Everything Everywhere All at Once*

Ash Asudeh

Department of Linguistics & Center for Language Sciences
University of Rochester

October 4, 2023
Moscow State University

lrfg.online

1 Background

- The framework of Lexical-Realizational Functional Grammar (L_RFG) was founded by me and Dan Siddiqi (Carleton University), but the L_RFG research group now includes many international collaborators, including Oleg Belyaev of this very institution.
- L_RFG seeks to develop a formal and theoretical framework that couples the constraint-based approach to syntax of Lexical-Functional Grammar (LFG; Bresnan et al. 2016a) with the realizational, morpheme-based approach to word-formation of Distributed Morphology (DM; Halle and Marantz 1993).
- While this synthesis may seem superficially quixotic, it has surprising strengths:
  1. L_RFG inherits a robust approach to *polysynthesis* from DM.
  2. L_RFG inherits a robust approach to *nonconfigurationality* from LFG.
  3. L_RFG offers a version of DM that promises more coherence, since it is 'constraints all the way down,' in contrast to standard DM approaches, which assume a derivational approach to syntax (typically, some version of Minimalism) coupled with a constraint-based approach to form (typically, some version of Optimality Theory).

*This talk is based on previous work with Tina Bögel (Konstanz/Frankfurt), Dan Siddiqi (Carleton University), and Paul Melchin (former Carleton University), in particular Asudeh et al. (2023), Asudeh and Siddiqi (2023), Melchin et al. (2020b), and Asudeh et al. (2021). We are also collaborating with Oleg Belyaev (Moscow State University), Bronwyn Bjorkman (Queen's), Mike Everdell (UT Austin), and Will Oxford (Manitoba). Further details about the L_RFG project and previous work can be found here: lrfg.online. I accept sole responsibility for any errors in this talk.

2 Goals of the talk

- The aim of this talk is to give a full explanation of the following diagram, which schematizes the L_RFG *exponence* function, ν.

(1) $\langle$ [C_1,...,C_n] , F ∪ G ∪ I $\rangle$ $\xrightarrow{\nu}$ [ ]$_{v\text{-}structure}$
distribution *function/meaning*

- The lefthand side of the exponence function is a bundle of information, the *exponendum*, split into a part about *distribution* and a part about *function/meaning*.
- The righthand side of the exponence function is another bundle of information, the *exponent*, represented as a feature structure called a v(ocabulary) structure.
- Along the way, I will provide aspects of L_RFG analyses of English, Spanish, and Ojibwe.
- I will answer the following questions:
  1. Why should exponenda be represented as tuples that separate distribution and function/meaning?
  2. What kind of information is necessary in exponents?
  3. What is the relationship between (*morphological*) *exponence* and (*phonological*) *realization*?

2.1 L_RFG and Distributed Morphology

- DM is a framework for morphological theory. Like any morphological framework, it assumes an interface with a syntactic module.
- However, unlike other realizational models of morphology, DM has always assumed a particular syntactic framework for providing those structures: standard Chomskyan syntax (here called Minimalism for short, even though DM predates Minimalism strictly construed).
- Given this consistent co-occurrence, it can be hard to easily identify where Minimalism ends and where DM begins.
  - For example, the Y-model is a feature of Minimalism that DM inherits, as is the fact that the consequences of operations such as head-movement live in the PF branch of that Y.
- So what is the essence of DM, without Minimalist or Minimalist-like assumptions about syntax?

1. **Morpheme-based morphosyntax**
   - In a morpheme-based model of morphology,
     (a) complex words have internal structure (we set aside here the completely confusing use of "morpheme" in the DM literature and adopt the standard definition of the morphemic hypothesis); and
     (b) morphology is (often strictly) concatenative.
   - This means that words are not atomic in DM, and paradigms are epiphenomenal.
   - Stump (2001) calls this property of morphological theories *lexical* (yes, this is also confusing).
   - Parsimony dictates that there should not be two structure-building modules (morphology and syntax), if a theory can achieve the same explanatory power with just one (see, for example, Lieber 1992).
   - This property of DM is often called "syntax all the way down" in the DM literature.

2. **Realization**

- A *realizational* model of morphology (Beard 1995) assumes that morphology *expresses* information rather than adding information.
- In contrast, *incremental* approaches (using Stump's opposition) to morphology assume that morphology is information-adding.
- Realizational approaches typically assume bundles of features as the input to morphological operations.
- The DM literature uniquely calls this property *late-insertion*, but in reality most contemporary models of morphology are realizational (see Siddiqi and Harley 2016, editors' notes, for discussion).
- Note that this property, realization, and the previous property, morpheme-based morphosyntax, are not linked properties.
  - For example, A-morphous Morphology (Anderson 1992) and Paradigm Function Morphology (Stump 2001) are word and paradigm realizational models.
  - In Stump's (2001) classification, such models are *Inferential-Realizational*.
  - DM is in contrast *Lexical-Realizational*.
  - Nanosyntax (Starke 2009) is another Lexical-Realizational model of morphology.

3. **Morphology as an interface**

- In contemporary DM, the ideal appears to be that morphology is an interface rather than a separate generative component of the grammar.

  > "Morphology has no proprietary categories, but deals only in morphs, understood as pieces of phonological material lexically specified with instructions for their use as exponents of syntactic properties." (Bermúdez-Otero and Luís 2016: 311)

- This is most relevant in DM's rejection of the *morphome* (Aronoff 1994), the word, and the paradigm.
- In the DM literature, this property is typically called *non-lexicalist* because the model of syntax that results from an interface with DM has been taken to necessarily reject the *Lexicalist Hypothesis/Lexical Integrity Hypothesis* (Chomsky 1970, Lapointe 1980).
- It is worth a moment to acknowledge that this is a not a settled view of DM.
  - In Halle and Marantz (1993), M-structure clearly had proprietary rules.
  - Indeed, a view of DM in the literature today is that the post-spellout, PF branch rules, (such as *Local Dislocation*, *Rebracketing*, *Impoverishment*, *Fusion*, *Fission*, *Enrichment*, and *Readjustment*) constitute a morphological component of the grammar, i.e. a Lexicon.[1]
  - However, it is very important for the present discussion that all these operations serve to address syntax-morphology mismatches.
    - We assume here that such mechanisms are not properties of DM *per se*, but rather are specifically properties of the DM–Minimalism interface, which is why they are typically referred to as PF operations.

4. **Three lists**

- Perhaps the most salient feature of DM that makes it distinct from other models of morphology is that it gave up on the claim that syntactic, semantic, and phonological domains of "word" (or morpheme) align on the same domain (see Marantz 1997 for discussion).

[1] Similarly, a version of Minimalist Syntax that does not assume a DM interface, such as those that appear in most syntax textbooks, does not assume any of these mechanisms.

- Indeed, Distributed Morphology gets its name from this property.
- Instead it created three distinct lists, one for each domain.
  (a) The *Vocabulary* is the list for phonological properties.
  (b) The *Encyclopedia* is the list for semantic properties.
  (c) The third list is frustratingly without an agreed-upon name in the literature.
    - Its members are the formal features that populate syntactic structures (these are the things called "morphemes" by some practitioners of DM).

5. **Elsewhere Principle**

- DM employs a common type of linguistic rule called the *Elsewhere Principle* (also called the *Paninian Principle*), which is almost standard in morphological theory and dates back to the 70s, when formal morphological theory saw it nascence (see for example Anderson 1969, Matthews 1972, Kiparsky 1973, and Aronoff 1976).
- The classic view of the elsewhere condition is seen in Aronoff (1976): a morphological rule with more specific conditions is applied before a general rule.
- This has long been the main means through which irregular rules block the application of regular rules, for example.
- In contemporary realizational models, this blocking is achieved through competition, where a more specific form outcompetes a more general form.
- In the DM literature, this is called the *Subset Principle*.

6. **Underspecification**

- A fundamental challenge of any morphological theory is syncretism or polysemy.
- This is the simple phenomenon that morphological forms appear in many different environments, often with slightly different meanings.
- These are a violation of bi-uniqueness (a unique form maps to a unique meaning; see Harris 2016 for detailed discussion).
- There are many ways that morphological models can deal with this, such as rules of referral (Stump 2016).
- DM, like many other models, capture this by assuming that a morphological form underdetermines its syntactic and semantic properties.

- In sum, for us, DM without a Minimalism interface is precisely the combination of those six properties, listed here:

  1. Morpheme-based morphosyntax
  2. Realization
  3. Morphology as an interface
  4. Three lists ("Distributed" Morphology)
  5. Elsewhere Principle
  6. Underspecification

- L_RFG also has all of these properties.

$\Rightarrow$ L_RFG is a *version of DM*, but one which is entirely constraint-based and not based on Minimalist syntactic assumptions.

3 L_RFG and LFG

- LFG is standardly a *lexicalist* syntactic theory (Chomsky 1970, Lapointe 1980):

  (2) *Lexicalist Hypothesis*
  No syntactic rule can refer to elements of morphological structure. (Lapointe 1980: 8)

- In LFG, this is captured in the *Lexical Integrity Principle*, through formulations like the following:

  (3) *Lexical Integrity*
  No constituent structure rule may order any element into or out of lexical categories such as N, A, V. That is, constituent structure rules are blind to the internal structure of lexical categories. (Simpson 1983: 74)

  (4) *Lexical Integrity*
  Words are built out of different structural elements and by different principles of composition than syntactic phrases. (Bresnan and Mchombo 1995: 181)

  (5) *Lexical Integrity*
  Morphologically complete words are leaves of the c-structure tree, and each leaf corresponds to one and only one c-structure node. (Bresnan et al. 2016b: 92)

- There are three distinct possible positions on lexicalism, with LFG standardly subscribing to the first and lexicalist derivational theories, such as Chomsky's own theories up to and including the Minimalist Program, subscribing to the second. Distributed Morphology (Halle and Marantz 1993) subscribes to the third (Marantz 1997):

  (6) *Strong Lexicalism*
  Morphology and syntax are entirely distinct: inflectional and derivational morphology are captured by a separate morphological component.

  (7) *Weak Lexicalism*
  Inflectional morphology is part of the syntactic component but derivational morphology is captured by a separate morphological component.

  (8) *Non-Lexicalism*
  Inflectional morphology and derivational morphology are part of the syntactic component: there is no separate morphological component.

- The obvious point of contrast between L_RFG and LFG concerns the Lexicalist Hypothesis in (2) above.

- L_RFG is a kind of non-lexicalist LFG.
  - This may sound contradictory, but it's important to separate LFG *theory*, which assumes strong lexicalism, from the LFG *formalism* used to state the theory.
  - The formalism and formal tools can equally well formalize a non-lexicalist version of LFG, like L_RFG.

- In fact, L_RFG even upholds part of LFG's Lexical Integrity Principle.

- The definition of Lexical Integrity in (5) has two parts:

  1. L_RFG *upholds* the part that states that "each leaf corresponds to one and only one c-structure node".

- This may contrast with Lexical Sharing (Wescoat 2002, 2005, 2007), in which portmanteau forms like *du* ('of.DEF.MASC.SG') in French appear to correspond to more than one c-structure node. We need to look under the hood carefully, though, to see what the formal definition of Lexical Sharing is rather than simply going by its graphical representation, which may be misleading. We haven't done this work yet.

2. L_RFG *rejects* the part that states that "morphologically complete words are leaves of the c-structure tree".
   - The c-structure leaves/terminals in L_RFG are not "morphologically complete words". The c-structure leaves/terminals are feature bundles that *map* to form, but the form itself is not part of the terminal node.

- However, notice that the notion *morphologically complete word* is left unanalyzed in the definition in (5).
- In fact, it is far from clear that "morphologically complete word" is a coherent notion (see, for example, Anderson 1982).
  - The essential problem is that there are multiple relevant notions of wordhood, and they don't align on a single type of object that we can point to and unambiguously and confidently call a word (Di Sciullo and Williams 1987).[2] In fact, there can be mismatches between the phonological, syntactic, and semantic aspects of words (Marantz 1997).

1. Portmanteau words are examples of things that are phonologically simple but semantically and syntactically complex.

(9) Tu bois **du** lait. French
you drink of.DEF.MASC.SG lait
'You drink/are drinking milk.'

(10) **I'mma** go. English dialect
1SG.FUT.PROX go
'I'm about to go.'

- Idiomatic expressions are phonologically and syntactically complex, but not necessarily semantically complex, and never in a way that maps entirely transparently to their phonology and syntax.

(11) I read **the shit out of** this book.
INTENSIFIER
'I thoroughly read this book.'

- Units of syntax can be phonologically or semantically dependent on their contexts.

(12) Je **l'**ai vu. French clitic
I 3SG.saw
'I saw it.'

(13) **The cat**'s been let out of the bag.

[2] This is a long and broad discussion that we cannot possibly do justice to here.

- L_RFG thus countenances three criteria for wordhood:
  1. A word as an unanalyzed phonological string (phonological criterion)
  2. A word as a lexicalized string with a non-compositional meaning (semantic criterion)
  3. A word as a syntactic atom (syntactic criterion)
- L_RFG thus assumes that there are three notions of wordhood that sometimes happen to align, but can diverge, i.e., there are mismatches between the three types of wordhood.
- With its focus on mismatches, L_RFG is therefore strongly in the spirit of LFG.
  - L_RFG uses the standard *co-description* mechanism of LFG (for recent exposition, see Dalrymple et al. 2019) to simultaneously state the phonological, syntactic and semantic aspects of formatives.
- Here are some possible points of comfort for an LFGer gazing on L_RFG's familiar yet alien landscape:
  1. L_RFG could be considered to be offering a morphological theory for LFG that had previously been captured by somewhat ad hoc devices like phrase structure rules for word formation; see, e.g., the discussions of Japanese and West Greenlandic in Bresnan et al. (2016b). In other words, LFG owes some kind of theory of word structure, which has generally been lacking until recently (see, e.g., Dalrymple 2015, Dalrymple et al. 2019), and L_RFG seeks to pay that debt.
  2. The vocabulary items of L_RFG contain much the same information as LFG's lexical entries, but without the commitment that morphophonological form is bundled as part of the lexical entry. It should be easy to specify an algorithm for translating L_RFG's VIs into LFG lexical entries.
  3. Related to the first two points, if one were to want to maintain some version of the Lexicalist Hypothesis, one could view L_RFG as offering a microscopic view of the structure of "words", in particular major categories like verb and noun. For example, the TP node in (15) in some sense *is* the verb, but the L_RFG c-structure shows its internal structure.

⇒ L_RFG is a *version of LFG*, formally, but is distinct from standard, syntactically lexicalist LFG.

4 Lexical-Realizational Functional Grammar

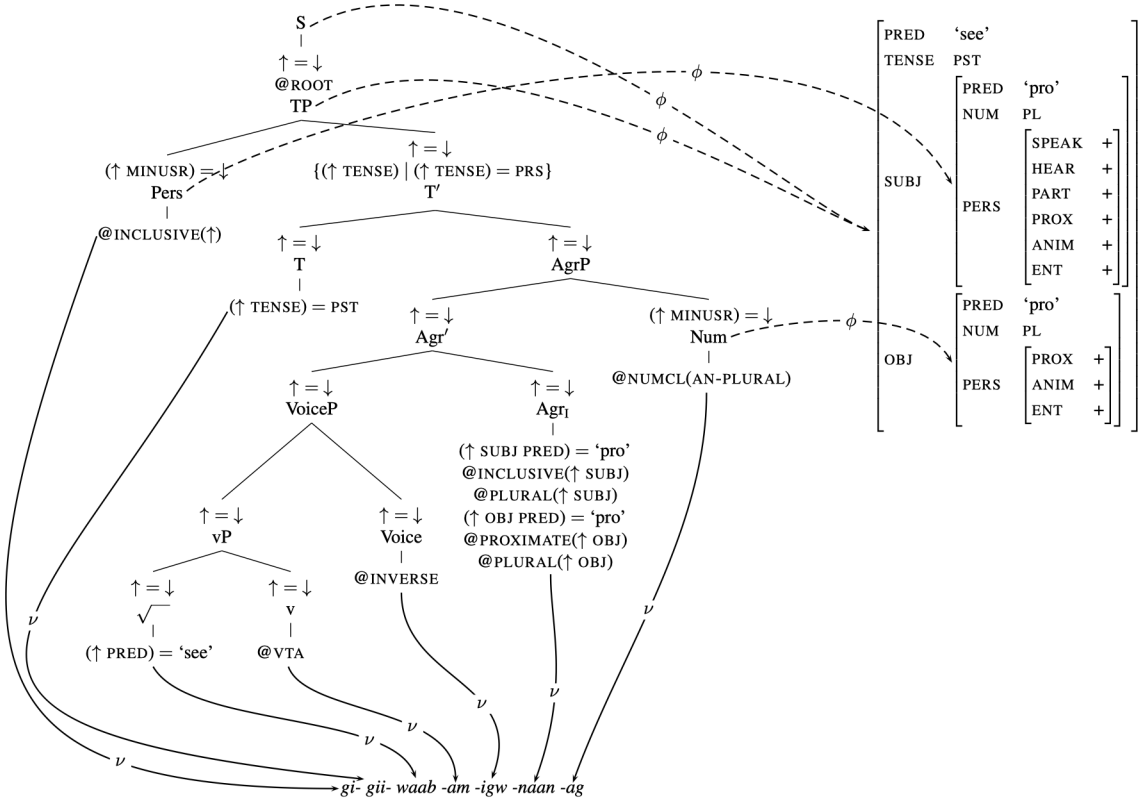
- L_RFG is similar to standard LFG, with changes to the c-structure and its relationship with morphosyntactic elements.
- The terminal nodes of c-structures *are not words*, but instead are exponenda, i.e. distribution–function/meaning pairs. Recall from (1) above:

(14) $\langle$ $[C_1,\ldots,C_n]$, $F \cup G \cup I$ $\rangle$
distribution *function/meaning*

- These c-structure terminal nodes are mapped by a correspondence function, ν, to *vocabulary structures* (*v-structures* for short).
- Here is an example from Ojibwe (15) to demonstrate the basics of an L_RFG analysis (16). Note that the exponent v-structures are here abbreviated by their form.

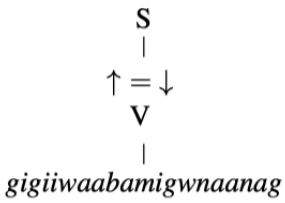
(15) gi- gii- waab -am -igw -naan -ag
2 PST see VTA INV 1PL 3PL
'They saw us(incl).'

(16)



- In contrast, the c-structure for (15) in standard LFG would like this:

(17)



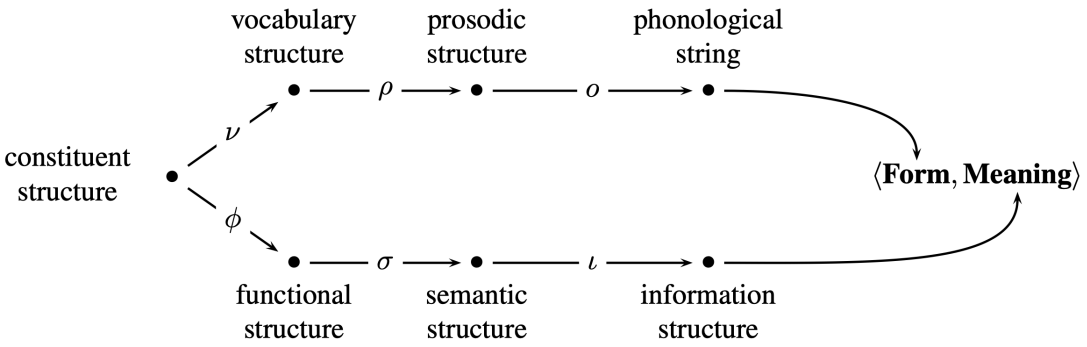
f-description from (16)

- While (17) is admirably syntactically simple, it is not particularly morphologically illuminating.
- LFG's morphological assumptions thus exhibits the typical tension between *word-based* and *morpheme-based* morphology:
  1. Word-based morphology is attractive for *fusional* languages and 'pure' morphology, as evinced very well by Indo-European.
     - It tends to be less illuminating for highly *agglutinative* (e.g., Finno-Ugric, Turkic) or *polysynthetic* languages (e.g., many North-American Languages).
  2. Morpheme-based morphology is attractive for agglutinative and polysynthetic languages.
     - It tends to be less illuminating for fusional languages and often denies that there is even any such thing as 'pure' morphology.
- LFG theory's assumption of strong lexicalism has pushed it towards a word-based realizational theory of morphology, such as PFM.

- L_RFG uses LFG's formal tools to instead build a morpheme-based theory of morphology, as in DM.
- The grammatical architecture of L_RFG looks like this:

(18)

Form: *Morphology-Prosody-Phonology (MPP) Path*



Meaning: *Structure/Function/Meaning (SFM) Path*

- The output of the grammar, ⟨**F, M**⟩, for any particular set of input formatives, is a form–meaning pair where the form incorporates prosody (still fed by constituent structure) and the meaning incorporates information structure (still fed by semantic structure).[3]
- The relationship between terminal nodes and VIs is many-to-one, using the mechanism of *Spanning* (Ramchand 2008, Merchant 2015, Haugen and Siddiqi 2016, Svenonius 2016); i.e. one VI may realize features of multiple terminal nodes.
- The result is similar to the Lexical Sharing model proposed for LFG by Wescoat (2002, 2005, 2007), but maintains the complex internal structures of words as part of syntax.
- Lastly, it should be mentioned that L_RFG is not the only realizational approach to morphology in LFG. The realizational framework of Spencer (2013) was designed with constraint-based theories like LFG in mind. Paradigm Function Morphology Stump (2001, 2016) has been used as a realizational framework for morphology by Sadler and Nordlinger (2006), Dalrymple (2015), Dalrymple et al. (2019), and Thomas (2021), among others.
  - A key distinction between L_RFG and these proposals is that L_RFG does not posit a separate class of morphological features for exponence, instead assuming that exponence is sensitive to regular syntactic features (f-structural information), compositional semantics, and information structure.
  - Positing an additional strictly morphological class of features generally results in a lot of redundancy, which is to be avoided if possible for reasons of theoretical elegance.

[3] Note that the *set* of all grammatical form-meaning pairs may have a given form recurring in several pairs, if it is ambiguous, or a given meaning recurring in several pairs, if it is expressible in alternative ways.

5 Exponence in L_RFG

- The theory of exponence is ultimately a theory of *exponents*.
- An exponent is a morphological representation that serves as the interface between an *exponendum* and a (*phonological*) *realization*.
- In L_RFG, an exponent is represented as a *vocabulary structure*, or *v-structure* for short.
- The *Vocabulary* in L_RFG is the mapping from the set of exponenda, the set of lefthand sides of *vocabulary items* (VIs), to the set of exponents, the set of righthand sides of vocabulary items, i.e. the set of v-structures.
- The Vocabulary is thus nothing more or less then a set of vocabulary items, i.e., a set of pairs of exponenda and exponents.
- A vocabulary item is represented as in (1) above, repeated here:

(1) $\langle$ [C$_1$,...,C$_n$] , F $\cup$ G $\cup$ I $\rangle$ $\xrightarrow{v}$ [ ]$_{\textit{v-structure}}$
distribution *function/meaning*

- The tuple in (1) is the representation of an exponendum.
- It is mapped by ν, the exponence function from exponenda to exponents, to its exponent, represented as a vocabulary structure.[4]

5.1 Distribution

- The first member of an exponendum pair is a list of categories, which represents some part of the terminal yield of an L_RFG c-structure.
  - Thus, the first member of the pair encodes the vocabulary item's syntactic *distribution*.
- L_RFG assumes the morphosyntactic operation of *spanning* (Ramchand 2008, Svenonius 2016, Merchant 2015, Haugen and Siddiqi 2016).
  - *Spans* are lists of c-structure categories that are involved in many-to-one cases of exponence in which a single v-structure expones multiple c-structure nodes.
  - This is why the *distribution* coordinate in vocabulary items is a list rather than a single category. There are two kinds of spanning in L_RFG:
    1. *Vocabulary Spanning*: the case where the category list in the first coordinate of an exponendum has length greater than one; i.e., vocabulary spanning is a matter of listing in the Vocabulary.
    2. *Pac-Man Spanning*: the case where some category would be left unexponed and is instead mapped to a neighbouring exponent. In other words, Pac-Man spanning is a matter of the ν-mapping being a total function from the domain of c-structure nodes to the co-domain of v-structures.
  - As a result of the two kinds of spanning, ν must be a many-to-one function.

[4] Note that although the exponence function ν expects a pair of a list and a set as its argument, as in $\nu(\langle[\alpha], \{\_\_\}\rangle) = \beta$, we often abbreviate this as $\nu(\alpha) = \beta$, since using the category is often sufficient for identification.

5.2 Function/meaning

- The second member of the pair on the lefthand side of the abstract vocabulary item in (1) is the union of a set of descriptions of f-structures, F, a set of descriptions of s-structures and Glue semantics meaning constructors, G, and a set of descriptions of i-structures, I.
- Any of these sets may be empty. This union represents the *function/meaning* of the vocabulary item.
- In order to make it easier to refer to this union, we call it a *fugui*.

5.3 Exponents

- Let's now turn to the output/righthand side of the ν-mapping in (1), a vocabulary structure.
- A v-structure is modelled as an attribute-value matrix, similarly to f-structure.
  - Attributes are symbols, like DEPENDENCE.
  - Values are symbols, strings, v-structures, or sets of symbols.[5]
  - On analogy with f-structures and f-descriptions, v-structures are *described* by v-descriptions, a set of defining equations and constraints that picks out the minimal satisfying v-structure, if any, as its model.
- This is the general form of an exponent/v-structure in L_RFG:

(19)

$$
\begin{array}{l}
v\text{-}s \left[\begin{array}{ll}
\text{PHON(OLOGICAL) REP(RESENTATION)} & \textit{phon. realization \& conditions} \\
 & \textbf{string} \\
\text{P(ROSODIC) FRAME} & \textit{prosodic unit} \\
 & \textbf{string wrapped around} \\
 & \textbf{ρ-correspondent of this v-s} \\
\text{P(ROSODIC) DOMAIN} & \textit{prosodic unit} \\
 & \textbf{string wrapped around} \\
 & \textbf{value of } \text{PFRAME} \\
\text{DEP(ENDENCE)} & \{\text{LT, RT}\} \\
 & \textbf{set of symbols} \\
\text{CLASS} & \{\textit{inflectional classes}\} \\
 & \textbf{set of symbols} \\
\text{HOST} & v\text{-}s \left[\begin{array}{ll}
\text{IDENT(ITY)} & + \\
 & \textbf{symbol} \\
\text{PHONREP} & \ldots \\
\text{PFRAME} & \ldots \\
\text{DEP} & \ldots \\
\text{CLASS} & \ldots
\end{array}\right]
\end{array}\right]
\end{array}
$$

[5] Sets can obviously be generalized to contain any of the other kinds of values.

- Thus, the general form of a vocabulary item (VI) in L_RFG is as follows:

(20) $\langle$ [C$_1$,…,C$_n$], F ∪ G ∪ I $\rangle$ $\xrightarrow{\nu}$ $\begin{bmatrix} \text{PHON(OLOGICAL) REP(RESENTATION)} & \textit{phon. realization \& conditions} \\ \text{P(ROSODIC) FRAME} & \textit{prosodic unit} \\ \text{P(ROSODIC) DOMAIN} & \textit{prosodic unit} \\ \text{DEP(ENDENCE)} & \{\text{LT, RT}\} \\ \text{CLASS} & \{\textit{inflectional classes}\} \\ & \textit{v-s} \begin{bmatrix} \text{IDENT(ITY)} & + \\ \text{PHONREP} & \dots \\ \text{PFRAME} & \dots \\ \text{DEP} & \dots \\ \text{CLASS} & \dots \end{bmatrix} \\ \text{HOST} & \textit{v-s} \end{bmatrix}$

distribution *function/ meaning*

- We adopt the convention of writing the value of a set-valued feature without set-brackets when it is a singleton set; e.g. [CLASS *weak*] instead of [CLASS {*weak*}].
- Similarly, in descriptions we will drop the ∈ feature in paths and write (v DEP) = LT instead of (v DEP ∈) = LT or LT ∈ (v DEP).

- The following example shows the VI for the English *-en* suffix that is a deadjectivizing verbalizer (as in, e.g., *blacken* and *soften*) case whose analysis I will provide and motivate:

(21) $\langle$ [v$_a$], { $\lambda P \lambda v.$**cause**(**become**(P))(v) : $((\uparrow_\sigma \text{E}) \multimap \uparrow_\sigma) \multimap ((\uparrow_\sigma \text{E}) \multimap \uparrow_\sigma)$ } $\rangle$ $\xrightarrow{\nu}$ $\begin{bmatrix} \text{PHONREP} & \text{/ən/} \\ \text{PFRAME} & ((\ \ )(\cdot)_\sigma)_{ft} \\ \text{PDOMAIN} & (\ (v\ \text{PFRAME})\ )_\omega \\ \text{DEP} & \text{LT} \\ \text{CLASS} & \text{WEAK} \\ \text{HOST} & \begin{bmatrix} \text{IDENT} & + \\ \text{PHONREP} & \text{/…([son])[obs]/} \\ \text{PFRAME} & (\ \rho(v\ \text{HOST})\ )_\sigma \end{bmatrix} \end{bmatrix}$

- The distribution coordinate of the exponendum in this VI is a simple singleton list: it is an exponent of category v$_a$.
- The function/meaning coordinate of the exponendum is a single Glue semantics meaning constructor (see, among others, Dalrymple 1999, Dalrymple et al. 2019, Asudeh 2012, 2022, 2023) that encodes its meaning.
  - There is no f-structural or information-structural information to express.
  - The semantics will be further fleshed out shortly.

5.4 Phonological and prosodic features

- The first lot of features in a v-structure concerns the phonology and prosody of the exponent.
  - The feature PHONREP (PHONO- LOGICAL REPRESENTATION) encodes the exponent's underlying phonological representation and any conditions it places on its phonological context.
  - The feature P(ROSODIC) FRAME specifies any conditions the exponent places on its prosodic context.
  - The feature P(ROSODIC) DOMAIN encodes the prosodic level at which the exponent 'prosodifies', i.e. is integrated into the surrounding prosodic environment.
  - The feature DEP(ENDENCE) encodes the direction of prosodic dependence of the exponent.

5.4.1 Phonological representation

- PHONREP takes a string as a value.
- The string encodes conditions on the phonological form and context of the exponent.
- We typically represent the phonological form as segments, but we assume that these segments are underlyingly feature bundles.
- This allows aspects of the phonology to be underspecified.
  - For example, much of English inflection is probably underspecified for [±voice].
- The value can also encode a memorized, conditioned list.
  - For example, the English indefinite determiners (*a/an*) are listed, phonologically conditioned allomorphs. This would also form part of our approach to French liaison.

5.4.2 Prosodic frame

- PFRAME takes as its value a string wrapped around the prosodic correspondent of the v-structure.
  - The prosodic correspondent is the value of $\rho(\bullet)$, where $\bullet$ is the v-structure in question.
- The value encodes any conditions that the exponent places on its mapping to prosody, i.e. on its prosodic context.
  - For example, FUCK-insertion in English is sensitive to foot structure: *(ábso)fucking(lùtely* but **(ab-fuck)(ingso)(lutely)*.
- Similarly, *-um-* infixation (Austronesian) is sensitive to syllable structure (Orgun and Sprouse 1999, Roark and Sproat 2007: 30, 39–41).

5.4.3 Prosodic domain

- PDOMAIN takes as a value a string wrapped around (• PFRAME), i.e. the v-structure's PFRAME value.
- The value specifies in which prosodic domain the v-structure's ρ-correspondent is integrated into prosody according to some definition of prosodic phrasing at p-structure (Bögel 2015, 2021).
  - For example, using $\cdot$ to represent the ρ-correspondent of the v-structure in question, English geminates can only appear at [PDOMAIN $(\ \cdot\ )_\iota$], i.e. above the level of the prosodic word.

- Similarly, some Germanic prefixes are metrical, [PDOMAIN (·)$_\omega$], while others are extrametrical, [PDOMAIN (·),( )$_\omega$].
  - We use the comma to represent the unordered concatenation of two intonational units; the following equality therefore holds: [PDOMAIN (·),( )$_\omega$] = { [PDOMAIN (·)( )$_\omega$] | [PDOMAIN ( )$_\omega$(·)] }.
  - The actual order of comma cases must be set by the DEPENDENCE feature.
- The prosodic domain account for the different stress and phonotactic restrictions on affixation promises a viable analysis of German prefixes.
  - Those whose domain is [PDOMAIN (·)$_\omega$] are stressed.

    (22) *uralten* ('very old') German prefix; *(úr)(alten)* not *(urálten)*
  - In contrast, German prefixes whose domain is [PDOMAIN (·)$_\iota$] are unstressed (they are extrametrical).

    (23) *gealtert* ('aged') German prefix; *ge(áltert)* not *(gé)(altert)*

5.4.4 DEPENDENCE

- DEP takes a set of symbols as its value.
- The symbols encode the direction of the prosodic dependency: left (suffixes and left-leaning clitics), right (prefixes and right-leaning clitics), or both (infixes and mesoclitics).[6]
  - The value {LT}, typically abbreviated without the set brackets, encodes that the exponent v-structure is dependent to its left; i.e. the exponent is a suffix or left-leaning clitic.
  - The value {RT}, again typically abbreviated without the set brackets, encodes that the exponent v-structure is dependent to its right; i.e. the exponent is a prefix or right-leaning clitic.
  - The value {LT,RT} encodes that the exponent v-structure is dependent to both its left and right; i.e., the exponent is an infix or a mesoclitic (Harris 2002, Luís and Spencer 2004, Bögel 2015). In sum, the presence of this feature entails prosodic/phonological dependence.

5.5 Morphosyntactic features

- The second lot of features concern the morphosyntax of the exponent.
  - The feature CLASS encodes purely morphological distinctions, such as inflectional classes.
  - The feature HOST plays an important role in our theory of affix exponence and realization.
    - This feature relates the v-structure of an affix directly to the v-structure of its host/stem.
    - This direct relationship between affixes and hosts allows us to very locally encode effects that in other DM frameworks are modelled by derivational operations, such as *head movement* (Travis 1984) or *lowering* (Bobaljik 1994), which are operationalizations of *morphological merger* (Marantz 1984).[7]

[6] We assume here that circumfixes can be handled as a prefix/suffix combination, as in finite-state approaches (see, e.g., Beesley and Karttunen 2003). However, Bill Foley (p.c.) has suggested to us that there may be 'true' circumfixes that cannot be handled this way. If so, we could supplement DEP values with values like LEDGE (left edge) and REDGE (right edge).

[7] This is an earlier idea that DM has adopted.

5.5.1 CLASS

- Any theory of morphology needs to have some way of capturing purely morphological restrictions on distribution, such as an affix appearing with only a certain class of stems.
- In L_RFG, this is the purview of the CLASS feature, which takes a set of symbols as a value. This set encodes inflectional class and other purely morphological selectional properties.
  - For example, this is where we capture verb classes and noun classes, such as Latin conjugations and declensions.
- Furthermore, the CLASS feature allows L_RFG to make room for *morphomic effects* while maintaining, like other DM frameworks, that 'morphomes' do not entail the existence of a separate generative morphological component (Aronoff 1976, 1994).

5.5.2 HOST

- Another thing that any theory of morphology needs to capture is that affixes are phonologically and morphosyntactically conditioned.
- In L_RFG, we conceive of this conditioning as the affix constraining the possible hosts with which it can co-occur.
- This is accomplished through the HOST feature in v-structure.
- HOST encodes the relationship between an affix and its host non-derivationally, through (modified) equality: in other words, the value of the HOST of the v-structure exponent of an affix is itself another exponent v-structure.
- Most of the features in HOST are features that we have already encountered: PHONREP, PFRAME, PDOMAIN, DEP, and CLASS (any of which can be underspecified as usual).
- The HOST can also be specified for the IDENT(ITY) feature, which is either present with the value + or not present at all.
  - Thus, IDENT is effectively privative.
- Note that the HOST feature cannot contain HOST. This is captured by the *Principle of Local* HOST *Identification* (LHI) in (33) below.
  - The LHI uses the *restriction* operator (Kaplan and Wedekind 1993) to ensure that when a HOST is identified, it brings with it all of its features *except* HOST (if it has one). The LHI ensures that an exponent can include information about its HOST, but not its HOST's HOST.
- Thus, even though HOST takes a v-structure as its value, only a limited one-level embedding is possible in v-structures.
- The effects of morphological merger are controlled by a feature that can occur only in HOST, [IDENT +].
- Other than this feature, a HOST can be specified by an affix to have any of the v-structure features except HOST itself.
- We assume that the ρ-mapping from v-structure to p-structure is sensitive to the HOST feature.

- If a v-structure α has a HOST v-structure β, then β's realization in p-structure must be prosodified in the PDOMAIN of α's realization.
  - This is captured by the *Principle of* HOST *Mapping*:

(24) HOST *Mapping*
For all v-structures v, v':
$(v \text{ HOST}) = v' \Rightarrow \rho(v') \in \rho(v \text{ PDOMAIN})$

5.5.2.1 HOST: IDENTITY

- The feature IDENT(ITY) takes a symbol as a value.
- Its value is constrained to be the symbol $+$.
  - Thus, the feature is either present as [IDENT $+$] or not present at all.
  - It is a privative feature.
- The IDENT feature captures locality conditions on the c-structural and f-structural context of the host.
- If [IDENTITY $+$] is present in the HOST, then the exponent in question constrains the identity of its host as follows:

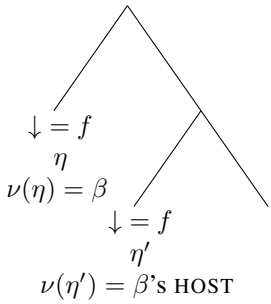
(25) HOST *Identification (Intuition)*
Given β, a v-structure containing the feature [HOST [IDENT $+$]], and η, a c-structure terminal node that maps to a vocabulary item that β expones, β's HOST is the v-structure that expones the *closest* c-structural terminal node to η that maps to the *same* f-structure as η.

- *Closest* is defined as follows:

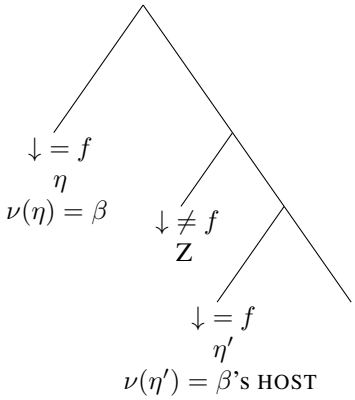
(26) Y is the closest c-structure node to X iff
  - X c-commands Y; and
  - there is no Z such that X c-commands Z and Z c-commands Y.

- The representations in (27–30) sketch two situations in which [IDENT $+$] is satisfied and two in which it is not.
- Note that in all cases, η is a c-structure node that corresponds to β; we abbreviate this as $\nu(\eta) = \beta$.

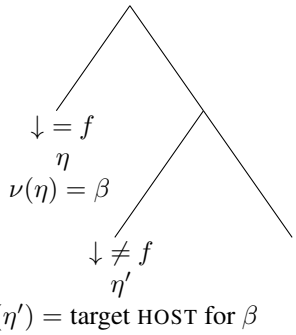
(27) [IDENT +] satisfied:
The node, η', that maps to β's HOST is the closest c-structure node to η that maps to the same f-structure as η.



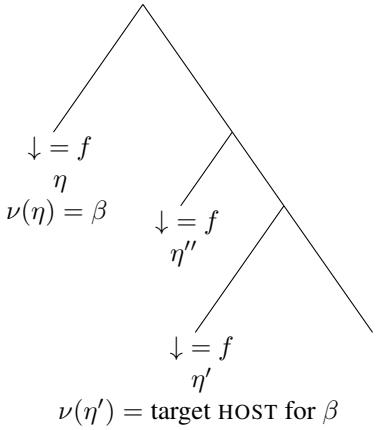
(28) [IDENT +] satisfied:
The node, η', that maps to β's HOST is the closest c-structure node to η that maps to the same f-structure as η.



(29) [IDENT +] not satisfied:
The node, η', cannot map to $\beta's$ HOST.
It is the closest terminal to η, but η and β's HOST do not map to same f-structure.



(30) [IDENT +] not satisfied:
The node, η', cannot map to $\beta's$ HOST.
It is not the closest c-structure node to η that maps to the same f-structure as η.



- We can use the term *f-domain* for the set of c-structure nodes that map to the same f-structure as some c-structure node α. We define a function to yield a node's f-domain.

  (31) For all c-structure nodes, n, in the set of c-structure nodes N for some c-structure,
  $$\textbf{f-domain}(n) = \{n' \mid n' \in N \wedge \phi(n') = \phi(n)\}$$

- Note that **f-domain** ensures that a node is in its own f-domain, since equality is a reflexive relation.
  - For example, **f-domain**(V), using the category label to stand in for the node, in a typical LFG analysis would include nodes labelled, V, V′, VP, T, T′, TP, C, C′, and CP.
- Since the function is reflexive, the minimal f-domain for any c-structure node that is ϕ-mapped is a singleton set containing that node itself. Only a c-structure node that is not mapped to f-structure can have an empty f-domain.
- We also define a function to calculate the closest c-structure terminal to a node, based on the informal definition in (26) above and an assumed standard definition of *c-command*.

- Note that since c-command is typically defined in terms of dominance and dominance is often construed to be a reflexive relation (see, e.g., Bresnan et al. 2016a: 136, fn.11), we explicitly exclude the case where a node reflexively satisfies **closest**.
  - A node should not count as the closest node to itself on both formal and theoretical grounds.
    - The formal objection is that every node would always be the closest node to itself, so it would rendeer the function pretty useless.
    - The theoretical objection is that this in turn would allow v-structures to be their own HOSTs, which fails to capture the intuition behind the notion.

(32) For all c-structure nodes, n, n', n'', in the set of c-structure terminal nodes T for some c-structure,
$$\textbf{closest}(n, n') \Leftrightarrow \textbf{c-command}(n, n') \wedge \neg[\textbf{c-command}(n, n'') \wedge \textbf{c-command}(n'', n')] \wedge n \neq n'$$

- We can capture the [IDENT +] constraint with the following global constraint on the c-structure/v-structure interface:

(33) *Local* HOST *Identification (LHI)*
For all c-structure nodes, n, n', in the set of c-structure nodes N for some c-structure,
$$(\nu(n') \text{ HOST IDENT}) = + \Rightarrow \textbf{closest}(n, n') \ \wedge \ n' \in \textbf{f-domain}(n) \ \wedge \ (\nu(n) \text{ HOST}) = \nu(n')\backslash\text{HOST}$$

- The definition in (33) uses the restriction operator (Kaplan and Wedekind 1993), $\backslash$, to state that $\nu(n)$'s HOST is the v-structure $\nu(n')$, *except* for any HOST information that $\nu(n')$ may contain.

- Note that this allows us to capture the notion of *bound stems*[8] as in:

(34) habl- Spanish
'talk'

  - Thus, a bound stem is a vocabulary item whose lefthand side contains a root and whose righthand side is listed as [IDENT +].

- That is, there are two ways for [IDENT +] to be marked on a v-structure:

1. By being specified as such on the righthand side of a vocabulary item; i.e. by being listed in the Vocabulary
2. By the v-structure being the HOST for some affix

- As a consequence, *exponence* (the ν-mapping) can be sensitive to [IDENT +] as a matter of being listed in the Vocabulary; this is the case of bound stems.

- But [IDENT +] can also be marked on an exponent that is not listed as such, due to operations in the grammar, namely HOST Mapping and Local HOST Identification.

[8] Bound stems are common in languages that require all roots to be inflected, such as Romance languages. Unqualified bound stems are harder to find in languages like English.

Interim summary

The general abstract representation of a vocabulary item is repeated below:

(20) $\langle [C_1,\ldots,C_n], F \cup G \cup I \rangle \xrightarrow{\nu}$

distribution *function/ meaning*

$$\left[\begin{array}{ll}
\textsc{phon(ological) rep(resentation)} & \textit{phon. realization \& conditions} \\
 & \textbf{string} \\
\textsc{p(rosodic) frame} & \textit{prosodic unit} \\
 & \textbf{string wrapped around} \\
 & \textbf{ρ-correspondent of this v-s} \\
\textsc{p(rosodic) domain} & \textit{prosodic unit} \\
 & \textbf{string wrapped around} \\
 & \textbf{value of } \textsc{pframe} \\
\textsc{dep(endence)} & \left\{\textsc{lt}, \textsc{rt}\right\} \\
 & \textbf{set of symbols} \\
\textit{v-s}\ \textsc{class} & \left\{\textit{inflectional classes}\right\} \\
 & \textbf{set of symbols} \\
 & \textit{v-s}\left[\begin{array}{ll}
\textsc{ident(ity)} & + \\
 & \textbf{symbol} \\
\textsc{phonrep} & \ldots \\
\textsc{pframe} & \ldots \\
\textsc{dep} & \ldots \\
\textsc{class} & \ldots
\end{array}\right] \\
\textsc{host} &
\end{array}\right]$$

5.6 MostSpecific

- L_RFG posits a constraint on the expression of phonological information, i.e. *morphophonology*, which we have called **MostSpecific** (Asudeh and Siddiqi 2023).
- **MostSpecific**(α, β) takes two exponents (v-structures), α and β, and returns whichever exponent has the most restrictions on its phonological context.
- The intuition behind **MostSpecific** is to prefer affixes over (otherwise compatible) clitics or free forms and to prefer clitics over (otherwise compatible) free forms.
  - In other words, we get the following preference order: affix $\gg$ clitic $\gg$ free form.
- In terms of information encoded in vocabulary items, choose the VI whose output v-structure contains more information, i.e. more features.
  - For example, if English comparative *-er*, an affix, and *more*, a free form, are in competition, then **MostSpecific** will select *-er*.
  - Similarly, if English verbal inflection *-s* and *does* are in competition, then **MostSpecific** will select *-s*.
- The proper subsumption relation on v-structures is used to formally capture the intuition behind **MostSpecific**: choose the exponent that contains the most information.

(35) Given two exponents (v-structures), α and β,

$$\textbf{MostSpecific}(\alpha,\beta) = \begin{cases} \alpha \textbf{ if } \beta\backslash\textsc{phonrep} \sqsubset \alpha\backslash\textsc{phonrep} \\ \beta \textbf{ if } \alpha\backslash\textsc{phonrep} \sqsubset \beta\backslash\textsc{phonrep} \\ \bot \textbf{ otherwise} \end{cases}$$

- In sum, the **MostSpecific** constraint is formalized as a function that takes two exponents, i.e. two v-structures, as arguments and returns whichever exponent contains the most information, using restriction to set aside PHONREP.
  - If neither candidate contains more information than the other (i.e., the two candidates are tied, but possibly have different PHONREPs) or they contain information that is incompatible with each other, the constraint returns $\bot$, meaning that neither candidate is better than the other with respect to the constraint.

6 Classifying forms: DEPENDENCE & IDENT

- Let's now see how v-structure features can be used to define various types of exponents.
- We'll also see how the particular features DEP and [IDENT +] can be used to form a factorial typology over types of exponents.
- The kinds of exponents that are of key interest for the typology are those for free forms, simple clitics, and affixes.
- In L_RFG, *free forms* are exponents that have the following features, with any value:

(36) $\begin{bmatrix} \textsc{phonrep} & \dots \\ \textsc{pframe} & \dots \end{bmatrix}$

  - Thus, free forms only specify their basic phonological features, in particular their underlying phonological form (PHONREP) and any constraints on their prosodic context (PFRAME).

- Adding features to this further constrains the exponent.
- The first added feature is DEP(ENDENCE), whose addition yields *simple clitics* or *leaners*. We arbitrarily call these *clitic*$_a$.
- These clitics have the following features:

(37) $\begin{bmatrix} \textsc{phonrep} & \dots \\ \textsc{pframe} & \dots \\ \textsc{dep} & \dots \end{bmatrix}$

  - For example, the English possessive *'s* and auxiliary *'ll* are specified as [DEP LT] because they lean on the preceding element.
  - We assume on general grounds that *'s* is the exponent of the category D and that *'ll* is the exponent of the category T.

(38) English possessive *'s*
  a. The car's fender
  b. The car you are in's fender
  c. The car you are exiting's fender

(39) English 'contractions'
  a. The person who arrives first'll leave last
  b. The person who passes out'll leave last
  c. The person who hides'll leave last
  d. The person who finds them'll leave last

- As the you can see in (38) and (39), leaners are not fussy about the category of the element that they attach to.
  - For example, in (38a), the leaner attaches to a noun, as is expected of a genitive marker, but in (38b) it attaches to a preposition, and in (38c) to a verb.
- Note that *'s* always happens to lean on a DP, but this is because there is always a DP in its specifier (Abney 1987).
- In contrast, *'ll* is not always preceded by a DP. For example, it can be preceded by a VP, provided the VP is the subject of the sentence:

  (40) To arrive on time'll always bring you happiness.
- The key thing to note is that the particular element at the right edge varies, and it's this element that is what the clitic's phonological form depends on; for example, it determines voicing assimilation: *the cat's meow* (voiceless) vs. *the car you are in's fender* (voiced).

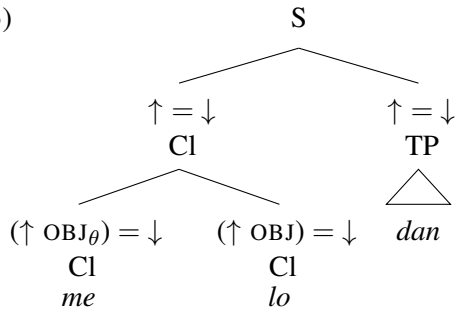
- Since the word 'clitic' is ambiguous throughout the literature, we want to distinguish these simple clitics from two other kinds, which we call respectively *phonological clitics* and *syntactic clitics*.
- Turning first to phonological clitics, these are a kind of clitic whose dependence properties are not determined by v-structure, but rather just by their phonology. We arbitrarily assign the term *clitic$_b$* to these *phonological clitics*.
  - For example, in the Frans Plank example, *drink a pint of milk*, the prosodic constituency is (*drinka*) (*pinta*) (*milk*) (Lahiri and Plank 2009).
  - The phonological dependence of these examples is entirely a product of prosodic structure i) footing together *drink* and the reduced form of the indefinite determiner *a* and ii) footing together *pint* and the reduced form of the preposition *of*.
  - In other words, this kind of prosodic phrasing is captured in p-structure (Bögel 2015, 2021), and simply arises from the fact that the relevant functional words (in this case, *a* and *of*) have /ə/ allomorphs.
- Therefore, the clitic$_b$ variety in fact does not have a DEP feature in v-structure at all, because its surface dependence is no more lexically conditioned than the surface dependence of *drink* or *pint*.
- Thus, the v-structure template for clitic$_b$ is identical to the one for free forms in (36) above.
- Next we turn to syntactic clitics, which we arbitrarily call *clitic$_c$*.
- Here we do not make reference to 'special clitics.'

  - We avoid this term simply because it tends to mean somewhat different things in different circles, although definitions overlap (see, for example, Spencer and Luis 2012).
  - Note that it is not our intent to treat *syntactic clitic* and *special clitic* as equivalent terms.
  - We expect a full theory of special clitics to deploy many of our morphosyntactic categories, including affixes and free forms.
- We define syntactic clitics as those elements that are associated with a clitic-specific syntactic category, Cl, in the c-structure (Bresnan et al. 2016b: 144–145, Arregi and Nevins 2018).
- This differentiates syntactic clitics from simple clitics (clitic_a) above.
- Indeed, elements of category Cl can be free-standing, affixal, or simple clitics/leaners, depending on their v-structure properties.
  - For example, this is how we would treat Romance object clitics (à la Arregi and Nevins 2018):[9,10]

(41) me lo d-a-n. Spanish
1.SG 3.SG.MASC give-TV-PL
'They give it to me.'

(42) $\text{Cl} \rightarrow \underset{(\uparrow \text{GF}) = \downarrow}{\text{Cl}} \quad \underset{(\uparrow \text{GF}) = \downarrow}{\text{Cl}}$

(43)



  - This is also how we have treated certain Ojibwe agreement clitics (Melchin et al. 2020b).
- Recall that leaners (clitic_a) arise from adding the feature DEP.
- Further specifying the v-structure by adding the feature [HOST [IDENT +]] yields the representation for an *affix*.
- Affixes arise from the combination of some DEP value and [IDENT +].
- These exponents have the following features, but note that the only possible value for IDENT is +.

(44)
$$\begin{bmatrix} \text{PHONREP} & \ldots \\ \text{PFRAME} & \ldots \\ \text{DEP} & \ldots \\ \text{HOST} & \begin{bmatrix} \text{IDENT} & + \end{bmatrix} \end{bmatrix}$$

- In sum, leaners (clitic_a) add the feature DEP to the features for free forms and affixes add the feature [HOST [IDENT +]] to the features for leaners, yielding a strict subsumption ordering:
free forms $\sqsubset$ leaners $\sqsubset$ affixes.
- The use of DEP and [IDENT +] in classifying forms yields a factorial typology of major morphological kinds, as shown in Table 1.

[9] Example (41) is declarative. In the imperative, *den=me=lo*, the clitics appear on the right side of the verb, rather than on the left as in (41), but the clitic constituent retains its order.

[10] The gloss TV stands for "theme vowel."

- Note that (• FEAT) and ¬(• FEAT) are standard LFG notation for indicating respectively the obligatory presence or absence of feature FEAT in the structure designated by •.
- Notice that in this factorial typology there is a possible combination of features that we have not considered above.
  - The combination in question — which is the presence of [IDENT +] in the absence of DEP — is shown in the bottom left cell of Table 1.
  - The occupants of this cell would be elements that care about their HOST and locality with respect to their host, but are not phonologically dependent.
  - The details of what it means to be hosted without being phonologically dependent we leave for future work, but we anticipate that certain particles and preposition might yield to this sort of analysis.

	[• IDENT +]	¬[• IDENT +]
[• DEP]	*affix*	*clitic*$_a$ *(leaner/simple clitic)*
¬[• DEP]	*some particles* *some prepositions*	*free form* *clitic*$_b$ *(phonological clitic)* *clitic*$_c$ *(syntactic clitic)*

Table 1: A factorial typology of major morphological kinds

7 An example: *-en*

- Let's now turn to a fully worked out example, an analysis of the English deadjectivizing verbalizer affix *-en*, repeated in (45) with the logical types fully indicated on the meaning constructor that is the sole member of its fugui.

(45) $\langle\, [\mathrm{v_a}], \{\, \lambda P_{et}\lambda v.\mathbf{cause}_{vt,vt}(\mathbf{become}_{et,vt}(P))(v) : \rangle \quad \stackrel{\nu}{\rightarrow}$

$$\begin{bmatrix} \text{PHONREP} & \text{/ən/} \\ \text{PFRAME} & ((\ \ )(\cdot)_\sigma)_{ft} \\ \text{PDOMAIN} & (\ (v\ \text{PFRAME})\ )_\omega \\ \text{DEP} & \text{LT} \\ \text{CLASS} & \text{WEAK} \\ \text{HOST} & \begin{bmatrix} \text{IDENT} & + \\ \text{PHONREP} & \text{/}\ldots([\mathbf{son}])[\mathbf{obs}]\text{/} \\ \text{PFRAME} & (\ \rho(v\ \text{HOST})\ )_\sigma \end{bmatrix} \end{bmatrix}$$

- This *-en* suffix occurs in words such as *blacken*, *quicken*, and *soften*.

- The suffix is simultaneously both very productive and quite restricted.
- It has many prosodic, phonological, and morphosyntactic restrictions on its host.
- It also places morphosemantic restrictions on the result of combination with its host, such that the host must be a property of events or entities.[11]
- However, if the full set of constraints is satisfied, the affix arises productively.
- We base the morphophonological and morphosyntactic constraints represented in the v-structure on the analysis of Halle (1973).
  - The host is monosyllabic and ends in an obstruent (optionally preceded by a sonorant).
  - According to Halle, this is a well-formedness condition on the output, which is why *soften* and *hasten* are allowed (the /t/ is deleted in these contexts).

- Let's next go through each of the affix's properties and how we capture them:

1. With respect to its prosody and morphophonology, this affix is consistently pronounced as a syllable with a reduced vowel and an alveolar nasal coda.
   - Therefore, its PHONREP value is /ən/.
   - The affix is a syllable that is the last in its foot.
   - Therefore its PFRAME value is $(\ldots(\,\cdot\,)_{\sigma})_{ft}$. The affix's form is subject to local word-level phonotactics. Therefore, its PDOMAIN value is $(\,\cdot\,)_{\omega}$.
2. With respect to dependency, the affix is a suffix, which means it is dependent to its left. Therefore, its DEP value is LT (short for {LT}).
3. We have also included the feature CLASS in the v-structure, even though it is probably not the case that CLASS is relevant to this affix.
   - Contemporary English probably does not synchronically have CLASS features; rather, it simply has regular verbs and irregular verbs.
   - However, for illustrative purposes, we can use CLASS, as might have been the case in the history of English, to capture the strong/weak distinction in verbs.
   - In this case, the resulting verb is a weak verb (in the Germanic sense); e.g. it is inflected with *-ed* in the past participle, unlike strong verbs like *take*, which is inflected in the past participle with the affix *-en*.
   - Again, just for the purpose of illustration, we identify two classes in English, *weak* and *strong*. Therefore, the value of CLASS is WEAK.
4. Conditions on the host:
   (a) The affix 'lowers' to the head of the complement of the affix.
   - Therefore, it contains the feature [HOST [IDENT +]].
   - As discussed above, this affix is subject to some kind of *morphological merger* operation in standard DM, such as head movement or lowering, because it is syntactically generated on the left of its host (given headedness in English), but appears on its right.
   - We capture this directly, rather than derivationally, through the combination of [HOST [IDENT +]], which requires the affix to attach to its host, and [DEP LT], which requires it to appear to the right of its host (i.e., the host must be on its left).

[11] This is captured through standard semantic typing; see Figure 2 below for details.

(b) The output form of the host must be no longer than one syllable. Therefore, the value of HOST PFRAME is $(\ldots)_{\sigma}$.

(c) The host must also end in an obstruent, optionally preceded by a sonorant (per Halle 1973).[12]

- For example, *soften* is legal despite a seemingly illegal base, because the final /t/ in the base is not present in the output [sɑfən].
- Furthermore, this restriction is a morphophonological constraint on the host and not a general phonological rule in English, because unaffixed forms with similar phonology are legal (e.g., **dryen* but *lion*, **dimmen* but *women*).
- Therefore, the value of HOST PHONREP is /…([**son**])[**obs**]/.

5. The affix is a deadjectivizing verbalizer.

- As is common in Distributed Morphology, we assume multiple subvarieties of categories, such as subvarieties of little v (for example, this is how we would capture theme vowel selection in Spanish).
- The fact that *-en* is deadjectivizing is a consequence of c-structural head adjunction of little a to the particular little v that *-en* is the exponent of.
- The use of adjunction allows the selectional history to be transmitted through the c-structure:[13]

(46) $\mathrm{v_a} \rightarrow \underset{\uparrow=\downarrow}{\mathrm{v_a}} \quad \underset{\uparrow=\downarrow}{\mathrm{a}}$

7.1 Formal analysis

- As is standard in LFG frameworks, $\mathrm{L_R FG}$ assumes that the ν-correspondence is defined and constrained by a description, which we can call a v-description.

- Thus, the exponent v-structure for *-en* can be described as follows, using $\bullet$ to represent "this v-structure" and $\cdot$ to represent "the p-structure correspondent of this v-structure," i.e. $\rho(\bullet)$:

(47)
$(\bullet\ \textsc{phonrep}) = /\text{ən}/$
$(\bullet\ \textsc{pframe}) = ((\quad)(\ \cdot\ )_{\sigma})_{ft}$
$(\bullet\ \textsc{pdomain}) = (\ \rho(\bullet\ \textsc{pframe})\ )_{\omega}$
$(\bullet\ \textsc{dependence}) = \textsc{lt}$
$(\bullet\ \textsc{class}) = \textsc{weak}$
$(\bullet\ \textsc{host ident}) = +$
$(\bullet\ \textsc{host phonrep}) =_c /\ldots([\textbf{son}])[\textbf{obs}]/$
$(\bullet\ \textsc{host pframe}) =_c (\ \rho(\bullet\ \textsc{host})\ )_{\sigma}$

- We can capture the general capacity to specify HOST content through this template:[14]

(48) $\textsc{host}(\mathrm{X,PR,PF,D,C}) :=$
$\mathrm{X} = + \Rightarrow (\bullet\ \textsc{host identity}) = +$
$\mathrm{PR} \neq \textbf{Id} \Rightarrow (\bullet\ \textsc{host phonrep}) =_c \mathrm{PR}$
$\mathrm{PF} \neq \textbf{Id} \Rightarrow (\bullet\ \textsc{host pframe}) =_c \mathrm{PF}$
$\mathrm{D} \neq \textbf{Id} \Rightarrow (\bullet\ \textsc{host dep}) =_c \mathrm{D}$
$\mathrm{C} \neq \textbf{Id} \Rightarrow (\bullet\ \textsc{host class}) =_c \mathrm{C}$

[12] We are presenting an unadulterated version of Halle's (1973) theory, but we are aware of complications, such as the well-formedness of *crispen*, which we set aside here.

[13] This phrase-structural approach replaces the feature TYPE in the previous brief presentation of v-structure in **??**. This allows us to capture an attested transitive property of this kind of selection that TYPE failed to capture (Oleg Belyaev, p.c.; Belyaev 2023).

[14] Note that we take the element **Id** to be whatever the appropriate *identity* element is for the argument in question. That is, an underspecified argument to a template returns whatever element is appropriate to combine with the value type in question to yield no change to the value. In the case of v-structure values, **Id** is the empty v-structure, since this can be thought of as unifying with any v-structure α to yield α. In the case of string values, such as the values of PHONREP and PFRAME, **Id** is the empty string, since this concatenates with any string α to yield α. In the case of set values, such as the values of DEP and CLASS, **Id** is the empty set, since this unions with any set A to return A.

- With (48) in hand, we can rewrite (47) as:

(49) (● PHONREP) = /ən/ (● DEPENDENCE) = LT
(● PFRAME) = (( )(·)$_\sigma$)$_{ft}$ (● CLASS) = WEAK
(● PDOMAIN) = (ρ(● PFRAME))$_\omega$ @HOST(+, /. . . ([**son**])[**obs**]/, ( )$_\sigma$, __, __)

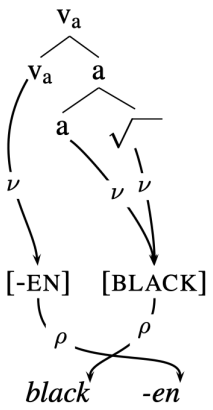
  - Any underspecified argument to a template is understood as an instance of the appropriate **Id** identity element (see footnote 14).

- Note that the re-ordering of the affix and host happens at p(rosodic)-structure, via the ρ correspondence function. The L_RFG c-structure with additional ρ-mapping indicated is sketched in (50).

- The less marked alternative is a zero-marked form. L_RFG does not employ zero affixes.

  - Zero-marking in L_RFG is a result of the fact that Pac-Man spanning is always available when overt exponence otherwise fails; see §5.6 above. Some examples are shown in (51) and (52).

(50)



(51)

Pac-Man spanning	*-en* affixation
to orange	**to redden**
to yellow	to blacken
to brown	* to brownen
* to red	* to orangen
* to black	* to yellowen

(52) a. The maple leaves **yellow**ed, **redden**ed, and finally **brown**ed in the sun.
b. Alex **redden**ed the mushrooms with food dye and then **brown**ed them in a skillet.

  - Pac-Man spanning results in portmanteaus, whenever the HOST requirements of *-en* are not satisfied.

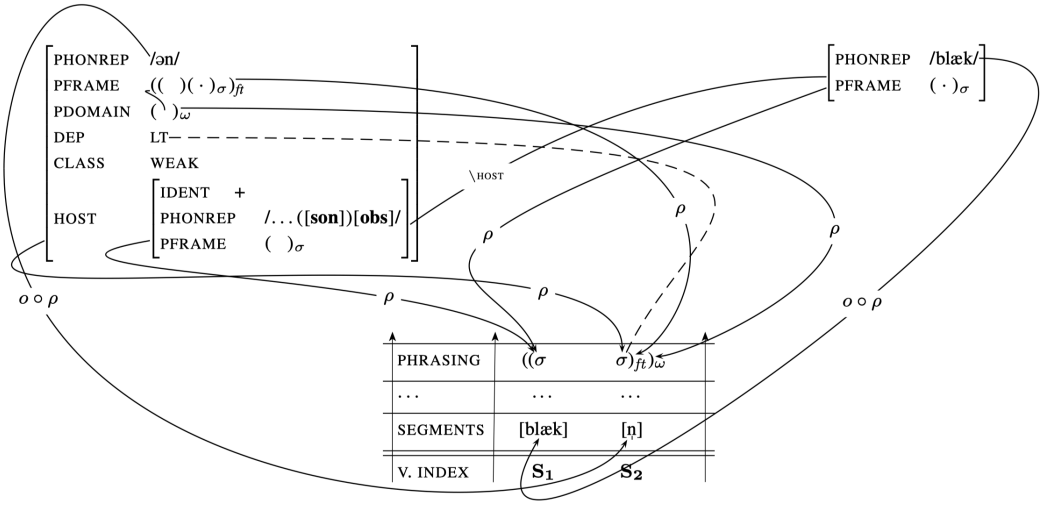
7.2 Mapping to Prosody

- The essence of our morphological analysis of *blacken* is captured by (45) and (50) above.
- However, now it is time to say more about the ρ-mapping, which we base on, e.g., Bögel (2015, 2021).
  - That is, given (45), how should the actual output of the ρ-mapping, the p-structure, be represented?
  - Similarly, how should the p-structure of the HOST, *black*, be represented?
- First, recall our principle (24), repeated here:

(24) HOST *Mapping*
For all v-structures v, v':
$(v \text{ HOST}) = v' \Rightarrow \rho(v') \in \rho(v \text{ PDOMAIN})$

- Given (24), the ρ-mapping must be as in (53), where the p-structure is represented as a *p-diagram* (see Bögel 2015, 2021).

(53)



- Note that the SEGMENTS in the p-diagram represent the output *p(honological)-string* in our correspondence architecture; see Figure 1.
  - These are *o*-mapped from p-structure, which is itself *ρ*-mapped from v-structure, so aspects of the phonological string can be mapped from v-structure using the composition of these mapping function, $o \circ \rho$.
  - This is why the mapping arrow from PHONREP in each v-structure is annotated $o \circ \rho$.

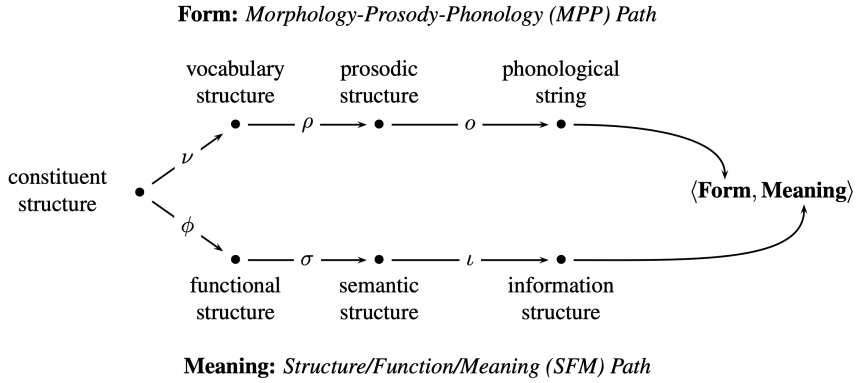


Figure 1: L_RFG Correspondence Architecture

- Figure 2 (page 28) shows what we call an *Everything Everywhere All at Once* (*EEAAO*) diagram for *blacken*; note that EEAAO is probably most easily pronounced as 'dubEdubAO.'
- A EEAAO diagram simultaneously represents the c-structure, v-structure(s), p-structure, f-structure, s-structure, mappings, and Glue proof(s) for an expression.
- It is a strength of the fully constraint-based ethos of L_RFG that one can simultaneously represent multiple kinds of grammatical information and how the different kinds of information relate to each other.

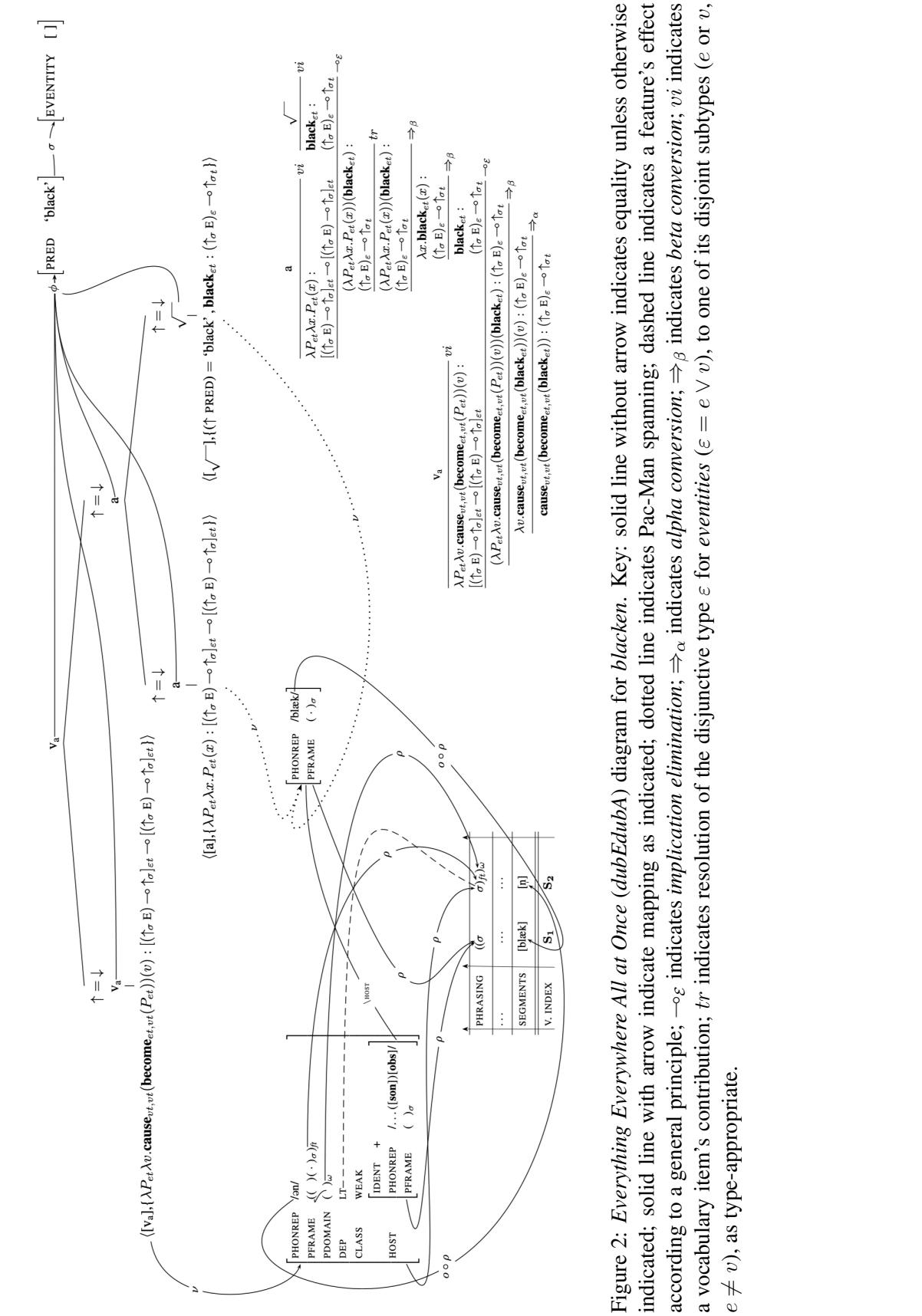


Figure 2: *Everything Everywhere All at Once* (*dubEdubA*) diagram for *blacken*. Key: solid line without arrow indicates equality unless otherwise indicated; solid line with arrow indicate mapping as indicated; dotted line indicates Pac-Man spanning; dashed line indicates a feature's effect according to a general principle; $-\circ_{\mathcal{E}}$ indicates *implication elimination*; $\Rightarrow_{\alpha}$ indicates *alpha conversion*; $\Rightarrow_{\beta}$ indicates *beta conversion*; vi indicates a vocabulary item's contribution; tr indicates resolution of the disjunctive type ε for *eventities* ($\varepsilon = e \vee v$), to one of its disjoint subtypes (e or v, $e \neq v$), as type-appropriate.

8 Conclusion

- I have shown how, in L_RFG, an *exponent* is a vocabulary structure that is ν-mapped from an *exponendum*.
- The exponendum is the lefthand side and the exponent is the righthand side of a *vocabulary item*, a listed mapping in the Vocabulary.
- The overall mapping thus looks like this, for any language L:

  (54) *exponendum* $\xrightarrow{\nu}$ *exponent* $\xrightarrow{o \circ \rho}$ *realization*

- Importantly, this demonstrates that *exponence* and *realization* are not conflated in L_RFG.
  - Exponence is about the mapping from c-structure to v-structure, as conditioned by the Vocabulary.
  - Realization is about the mapping from v-structure to prosody and phonology.
  - In other words, exponence concerns part of the morphology-prosody-phonology path in the architecture, namely the morphological interface between syntax and form that is represented by v-structure, whereas realization concerns the rest of the MPP path, the ρ-interfaces between morphology and prosody and the o-interface between prosody and phonology.
- Lastly, it is a strength of the L_RFG framework that we can represent the various structures and the interfaces between them simultaneously, as in the *everything everywhere all at once* diagram in Figure 2.

References

Abney, Steven Paul. 1987. The English Noun Phrase in its Sentential Aspect. Ph.D. thesis, Massachusetts Institute of Technology.

Anderson, Stephen R. 1969. West Scandanavian Vowel Systems and the Ordering of Phonological Rules. Ph.D. thesis, MIT.

—. 1982. Where's Morphology? *Linguistic Inquiry* 13(4): 571–612.

—. 1992. *A-Morphous Morphology*. Cambridge: Cambridge University Press.

Aronoff, Mark. 1976. *Word Formation in Generative Grammar*. Cambridge, MA: MIT Press.

—. 1994. *Morphology by Itself: Stems and Inflectional Classes*. Cambridge, MA: MIT Press.

Arregi, Karlos, and Andrew Nevins. 2018. Beware Occam's Syntactic Razor: Morphotactic Analysis and Spanish Mesoclisis. *Linguistic Inquiry* 49(4): 625–683.

Asudeh, Ash. 2012. *The Logic of Pronominal Resumption*. Oxford: Oxford University Press.

—. 2022. Glue Semantics. *Annual Review of Linguistics* 8: 321–341.

—. 2023. Glue Semantics. Forthcoming in Dalrymple (forthcoming).

Asudeh, Ash, Tina Bögel, and Daniel Siddiqi. 2023. Modelling Exponents. Forthcoming.

Asudeh, Ash, Paul B. Melchin, and Dan Siddiqi. 2021. Constraints All the Way Down: DM in a Representational Model of Grammar. Ms., Carleton University and University of Rochester. To appear in the *Proceedings of the 37th West Coast Conference on Formal Linguistics*.

Asudeh, Ash, and Daniel Siddiqi. 2023. Morphology in LFG. Forthcoming in Dalrymple (forthcoming).

Beard, Robert. 1995. *Lexeme-Morpheme Base Morphology*. Albany, NY: SUNY Press.

Beesley, Kenneth R., and Lauri Karttunen. 2003. *Finite State Morphology*. Stanford, CA: CSLI Publications.

Belyaev, Oleg. 2023. Ossetic verb morphology in L_RFG. Presented in the LFG 2023 workshop, *Revisiting Lexical Integrity*, July 24, University of Rochester.

Bermúdez-Otero, Ricardo, and Ana R. Luís. 2016. A View of the Morphome Debate. In Ana R. Luís and Ricardo Bermúdez-Otero, eds., *The Morphome Debate*, 309–340. Oxford: Oxford University Press.

Bobaljik, Jonathan David. 1994. *What Does Adjacency Do?*, vol. 22 of *MIT Working Papers in Linguistics*.

Bögel, Tina. 2015. The Syntax-Prosody Interface in Lexical Functional Grammar. Ph.D. thesis, University of Konstanz.

—. 2021. Function Words at the Interface: A Two-Tier Approach. *Languages* 6(4): 197.

Bresnan, Joan, Ash Asudeh, Ida Toivonen, and Stephen Wechsler. 2016a. *Lexical-Functional Syntax*. Chichester, UK: Wiley-Blackwell, 2nd edn.

—. 2016b. *Lexical-Functional Syntax*. Malden, MA: Wiley-Blackwell, 2nd edn.

Bresnan, Joan, and Sam A. Mchombo. 1995. The Lexical Integrity Principle: Evidence from Bantu. *Natural Language and Linguistic Theory* 13(2): 181–254.

Chomsky, Noam. 1970. Remarks on Nominalization. In Roderick A. Jacobs and Peter S. Rosenbaum, eds., *Readings in English Transformational Grammar*, 184–221. Waltham, MA: Ginn and Company.

Dalrymple, Mary, ed. 1999. *Semantics and Syntax in Lexical Functional Grammar: The Resource Logic Approach*. Cambridge, MA: MIT Press.

Dalrymple, Mary. 2015. Morphology in the LFG Architecture. In Miriam Butt and Tracy Holloway King, eds., *Proceedings of the LFG15 Conference*, 65–83. Stanford, CA: CSLI Publications.

Dalrymple, Mary, ed. forthcoming. *The Handbook of Lexical Functional Grammar*. Berlin: Language Science Press.

Dalrymple, Mary, John J. Lowe, and Louise Mycock. 2019. *The Oxford Reference Guide to Lexical Functional Grammar*. Oxford: Oxford University Press.

Di Sciullo, Anna Maria, and Edwin Williams. 1987. *On the Definition of Word*. Cambridge, MA: MIT Press.

Everdell, Michael, and Paul B. Melchin. 2021. Control the sentence, subordinate the pronoun: on the status of controlled versus non-controlled complement clauses in Odam. In *Proceedings of the 39th West Coast Conference on Formal Linguistics*. Cascadilla Press. Forthcoming.

Everdell, Michael, Paul B. Melchin, Ash Asudeh, and Daniel Siddiqi. 2021. Beyond c-structure and f-structure: On the argument-adjunct distinction in Odam. In Miriam Butt, Jamie Y. Findlay, and Ida Toivonen, eds., *Proceedings of the LFG21 Conference*, 125–145. Stanford, CA: CSLI Publications.

Halle, Morris. 1973. Prolegomena to a Theory of Word-Formation. *Linguistic Inquiry* 4(1): 3–16.

Halle, Morris, and Alec Marantz. 1993. Distributed morphology and the pieces of inflection. In Samuel Jay Keyser and Ken Hale, eds., *The View from Building 20*, 111–176. Cambridge, MA: MIT Press.

Harris, Alice C. 2002. *Endoclitics and the Origins of Udi Morphosyntax*. Oxford: Oxford University Press.

—. 2016. *Multiple Exponence*. Oxford: Oxford University Press.

Haugen, Jason D., and Daniel Siddiqi. 2016. Towards a restricted realizational theory: Multimorphemic monolistemicity, portmanteaux, and post-linearization spanning. In Siddiqi and Harley 2016, 343–386.

Kaplan, Ronald M., and Jürgen Wedekind. 1993. Restriction and Correspondence-Based Translation. In *Proceedings of the 6th Meeting of the EACL*. European Chapter of the Association of Computational Linguistics, University of Utrecht.

Kiparsky, Paul. 1973. Elsewhere in Phonology. In Stephen Anderson and Paul Kiparsky, eds., *A Festschrift for Morris Halle*. New York: Holt, Rinehart and Winston.

Lahiri, Aditi, and Frans Plank. 2009. *What Linguistics Universals can be True of*. Berlin: Springer.

Lapointe, Steven. 1980. A Theory of Grammatical Agreement. Ph.D. thesis, University of Massachusetts, Amherst.

Lieber, Rochelle. 1992. *Deconstructing Morphology*. Chicago: University of Chicago Press.

Luís, Ana, and Andrew Spencer. 2004. A paradigm function account of 'mesoclisis' in European Portuguese. *Yearbook of Morphology 2004* 177–228.

Marantz, Alec. 1984. *On the Nature of Grammatical Relations*. Cambridge, MA: MIT Press.

Marantz, Alec P. 1997. No Escape from Syntax: Don't Try Morphological Analysis in the Privacy of Your Own Lexicon. In Alexis Dimitriadis and Laura Siegel, eds., *University of Pennsylvania Working Papers in Linguistics*, vol. 4.2, 201–226. University of Pennsylvania.

Matthews, Peter H. 1972. *Inflectional Morphology: A Theoretical Study Based on Aspects of Latin Verb Conjugation*. Cambridge: Cambridge University Press.

Melchin, Paul B., Ash Asudeh, and Dan Siddiqi. 2020a. Ojibwe Agreement in a Representational, Morpheme-based Framework. In Angelica Hernández and M. Emma Butterworth, eds., *Proceedings of the 2020 Canadian Linguistic Association*. Canadian Linguistic Association.

—. 2020b. Ojibwe Agreement in Lexical-Realizational Functional Grammar. In Miriam Butt and Ida Toivonen, eds., *Proceedings of the LFG20 Conference*, 268–288. Stanford, CA: CSLI Publications.

Merchant, Jason. 2015. How much context is enough? Two cases of span-conditioned stem allomorphy. *Linguistic Inquiry* 46(2): 273–303.

Orgun, Cemil Orhan, and Ronald L. Sprouse. 1999. From "MParse" to "Control": Deriving Ungrammaticality. *Phonology* 16(2): 191–224.

Ramchand, Gillian Catriona. 2008. *Verb Meaning and the Lexicon: A First Phase Syntax*. Cambridge: Cambridge University Press.

Roark, Brian, and Richard Sproat. 2007. *Computational Approaches to Morphology and Syntax*. Oxford: Oxford University Press.

Sadler, Louisa, and Rachel Nordlinger. 2006. Case Stacking in Realizational Morphology 44(3): 459–487.

Siddiqi, Daniel. 2021. On the Taxonomy of Root Suppletion. In *Proceedings of the 39th West Coast Conference on Formal Linguistics*. Cascadilla Press. Forthcoming.

Siddiqi, Daniel, and Heidi Harley, eds. 2016. *Morphological Metatheory*. Amsterdam: John Benjamins.

Simpson, Jane. 1983. Aspects of Warlpiri Morphology and Syntax. Ph.D. thesis, MIT.

Spencer, Andrew. 2013. *Lexical Relatedness: A Paradigm-Based Model*. Oxford: Oxford University Press.

Spencer, Andrew, and Ana Luis. 2012. *Clitics: An Introduction*. Cambridge: Cambridge University Press.

Starke, Michal. 2009. Nanosyntax: A short primer to a new approach to language. *Nordlyd* 36(1): 1–6.

Stump, Gregory T. 2001. *Inflectional Morphology: A Theory of Paradigm Structure*. Cambridge: Cambridge University Press.

—. 2016. *Inflectional Paradigms: Content and Form at the Syntax-Morphology Interface*. Cambridge: Cambridge University Press.

Svenonius, Peter. 2016. Spans and words. In Siddiqi and Harley 2016, 201–222.

Thomas, Amanda. 2021. Divergence and Mismatches in Portuguese Morphology and Syntax. Ph.D. thesis, University of Oxford.

Travis, Lisa. 1984. Parameters and Effects of Word Order Variation. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA.

Wescoat, Michael T. 2002. On Lexical Sharing. Ph.D. thesis, Stanford University.

—. 2005. English nonsyllabic auxiliary contractions: An analysis in LFG with lexical sharing. In Miriam Butt and Tracy Holloway King, eds., *Proceedings of the LFG05 Conference*, 469–486. Stanford, CA: CSLI.

—. 2007. Preposition-determiner contractions: An analysis in Optimality-Theoretic Lexical-Functional Grammar with lexical sharing. In Miriam Butt and Tracy Holloway King, eds., *Proceedings of the LFG07 Conference*, 439–459. Stanford, CA: CSLI.