scott, S. K. (2021). Error in the superior temporal gyrus? a systematic review and activation likelihood estimation meta-analysis of speech production studies. *Journal of Cognitive Neuroscience*, 33(3):422–444.

Mudrik, L., Lamy, D., and Deouell, L. Y. (2010). Erp evidence for context congruity effects during simultaneous object–scene processing. *Neuropsychologia*, 48(2):507–517.

Nedergaard, J. S., Martínez-Ferreiro, S., Fortescue, M. D., and Boye, K. (2020). Non-fluent aphasia in a polysynthetic language: Five case studies. *Aphasiology*, 34(6):675–694.

Nordlinger, R. and Sadler, L. (2004). Nominal tense in crosslinguistic perspective. *Language*, pages 776–806.

Novick, J. M., Trueswell, J. C., and Thompson-Schill, S. L. (2010). Broca's area and language processing: Evidence for the cognitive control connection. *Language and Linguistics Compass*, 4(10):906–924.

Noyer, R. (1998). Vietnamese 'morphology' and the definition of word. *University of Pennsylvania Working Papers in Linguistics*, 5(2):5.

Papillon, M. (2020). *Precedence and the lack thereof: Precedence-relation-oriented phonology*. PhD thesis, University of Maryland, College Park.

Penke, M. and Westermann, G. (2006). Broca's area and inflectional morphology: Evidence from Broca's aphasia and computer modeling. *Cortex*, 42(4):563–576.

Pierrehumbert, J. (2000). Tonal elements and their alignment. In *Prosody: Theory and Experiment: Studies Presented to Gösta Bruce*, pages 11–36. Springer.

Pinker, S. (1998). Words and rules. *Lingua*, 106:219–242.

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

Pinker, S. and Prince, A. (1988b). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, 28(1-2):73–193.

Pinker, S. and Ullman, M. T. (2002). The past and future of the past tense. *Trends in Cognitive Sciences*, 6(11):456–463.

Platzack, C. (2013). Head movement as a phonological operation. *Diagnosing syntax*, 46:21–43.

Plaut, D. C. (1995). Double dissociation without modularity: Evidence from connectionist neuropsychology. *Journal of Clinical and Experimental Neuropsychology*, 17:291–321.

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

Plunkett, K. and Marchman, V. A. (1991). U-shaped learning and frequency effects in a multilayered perceptron: Implications for child language acquisition. *Cognition*, 38:43–102.

Pollard, C. and Sag, I. A. (1994). *Head-driven phrase structure grammar*. University of Chicago Press.

Pollock, J.-Y. (1989). Verb movement, universal grammar, and the structure of IP. *Linguistic Inquiry*, 20(3):365–424.

Prather, P., Zurif, E., Love, T., and Brownell, H. (1997). Speed of lexical activation in nonfluent Broca's aphasia and fluent Wernicke's aphasia. *Brain and language*, 59(3):391–411.

Preminger, O. (2021). Natural language without semiosis. In *Presentation at the 14th annual conference on Syntax, Phonology and Language Analysis (SinFonIJA 14)*.

Pullum, G. K. (1989). The great Eskimo vocabulary hoax. *Natural Language & Linguistic Theory*, pages 275–281.

Rafferty, M. B., Saltuklaroglu, T., Reilly, K., Paek, E. J., and Casenhiser, D. M. (2023). Neural synchrony reflects closure of jabberwocky noun phrases but not predictable pseudoword sequences. *European Journal of Neuroscience*, 57(11):1834–1847.

Raimy, E. (2000). *The phonology and morphology of reduplication*, volume 52. Walter de Gruyter.

Rajalingham, R., Issa, E. B., Bashivan, P., Kar, K., Schmidt, K., and DiCarlo, J. J. (2018). Large-scale, high-resolution comparison of the core visual object recognition behavior of humans, monkeys, and state-of-the-art deep artificial neural networks. *Journal of Neuroscience*, 38(33):7255–7269.

Rauschecker, J. P. and Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12(6):718–724.

Roelofs, A., Meyer, A. S., and Levelt, W. J. (1998). A case for the lemma/lexeme distinction in models of speaking: Comment on caramazza and miozzo (1997). *Cognition*, 69(2):219–230.

Rumelhart, D. and McClelland, J. (1986). On learning the past tenses of English verbs. In McClelland, J., Rumelhart, D., and Group, t. P. R., editors, *Parallel Distributed Processing: Explorations in the microstructure of cognition. Volume 2: Psychological and biological models*, pages 216–271. MIT Press, Cambridge, MA.

Sag, I. A., Baldwin, T., Bond, F., Copestake, A., and Flickinger, D. (2002). Multiword expressions: A pain in the neck for NLP. In *International conference on intelligent text processing and computational linguistics*, pages 1–15. Springer.

Sahin, N. T., Pinker, S., Cash, S. S., Schomer, D., and Halgren, E. (2009). Sequential processing of lexical, grammatical, and phonological information within broca's area. *Science*, 326(5951):445–449.

Sánchez, J., Trueman, A., Leyva, M. F., Alvarez, S. L., Blanco, M. T., Jung, H.-K., St Amour, L., and Harley, H. (2017). *An Introduction to Hiaki Grammar: Hiaki Grammar for Learners and Teachers, Volume 1*. CreateSpace Independent Publishing Platform, Scotts Valley, California.

Sandler, W. (1999). Cliticization and prosodic words in a sign language. *Studies on the phonological word*, 174:223.

Schriefers, H., Teruel, E., and Meinshausen, R.-M. (1998). Producing simple sentences: Results from picture–word interference experiments. *Journal of Memory and Language*, 39(4):609–632.

Selkirk, E. (1995). Sentence prosody: Intonation, stress, and phrasing. *The handbook of phonological theory*, 1:550–569.

Selkirk, E. (2014). The prosodic structure of function words. In *Signal to syntax*, pages 199–226. Psychology Press.

Shih, S. S. and Zuraw, K. (2017). Phonological conditions on variable adjective and noun word order in Tagalog. *Language*, 93(4):e317–e352.

Slobin, D. I. (1996). *From 'thought and language' to 'thinking for speaking.'*, volume 17 of *Studies in the social and cultural foundations of language*, page 70–96. Cambridge University Press.

Smolensky, P. (2012). Symbolic functions from neural computation. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 370(1971):3543–3569.

Smolensky, P. and Legendre, G. (2006). *The harmonic mind: From neural computation to optimality-theoretic grammar (Cognitive architecture), Vol. 1.* MIT press.

Sprenger, S. A., Levelt, W. J., and Kempen, G. (2006). Lexical access during the production of idiomatic phrases. *Journal of Memory and Language*, 54(2):161–184.

Strijkers, K. and Costa, A. (2011). Riding the lexical speedway: A critical review on the time course of lexical selection in speech production. *Frontiers in Psychology*, 2.

Strijkers, K., Costa, A., and Thierry, G. (2010). Tracking lexical access in speech production: Electrophysiological correlates of word frequency and cognate effects. *Cerebral Cortex*, 20(4):912–928.

Takashima, A., Konopka, A., Meyer, A., Hagoort, P., and Weber, K. (2020). Speaking in the brain: the interaction between words and syntax in sentence production. *Journal of Cognitive Neuroscience*, 32(8):1466–1483.

Travis, L. d. (1984). *Parameters and effects of word order variation*. PhD thesis, Massachusetts Institute of Technology.

Turk, A. and Shattuck-Hufnagel, S. (2014). Timing in talking: what is it used for, and how is it controlled? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1658):20130395.

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, 92(1-2):231–270.

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

Ünal, E. and Papafragou, A. (2016). Production–comprehension asymmetries and the acquisition of evidential morphology. *Journal of Memory and Language*, 89:179–199.

Ussishkin, A., Dawson, C. R., Wedel, A., and Schluter, K. (2015). Auditory masked priming in maltese spoken word recognition. *Language, Cognition and Neuroscience*, 30(9):1096–1115.

Venezia, J. H., Vaden Jr, K. I., Rong, F., Maddox, D., Saberi, K., and Hickok, G. (2017). Auditory, visual and audiovisual speech processing streams in superior temporal sulcus. *Frontiers in Human Neuroscience*, 11:174.

Vosse, T. and Kempen, G. (2000). Syntactic structure assembly in human parsing: A computational model based on competitive inhibition and a lexicalist grammar. *Cognition*, 75(2):105–143.

Wagner, M. (2010). Prosody and recursion in coordinate structures and beyond. *Natural Language & Linguistic Theory*, 28:183–237.

Walker, J. P., Joseph, L., and Goodman, J. (2009). The production of linguistic prosody in subjects with aphasia. *Clinical Linguistics & Phonetics*, 23(7):529–549.

Williams, E. (1981). On the notions "lexically related" and "head of a word". *Linguistic Inquiry*, 12(2):245–274.

Wilson, S. M., Bautista, A., and McCarron, A. (2018). Convergence of spoken and written language processing in the superior temporal sulcus. *Neuroimage*, 171:62–74.

Wittenberg, E., Paczynski, M., Wiese, H., Jackendoff, R., and Kuperberg, G. (2014). The difference between “giving a rose” and “giving a kiss”: Sustained neural activity to the light verb construction. *Journal of memory and language*, 73:31–42.

Xiang, H.-D., Fonteijn, H. M., Norris, D. G., and Hagoort, P. (2010). Topographical functional connectivity pattern in the perisylvian language networks. *Cerebral cortex*, 20(3):549–560.

Yeaton, J. (2023). Evidence for a syntactic production-comprehension asymmetry in post-stroke aphasia. In *Annual meeting of the Society for the Neurobiology of Language*.

Yi, H. G., Leonard, M. K., and Chang, E. F. (2019). The encoding of speech sounds in the superior temporal gyrus. *Neuron*, 102(6):1096–1110.

Yu, X., Tian, X., and Lau, E. (2024). Electrophysiological responses to syntactic and ‘morphological’ structures: evidence from mandarin chinese. *bioRxiv*, pages 2024–01.

Zeiler, M. D. (2012). Adadelta: an adaptive learning rate method. *arXiv preprint arXiv:1212.5701*.