WP for sign-based CGs: a focus on morphotactics

Noah Diewald

March 26, 2019

WP for sign-based CGs: a focus on morphotactics

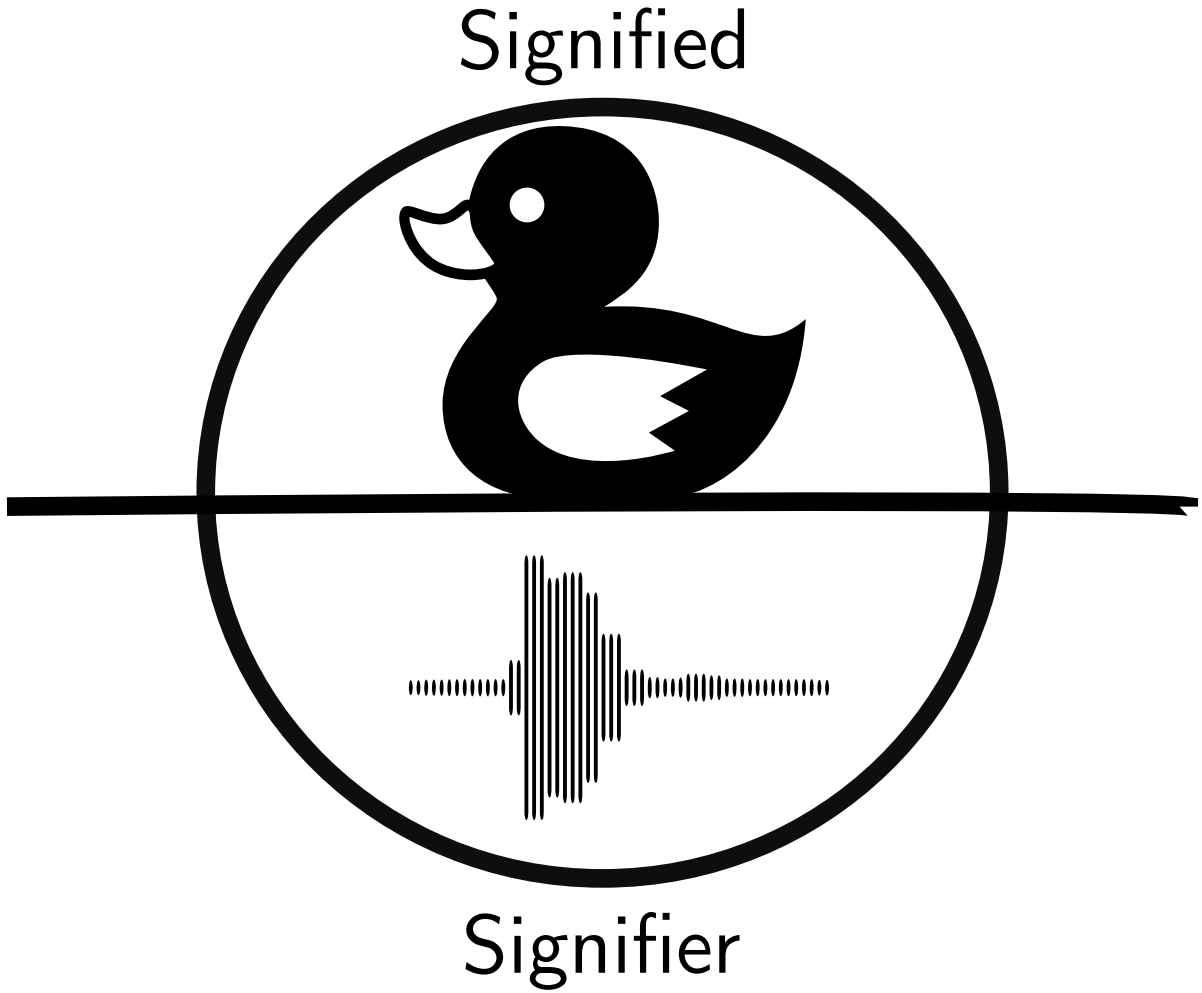
WP for sign-based CGs: a focus on morphotactics

Noah Diewald

March 26, 2019

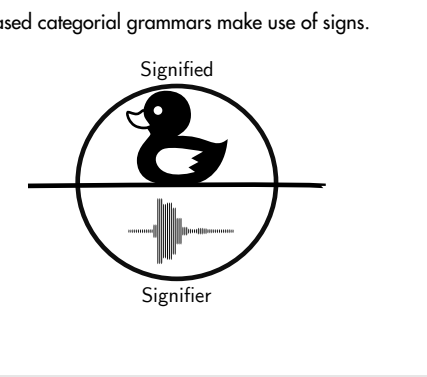
- This talk describes a aspects of a formal theory of separationist Word and Paradigm (WP)(Robins, 1959) morphology that interfaces well with categorial grammars (CG) (Bar-Hillel, 1953; Ajdukiewicz, 1935; Lambek, 1958) that are sign-based, in particular Linear Categorial Grammar (LCG) (Mihaliček and Pollard, 2012).
- The focus will be on morphotactics.
- The exposition to follow is intended as a simplified overview of particular themes that are central to this work. There is a lot that is being left out. This includes details of the formalism. Additionally, only some of the motivations are specified.

Sign-based categorial grammars make use of signs.



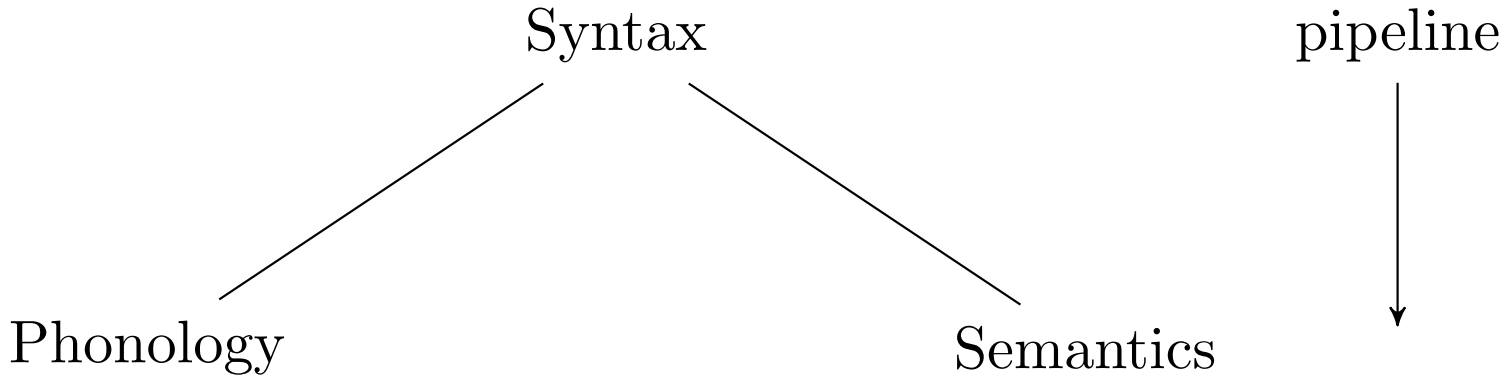
WP for sign-based CGs: a focus on morphotactics

Sign-based categorial grammars make use of signs.



- LCG is a sign-based categorial grammar.
- Sign-based Categorial Grammars, perhaps necessarily, combine two concepts.
  - They are Curry-esque, meaning there is some notion of phenogrammar-tectogrammar distinction (Curry, 1961).
  - They utilize an extension of the Saussurian sign concept (de Saussure, 2011).
- The Saussurian sign concept is essentially the pairing of form, the signifier, and meaning, the signified.

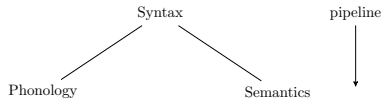
Saussurian signs do not capture "levels of grammar".



WP for sign-based CGs: a focus on morphotactics

Saussurian signs do not capture ”levels of grammar”.

Saussurian signs do not capture "levels of grammar".

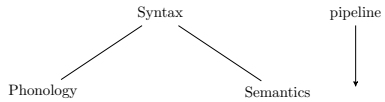


- As shown here, a sign is missing a number of properties associated with the modern concept of “levels of grammar.” This is a concept that could be said to have existed, in some sense, in the Saussurian concepts of syntagmatic and paradigmatic levels. The modern conception was evident in the American structuralist tradition (Harris, 1951). It was further developed by generative grammarians (Chomsky, 1957).
    - In the sign, there is no pipeline from meaning or syntax to form (or from phonetics upward as might be conceived by some structuralists.)
    - In the sign, there is no glue, or syntax mediating form and meaning.

WP for sign-based CGs: a focus on morphotactics

Saussurian signs do not capture ”levels of grammar”.

Saussurian signs do not capture "levels of grammar".

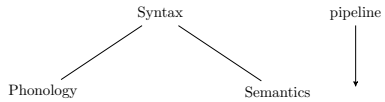


- In a sign-based categorial grammar, like LCG, we eschew the pipeline. Form and meaning are built up in parallel. Signs build signs. This is similar in many ways to HPSG (Pollard and Sag, 1994), which should be unsurprising given a closely shared heritage.

WP for sign-based CGs: a focus on morphotactics

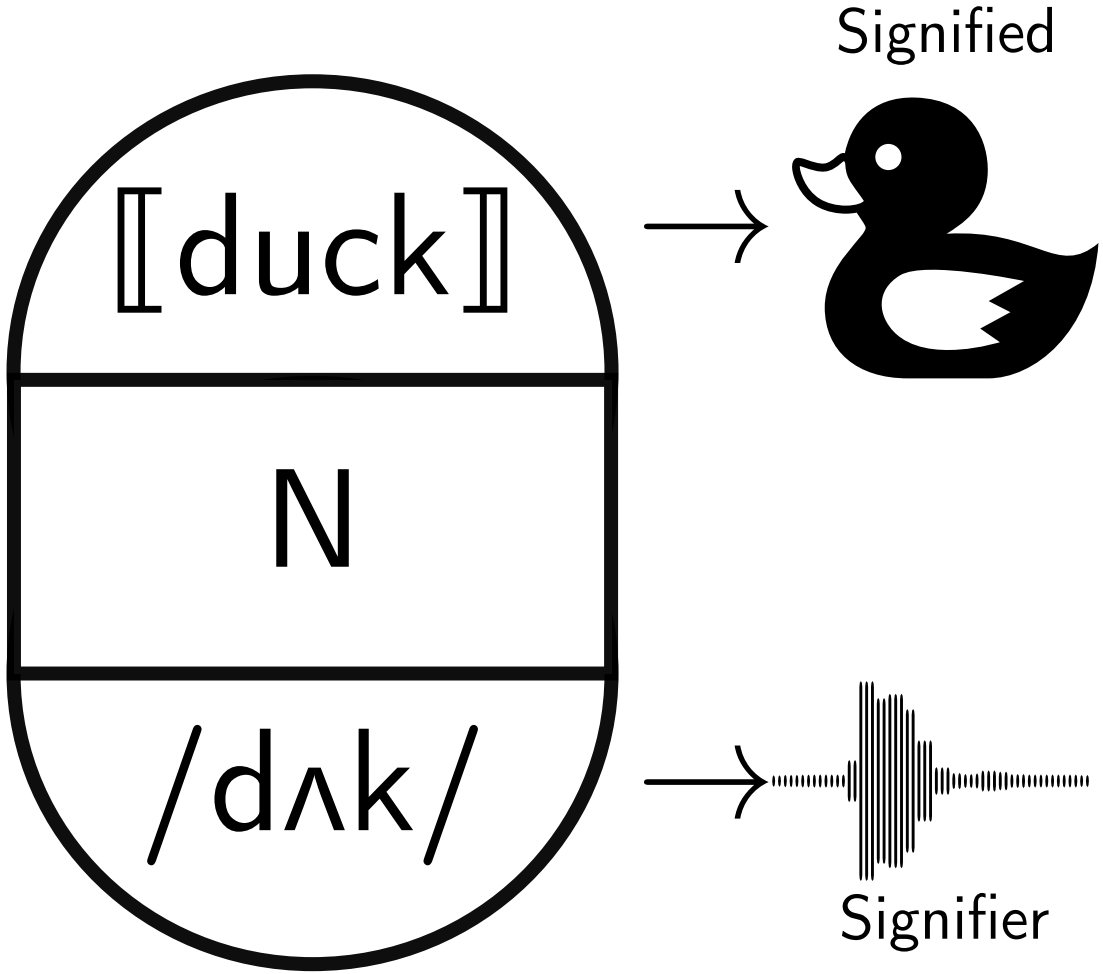
Saussurian signs do not capture ”levels of grammar”.

Saussurian signs do not capture "levels of grammar".



- To provide the glue (syntax) between meaning and form, we take advantage of a distinction made by Haskell Curry.
- Haskell Curry made a distinction between a concept of phenogrammar and tectogrammar.
- This is interpreted in LCG as a separation of combinatorics, handled by the tectogrammatical component, from word order, handled by the phenogrammatical component.
- By conflating the serialization concept shared by phenogrammar and phonology, and looking at phonology as ultimately being about form, we can augment our concept of sign with an interface to phonology.
- The tectogrammar serves the role of providing syntactic categories and specifying what may or may not combine.

LCG grammatical signs include a tectogrammar and a more indirect relationship to concept and form.



An LCG sign is a triple of grammatical information

$$\langle \underbrace{\text{dʌk}}_{\text{Pheno}}, \underbrace{\mathsf{N}}_{\text{Tecto}}, \underbrace{\textit{duck}}_{\text{Semantics}} \rangle$$

WP for sign-based CGs: a focus on morphotactics

An LCG sign is a triple of grammatical information

An LCG sign is a triple of grammatical information

$$\langle \underbrace{\text{dʌk}}_{\text{Pheno}}, \underbrace{\mathsf{N}}_{\text{Tecto}}, \underbrace{\mathit{duck}}_{\text{Semantics}} \rangle$$

- Here is the DUCK example translated to LCG notation.
- I shorten the names of the components of the signs here, they are phenogrammar, tectogrammar and semantics, respectively.
- Formally, these shortened component names will be the names of the types of the sign, i.e. $\mathit{Pheno} \times \mathit{Tecto} \times \mathit{Semantics}$.
- In general, I will use common orthography in the pheno.

A more relevant example is walk

$$\langle \underbrace{\lambda \mathrm{s.s} \bullet \mathrm{walks}}_{\text{Pheno}}, \underbrace{\mathsf{Nom} \multimap \mathsf{Fin}}_{\text{Tecto}}, \underbrace{\mathit{walk}}_{\text{Semantics}} \rangle$$

A more relevant example is WALK

A more relevant example is walk

$\langle \underbrace{\lambda s.s \bullet \text{walks}}_{\text{Pheno}}, \underbrace{\text{Nom} \multimap \text{Fin}}_{\text{Tecto}}, \underbrace{\mathit{walk}}_{\text{Semantics}} \rangle$

- The dot in the pheno is a concatenation operator. The function here takes some pheno term and concatenates it before walks.
- The lollipop symbol ($\multimap$) is used to indicated that the combinatorics says that a Nom (nominative NP) is required to make a Fin (finite clause).
- The semantics, which will be simplified in much of this work, is indicated with the predicate, *walk*.

The type of the sign triple.

$$\mathit{Pheno} \times \mathit{Tecto} \times \mathit{Semantics}$$

WP for sign-based CGs: a focus on morphotactics

The type of the sign triple.

The type of the sign triple.

Pheno × *Tecto* × *Semantics*

- The type of signs in given here.
- We can safely ignore some of the details but what we can see is that it is very broad.
- It says that anything that is a phenogrammatical term can combine with anything that is a tectogrammatical term and any meaning and be a sign.
- This type covers not only grammatical signs but ungrammatical and non-sense signs.

The type itself does not capture grammaticality.

$$\langle \underbrace{\text{dʌk} \bullet \text{dʌk}}_{\text{Pheno}}, \underbrace{\text{Fin}}_{\text{Tecto}}, \underbrace{\textit{kick}}_{\text{Semantics}} \rangle$$

WP for sign-based CGs: a focus on morphotactics

The type itself does not capture grammaticality.

The type itself does not capture grammaticality.

$\langle \underbrace{\text{dʌk} \bullet \text{dʌk}}_{\text{Pheno}}, \underbrace{\mathsf{Fin}}_{\text{Tecto}}, \underbrace{\mathit{kick}}_{\text{Semantics}} \rangle$

- The point of LCG, in large part, is to pick out which of these signs are part of a grammar for a language.

LCG seeks to derive only grammatical signs.

$$\frac{\vdash \lambda.\mathrm{s} \bullet \mathrm{walks}; \mathsf{NP} \multimap \mathsf{Fin}; \mathit{walk} \qquad \vdash \mathrm{howard}; \mathsf{NP}; \mathit{h}}{\vdash \mathrm{howard} \bullet \mathrm{walks}; \mathsf{Fin}; \mathit{walk}(\mathit{h})}\ \mathrm{MP}$$

* The semicolons, rather than commas, are more common when using signs in proofs.

WP for sign-based CGs: a focus on morphotactics

LCG seeks to derive only grammatical signs.

LCG seeks to derive only grammatical signs.

$$\frac{\vdash \lambda s \bullet \text{walks}; \mathsf{NP} \multimap \mathsf{Fin}; \mathit{walk} \qquad \vdash \text{howard}; \mathsf{NP}; h}{\vdash \text{howard} \bullet \text{walks}; \mathsf{Fin}; \mathit{walk}(h)}\ \text{MP}$$

* The semicolons, rather than commas, are more common when using signs in proofs.

- What you see in this slide is a proof about signs.
- We see the sign for WALK being combined, via modus ponens, with the sign for HOWARD.
- Note that the result is a new sign.
- There is no notion that any component of the sign is prior to another as would be the case in a pipeline.

Lexical entries are assumed, rather than derived.

$$\vdash \langle \lambda \mathrm{s.s} \bullet \mathrm{walks}, \mathsf{Nom} \multimap \mathsf{Fin}, \mathit{walk} \rangle$$

WP for sign-based CGs: a focus on morphotactics

Lexical entries are assumed, rather than derived.

- Not all useful signs are derived in LCG.
- Some are specified as axioms.
- Some such axioms are the *lexical entries* of the theory.

Lexical entries are assumed, rather than derived.

$\vdash \langle \lambda s.s \bullet \text{walks}, \mathsf{Nom} \multimap \mathsf{Fin}, \mathit{walk} \rangle$

There are a fixed number of such axioms for a lexeme, forming a paradigm.

Designation	**Lexical Entry**
present tense	$\langle \lambda s.s \bullet \text{jump}, \mathsf{Nom}_{\neg 3s} \multimap \mathsf{Fin}, \mathit{jump} \rangle$
third person	$\langle \lambda s.s \bullet \text{jumps}, \mathsf{Nom} \multimap \mathsf{Fin}, \mathit{jump} \rangle$
past tense	$\langle \lambda s.s \bullet \text{jumped}, \mathsf{Nom} \multimap \mathsf{Fin}, \mathit{past}(\mathit{jump}) \rangle$
base	$\langle \text{jump}, \mathsf{PRO} \multimap \mathsf{Bse}, \mathit{jump} \rangle$
past participle	$\langle \text{jumped}, \mathsf{PRO} \multimap \mathsf{Prp}, \mathit{jump} \rangle$
present participle	$\langle \text{jumping}, \mathsf{PRO} \multimap \mathsf{Psp}, \mathit{jump} \rangle$

WP for sign-based CGs: a focus on morphotactics

There are a fixed number of such axioms for a lexeme, forming a paradigm.

There are a fixed number of such axioms for a lexeme, forming a paradigm.

Designation	Lexical Entry
present tense	$\langle \lambda s.s \bullet \text{jump}, \text{Nom}_{\text{-3s}} \multimap \text{Fin}, \textit{jump} \rangle$
third person	$\langle \lambda s.s \bullet \text{jumps}, \text{Nom} \multimap \text{Fin}, \textit{jump} \rangle$
past tense	$\langle \lambda s.s \bullet \text{jumped}, \text{Nom} \multimap \text{Fin}, \textit{past}(\textit{jump}) \rangle$
base	$\langle \text{jump}, \text{PRO} \multimap \text{Bse}, \textit{jump} \rangle$
past participle	$\langle \text{jumped}, \text{PRO} \multimap \text{Prp}, \textit{jump} \rangle$
present participle	$\langle \text{jumping}, \text{PRO} \multimap \text{Psp}, \textit{jump} \rangle$

- There are a fixed number of lexical entries for a lexeme.
- In LCG, a typical English intransitive verb will need about 6 lexical entries.
- A lexical entry is a generalization of the distributional possibilities of a lexeme's sign-based manifestation in the grammar.
- The lexical entries for a particular lexeme are its *lexical entry paradigm*.

There are patterns across lexical entry paradigms.

$\langle \lambda s.s \bullet \text{jump}, \mathsf{Nom}_{\neg 3s} \multimap \mathsf{Fin}, \mathit{jump} \rangle$

$\langle \lambda s.s \bullet \text{jumps}, \mathsf{Nom} \multimap \mathsf{Fin}, \mathit{jump} \rangle$

$\langle \lambda s.s \bullet \text{jumped}, \mathsf{Nom} \multimap \mathsf{Fin}, \mathit{past}(\mathit{jump}) \rangle$

$\langle \text{jump}, \mathsf{PRO} \multimap \mathsf{Bse}, \mathit{jump} \rangle$

$\langle \text{jumped}, \mathsf{PRO} \multimap \mathsf{Prp}, \mathit{jump} \rangle$

$\langle \text{jumping}, \mathsf{PRO} \multimap \mathsf{Psp}, \mathit{jump} \rangle$

$\langle \lambda s.s \bullet \text{walk}, \mathsf{Nom}_{\neg 3s} \multimap \mathsf{Fin}, \mathit{walk} \rangle$

$\langle \lambda s.s \bullet \text{walks}, \mathsf{Nom} \multimap \mathsf{Fin}, \mathit{walk} \rangle$

$\langle \lambda s.s \bullet \text{walked}, \mathsf{Nom} \multimap \mathsf{Fin}, \mathit{past}(\mathit{walk}) \rangle$

$\langle \text{walk}, \mathsf{PRO} \multimap \mathsf{Bse}, \mathit{walk} \rangle$

$\langle \text{walked}, \mathsf{PRO} \multimap \mathsf{Prp}, \mathit{walk} \rangle$

$\langle \text{walking}, \mathsf{PRO} \multimap \mathsf{Psp}, \mathit{walk} \rangle$

WP for sign-based CGs: a focus on morphotactics

There are patterns across lexical entry paradigms.

There are patterns across lexical entry paradigms.

⟨λs.s • jump, Nom$_{\neg 3s}$ ⊸ Fin, *jump*⟩
⟨λs.s • jumps, Nom ⊸ Fin, *jump*⟩
⟨λs.s • jumped, Nom ⊸ Fin, *past*(*jump*)⟩
⟨jump, PRO ⊸ Bse, *jump*⟩
⟨jumped, PRO ⊸ Prp, *jump*⟩
⟨jumping, PRO ⊸ Psp, *jump*⟩

⟨λs.s • walk, Nom$_{\neg 3s}$ ⊸ Fin, *walk*⟩
⟨λs.s • walks, Nom ⊸ Fin, *walk*⟩
⟨λs.s • walked, Nom ⊸ Fin, *past*(*walk*)⟩
⟨walk, PRO ⊸ Bse, *walk*⟩
⟨walked, PRO ⊸ Prp, *walk*⟩
⟨walking, PRO ⊸ Psp, *walk*⟩

- When we compare the paradigms of different lexemes, clear patterns associated with forms in the phenogrammar emerge.
- Signs with the same tectogrammatical and semantic components also follow a pattern.
- Only the stem and predicate are unique to the individual entries.
- We say that words that share patterns in their paradigms share an inflection class, though we'll talk very little about such classes today.
- The inflection class for English that is shared by both JUMP and WALK is called *weak* or colloquially *regular*.

WP for sign-based CGs: a focus on morphotactics

There are patterns across lexical entry paradigms.

There are patterns across lexical entry paradigms.

$\langle \lambda s.s \bullet \text{jump}, \text{Nom}_{\neg 3s} \multimap \text{Fin}, \textit{jump} \rangle$
$\langle \lambda s.s \bullet \text{jumps}, \text{Nom} \multimap \text{Fin}, \textit{jump} \rangle$
$\langle \lambda s.s \bullet \text{jumped}, \text{Nom} \multimap \text{Fin}, \textit{past}(\textit{jump}) \rangle$
$\langle \text{jump}, \text{PRO} \multimap \text{Bse}, \textit{jump} \rangle$
$\langle \text{jumped}, \text{PRO} \multimap \text{Prp}, \textit{jump} \rangle$
$\langle \text{jumping}, \text{PRO} \multimap \text{Psp}, \textit{jump} \rangle$

$\langle \lambda s.s \bullet \text{walk}, \text{Nom}_{\neg 3s} \multimap \text{Fin}, \textit{walk} \rangle$
$\langle \lambda s.s \bullet \text{walks}, \text{Nom} \multimap \text{Fin}, \textit{walk} \rangle$
$\langle \lambda s.s \bullet \text{walked}, \text{Nom} \multimap \text{Fin}, \textit{past}(\textit{walk}) \rangle$
$\langle \text{walk}, \text{PRO} \multimap \text{Bse}, \textit{walk} \rangle$
$\langle \text{walked}, \text{PRO} \multimap \text{Prp}, \textit{walk} \rangle$
$\langle \text{walking}, \text{PRO} \multimap \text{Psp}, \textit{walk} \rangle$

- WP morphology seeks to capture patterns such as these and generalize over them.
- The theory of WP morphology to be presented here, among other things, seeks to capture generalizations in the lexicon that are not captured in the signs themselves.
- There is work in HPSG-like formalisms that sees the morphology-syntax relationship similarly (Koenig and Jurafsky, 1994).

There are often repetitions of forms in paradigms.

$\langle \lambda s.s \bullet \text{jump}, \mathsf{Nom}_{\neg 3s} \multimap \mathsf{Fin}, \mathit{jump} \rangle$

$\langle \text{jump}, \mathsf{PRO} \multimap \mathsf{Bse}, \mathit{jump} \rangle$

$\langle \lambda s.s \bullet \text{jumped}, \mathsf{Nom} \multimap \mathsf{Fin}, \mathit{past}(\mathit{jump}) \rangle$

$\langle \text{jumped}, \mathsf{PRO} \multimap \mathsf{Prp}, \mathit{jump} \rangle$

WP for sign-based CGs: a focus on morphotactics

There are often repetitions of forms in paradigms.

There are often repetitions of forms in paradigms.

$\langle \lambda s.s \bullet \text{jump}, \text{Nom}_{-3s} \multimap \text{Fin}, \textit{jump} \rangle$
$\langle \text{jump}, \text{PRO} \multimap \text{Bse}, \textit{jump} \rangle$
$\langle \lambda s.s \bullet \text{jumped}, \text{Nom} \multimap \text{Fin}, \textit{past}(\textit{jump}) \rangle$
$\langle \text{jumped}, \text{PRO} \multimap \text{Prp}, \textit{jump} \rangle$

- Syncretism is the repetition of form within a lexical entry paradigm.
- Other theories, even if they differ in the nature of their lexical theory, recognize similar occurrences of syncretism.
- Syncretism is often used as an argument against associating affixes and other types of morphological processes with particular meanings.
- It shows that the same form can be associated with many entries.

It is not unusual to have too many forms in a paradigm.

$$\langle \lambda \mathrm{s.s} \bullet \mathrm{dived}, \mathsf{Nom} \multimap \mathsf{Fin}, \mathit{past}(\mathit{dive}) \rangle$$
$$\langle \lambda \mathrm{s.s} \bullet \mathrm{dove}, \mathsf{Nom} \multimap \mathsf{Fin}, \mathit{past}(\mathit{dive}) \rangle$$

WP for sign-based CGs: a focus on morphotactics

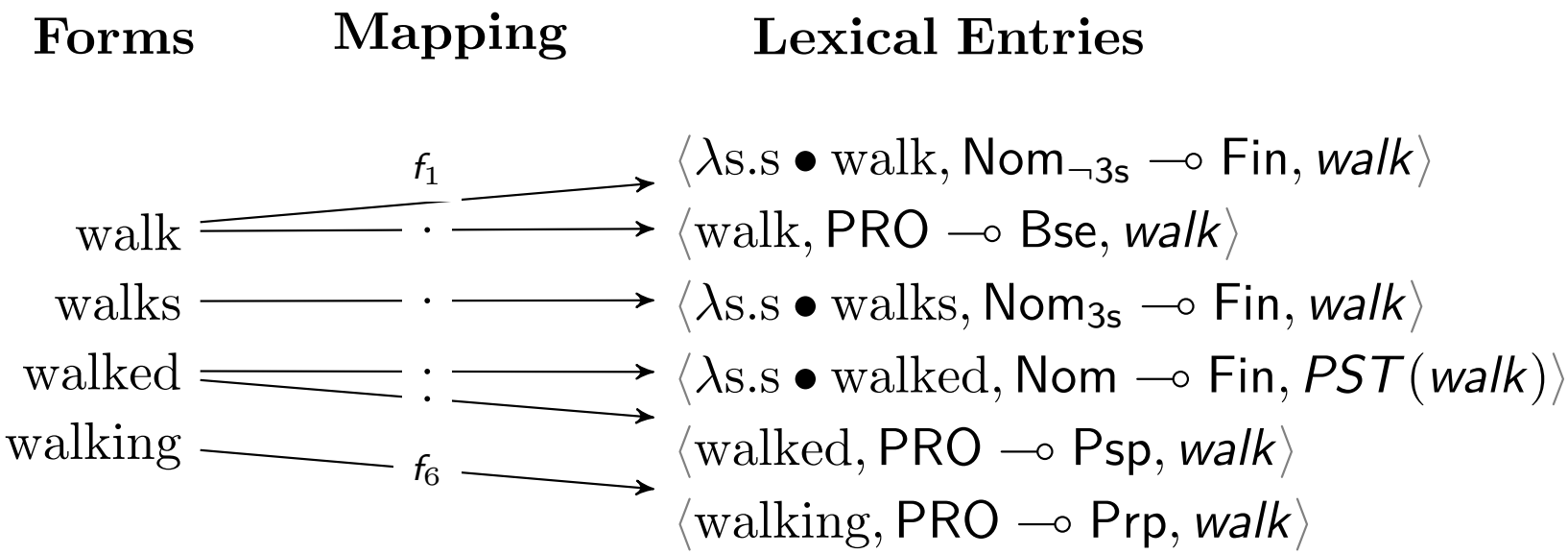
It is not unusual to have too many forms in a paradigm.

It is not unusual to have too many forms in a paradigm.

$\langle \lambda s.s \bullet \text{dived}, \text{Nom} \multimap \text{Fin}, \textit{past}(\textit{dive}) \rangle$
$\langle \lambda s.s \bullet \text{dove}, \text{Nom} \multimap \text{Fin}, \textit{past}(\textit{dive}) \rangle$

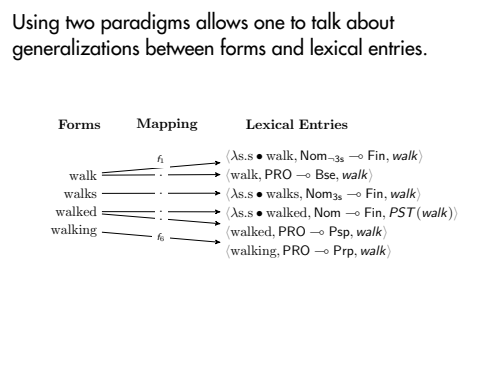
- The type of redundancy demonstrated in this slide is called over-abundance (Thornton, 2011).
- It is further evidence that there is some mismatch between forms and lexical entry paradigms.
- Syncretism showed that the same form can be associated with many entries.
- Overabundance shows that the same entry type can be associated with many forms.

Using two paradigms allows one to talk about generalizations between forms and lexical entries.



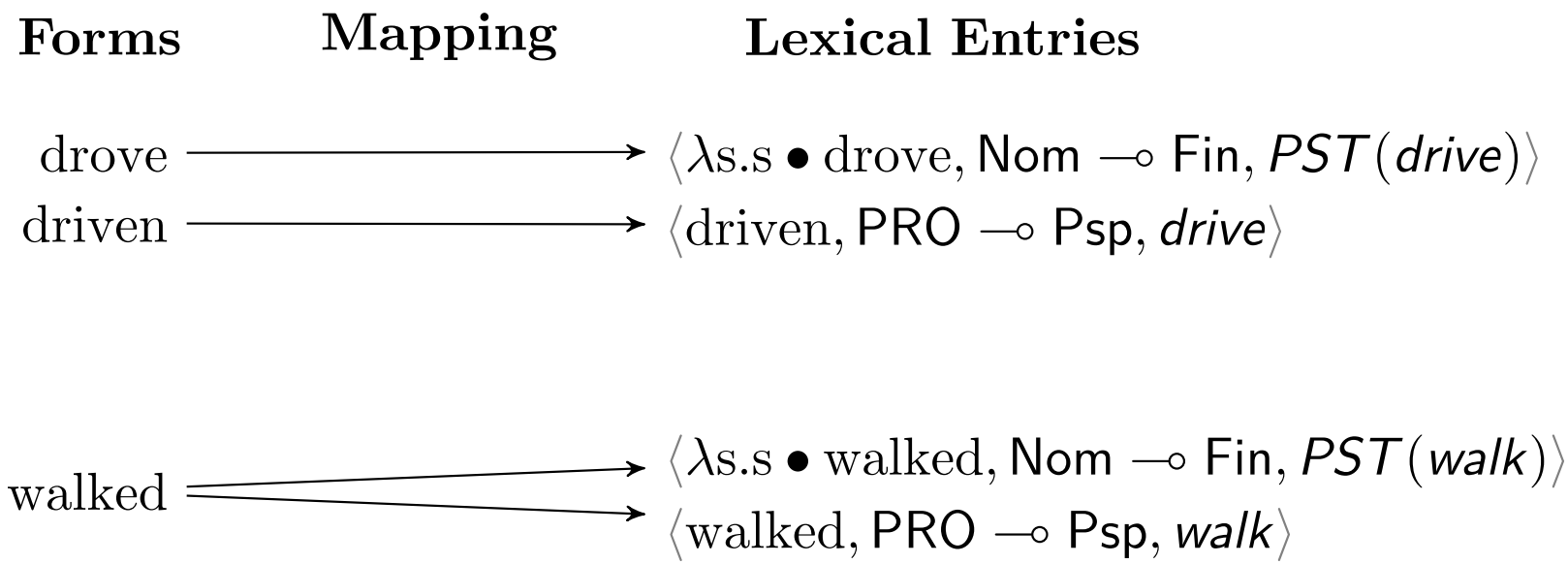
WP for sign-based CGs: a focus on morphotactics

Using two paradigms allows one to talk about generalizations between forms and lexical entries



- For this and other reasons, a tactic taken in some WP theories, such as Paradigm Linking Theory (Stump, 2016) is to propose two layers of paradigms.
- In the theory presented here, there is a *form paradigm* and a lexical entry paradigm.
- The form paradigm is determined by the inflected forms associated with a lexeme. There is no duplication of form in this paradigm.
- This arrangement serves nothing more extraordinary than capturing the fact that there appear to be a finite collection of forms associate with a lexeme and that there is a pattern to how these forms distribute in syntax.
- That pattern is captured by *paradigm mapping functions*, which are represented by arrows labeled f_1 to f_6 in the diagram.
- The very same mapping functions can be used to capture patterns between forms and lexical entries for many words.

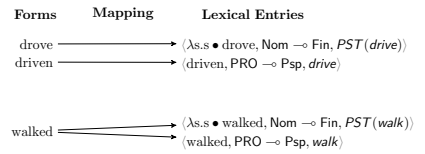
'Drive' and 'walk' exhibit different patterns with respect to the past and past participle.



WP for sign-based CGs: a focus on morphotactics

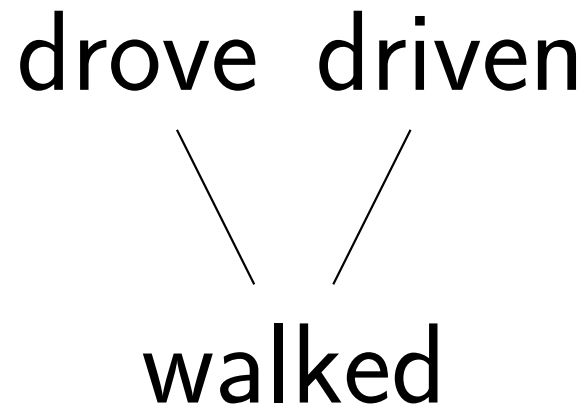
└'Drive' and 'walk' exhibit different patterns with respect to the past and past participle.

'Drive' and 'walk' exhibit different patterns with respect to the past and past participle.



- These paradigm mappings are not always obviously general.
- The pattern that 'drive' follows is not unique, with a number of verbs having different forms for the past and past participle.
- With respect to the weak paradigm, it seems a generalization is being missed.
- The weak past form is serving the role of both the drive-like participle and past forms.

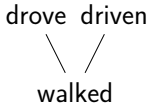
The weak past form serves the role of both past and past participle forms.



WP for sign-based CGs: a focus on morphotactics

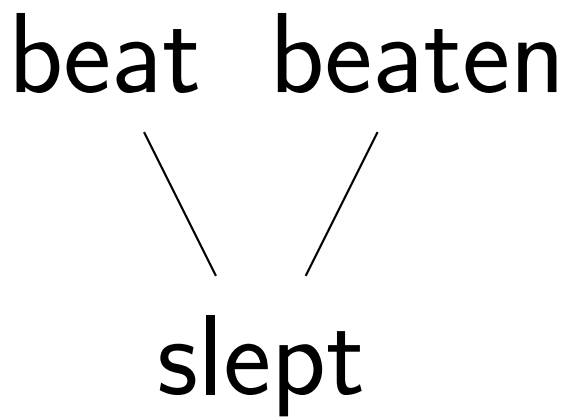
The weak past form serves the role of both past and past participle forms.

The weak past form serves the role of both past and past participle forms.



- The pattern schematized in this slide is seen throughout the lexicon.

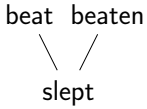
This same pattern follows for other verbs. Some are not even weak verbs.



WP for sign-based CGs: a focus on morphotactics

This same pattern follows for other verbs. Some are not even weak verbs.

This same pattern follows for other verbs. Some are not even weak verbs.

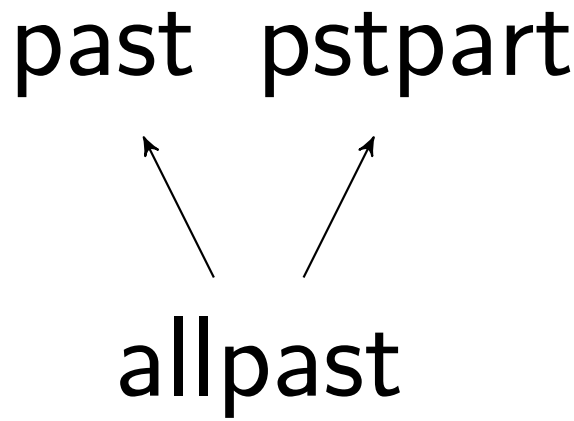


- What we're seeing in these slides is a relationship among categories, often called meta-syncretism (Williams, 1981; Williams, 1994; Bobaljik, 2002; Harley, 2008).
- We would like to provide is some designation for the categories.
- We would also like to find some means of formalizing the relationships between categories.

We can generalize this pattern and formalize the relationship.

$$\mathsf{allpast} \leq \mathsf{past}$$

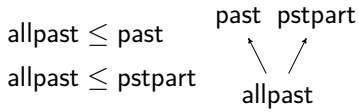
$$\mathsf{allpast} \leq \mathsf{pstpart}$$



WP for sign-based CGs: a focus on morphotactics

We can generalize this pattern and formalize the relationship.

We can generalize this pattern and formalize the relationship.



- I will name the category that is specific to the past tense forms *past*.
- I will name the category that is specific to the past participle forms *pstpart*.
- The category that is the union of the other two, I'll call *allpast*.
- Their relationship will be represented by a partial order, a reflexive, transitive and antisymmetric relation, shown as a less than or equal to symbol.
- In the diagram we represent this relationship as a directed acyclic graph, where an edge points toward the greater objects.

We tag every form paradigm entry with a category marker that we'll call a state.

$\langle \text{simple}, \text{walk} \rangle$

$\langle \text{third}, \text{walks} \rangle$

$\langle \text{allpast}, \text{walked} \rangle$

$\langle \text{prspart}, \text{walking} \rangle$

$\langle \text{simple}, \text{drive} \rangle$

$\langle \text{third}, \text{drives} \rangle$

$\langle \text{past}, \text{drove} \rangle$

$\langle \text{pstpart}, \text{driven} \rangle$

$\langle \text{prepart}, \text{driving} \rangle$

WP for sign-based CGs: a focus on morphotactics

We tag every form paradigm entry with a category marker that we'll call a state.

We tag every form paradigm entry with a category marker that we'll call a state.

⟨simple, walk⟩
⟨third, walks⟩
⟨allpast, walked⟩
⟨prspart, walking⟩

⟨simple, drive⟩
⟨third, drives⟩
⟨past, drove⟩
⟨pstpart, driven⟩
⟨prepart, driving⟩

- In this slide, I augment the form paradigm with these categories, which I will call *states* from here on out because of similarities that this system will have with finite state machines.
- Note, states are only names.
- They are not grammatical features or meanings.
- I might name them all 'state1', 'state2', etc. but that quickly gets confusing.

We use dependent types to create mappings that are both general and constrained.

$$\Sigma_{x:\mathit{form}_{\mathit{entry}}}(\mathit{state}(x) \leq \mathsf{past}) \rightarrow \mathit{lexical}_{\mathit{entry}}$$

WP for sign-based CGs: a focus on morphotactics

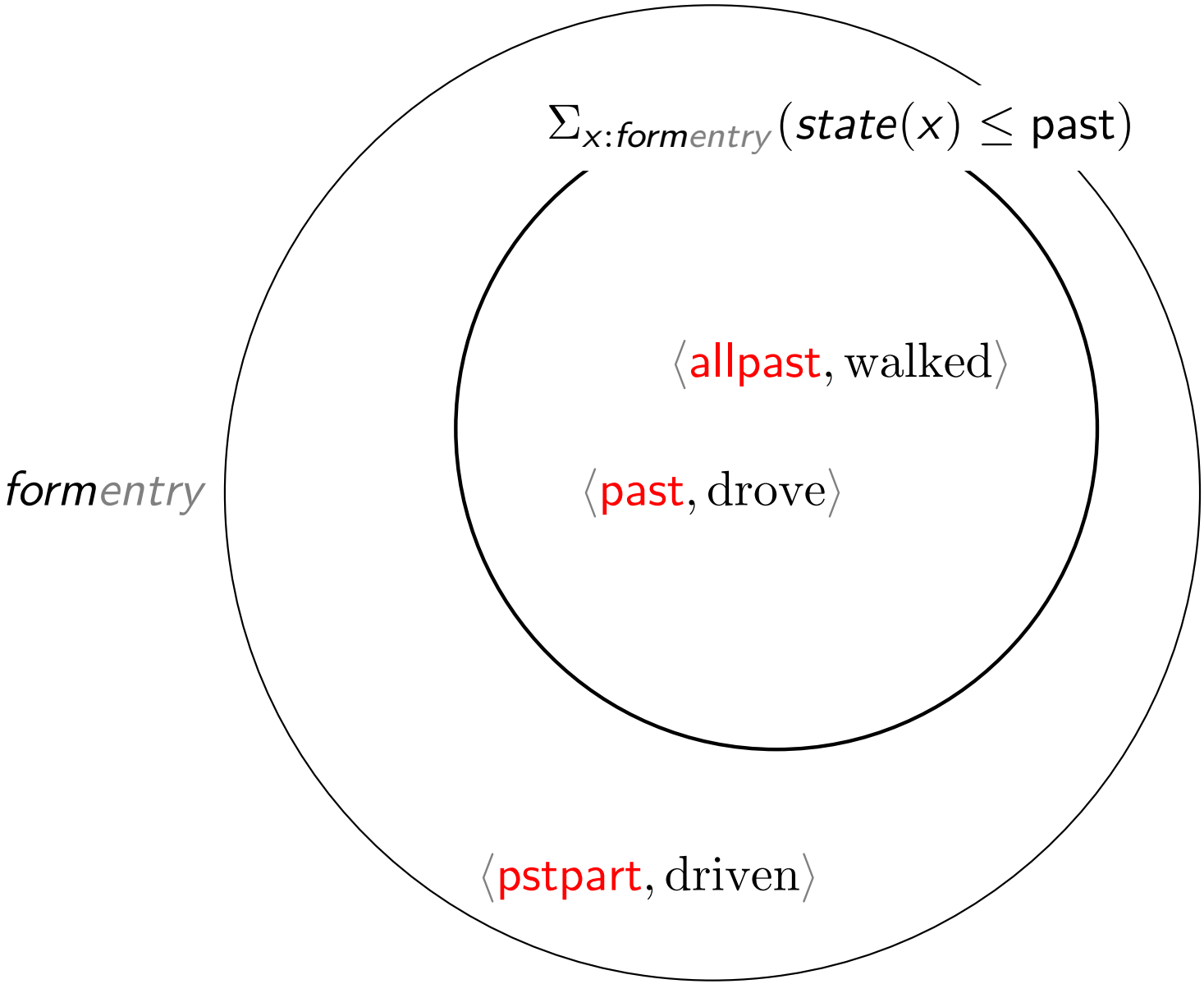
We use dependent types to create mappings that are both general and constrained.

We use dependent types to create mappings that are both general and constrained.

$\Sigma_{x:formentry}(state(x) \leq \text{past}) \rightarrow lexicalentry$

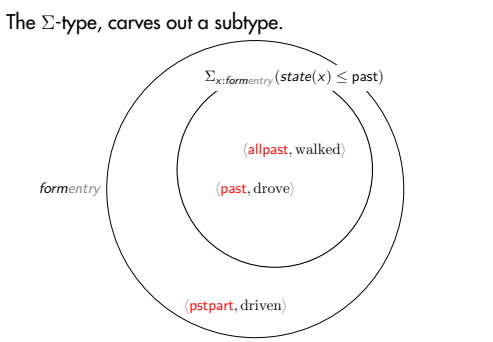
- In order to capture the relationships between states, I make use of a mechanism from dependent type theory (Martin-Löf, 1984).
- What you see in this slide is the type of a function.
- It is a dependent type because the type depends on the property of a term of type *formentry*.
- A term of type *formentry* is simply a pairing of a state and a form, as we saw in the previous slide.
- The type contains a predicate that makes a comparison.
- The $state(x)$ function retrieves the state information from the term of type *formentry*.
- This is then compared against the 'past' state.
- If the input state is less than or equal to 'past', the type given in the slide is correct for the term.
- The function corresponding to this type is a mapping function and will map this form to the lexical entry corresponding to a past tense verb.

The Σ-type, carves out a subtype.



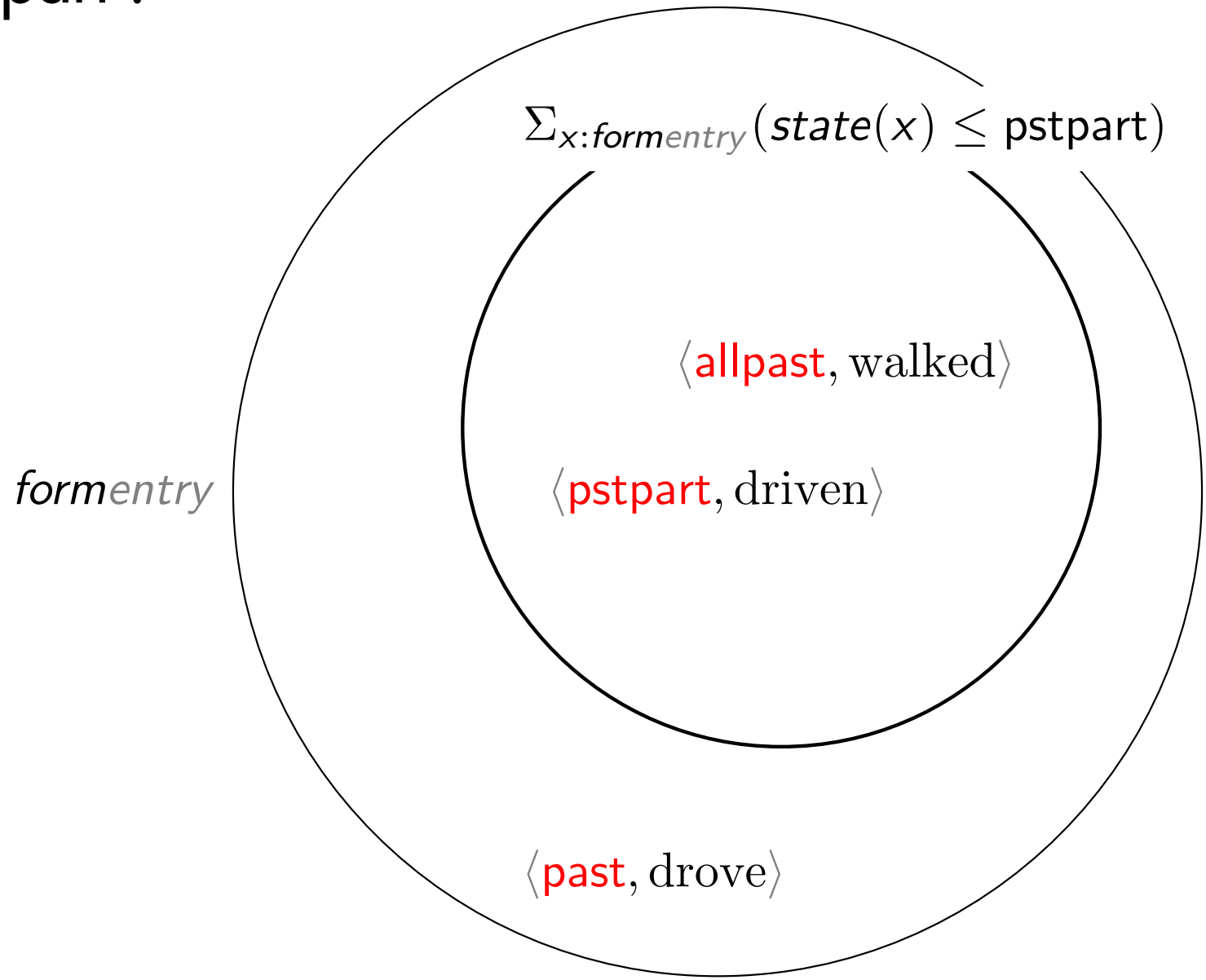
WP for sign-based CGs: a focus on morphotactics

The Σ-type, carves out a subtype.



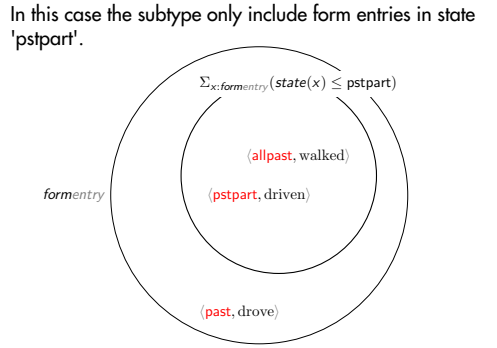
- We can see here that the effect of a sigma type is to carve out a subtype from a larger type.
- In this case the sigma type is the type of both ‘allpast’ entries and ‘past’ entries in the form paradigm.

In this case the subtype only include form entries in state 'pstpart'.



WP for sign-based CGs: a focus on morphotactics

In this case the subtype only include form entries in state 'pstpart'.



- In this case the sigma type is the type of both ‘allpast’ entries and ‘pstpart’ entries in the form paradigm.

Given this system of subtyping, we can provide two mapping rules.

$$\mathit{pastmap} : \Sigma_{x:\mathit{formentry}}(\mathit{state}(x) \leq \mathsf{past}) \rightarrow \mathit{lexicalentry}$$

$$\mathit{partmap} : \Sigma_{x:\mathit{formentry}}(\mathit{state}(x) \leq \mathsf{pstpart}) \rightarrow \mathit{lexicalentry}$$

WP for sign-based CGs: a focus on morphotactics

└Given this system of subtyping, we can provide two mapping rules.

Given this system of subtyping, we can provide two mapping rules.

$pastmap : \Sigma_{x:form_{entry}}(state(x) \leq \text{past}) \rightarrow lexical\,entry$

$partmap : \Sigma_{x:form_{entry}}(state(x) \leq \text{pstpart}) \rightarrow lexical\,entry$

- Given this subtyping mechanism, we need only provide two paradigm mapping functions that cover both weak verbs and verbs that differ in the past and past participle.
- Only the types will be shown here to avoid going into too many details.

WP for sign-based CGs: a focus on morphotactics

Given this system of subtyping, we can provide two mapping rules.

$pastmap : \Sigma_{x:form_{entry}}(state(x) \leq \text{past}) \rightarrow lexical\,entry$

$partmap : \Sigma_{x:form_{entry}}(state(x) \leq \text{pstpart}) \rightarrow lexical\,entry$

- This emphasis on relationships is what makes WP WP.
- Paradigms are organizations of information associated with a word, which for our purposes, today, can be thought of as synonymous with a lexeme. There are relationships that we want to capture between such paradigms and within such paradigms.
- By doing so, we're able to take what were axioms in LCG and turn them into the results of systematic generalizations.

There are still generalizations to capture between form paradigms.

$\langle$simple, walk$\rangle$	$\langle$simple, jump$\rangle$
$\langle$third, walk**s**$\rangle$	$\langle$third, jump**s**$\rangle$
$\langle$allpast, walk**ed**$\rangle$	$\langle$allpast, jump**ed**$\rangle$
$\langle$prspart, walk**ing**$\rangle$	$\langle$prspart, jump**ing**$\rangle$

WP for sign-based CGs: a focus on morphotactics

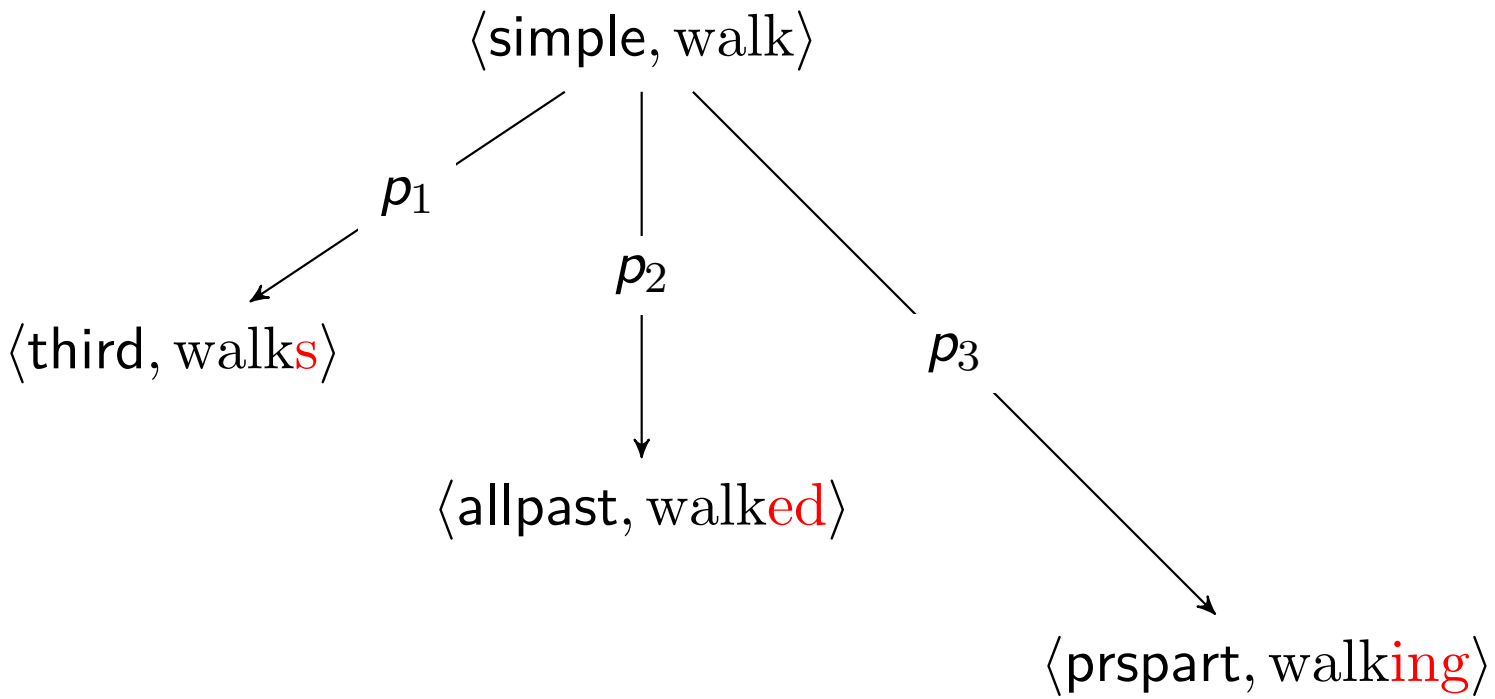
There are still generalizations to capture between form paradigms.

There are still generalizations to capture between form paradigms.

⟨simple, walk⟩	⟨simple, jump⟩
⟨third, walks⟩	⟨third, jumps⟩
⟨allpast, walked⟩	⟨allpast, jumped⟩
⟨prspart, walking⟩	⟨prspart, jumping⟩

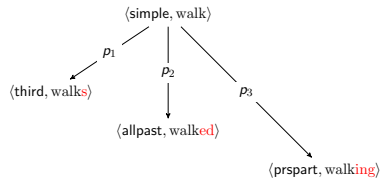
- We have so far seen examples of the theory's capabilities to capture generalizations between form entries and lexical entries.
- The form paradigms of lexemes also contain patterns that we'd like to capture.
- In this slide we can see that across form paradigms there are regularities in form.
- We'd like to make generalizations such that for any lexeme, once we know its stem, we can know its form paradigm.

We use processes to express the form paradigm internal relationships.



WP for sign-based CGs: a focus on morphotactics

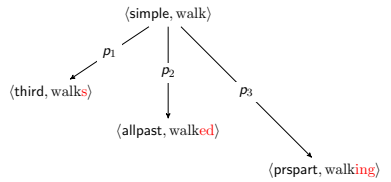
We use processes to express the form paradigm internal relationships.



- Each of the arrows emanating from *walk* represents the application of a *process function*.
- The labels on each arrow are the process function names.
- A process associates the form and state in the form entry with a form and state in a different entry.
- I will sometimes slip into "imperative" language when talking about processes "doing" things. This is only a convenient metaphor. Processes functionally relate things.
- The English weak verb example is fairly simple. In the WALK example, from a base stem, we use 3 processes that result in four forms total. The paradigm mapping functions will map to 6 lexical entry paradigm members.

WP for sign-based CGs: a focus on morphotactics

We use processes to express the form paradigm internal relationships.



- Processes can only apply to form paradigm members with particular states.
- For instance, all of the processes in the diagram in this slide can only apply to entries with the state ‘simple’.
- Processes and states are the primary ingredients of the system of *morphotactics*.
- Morphotactics can be thought of as the syntax of morphological form building.
- For instance, morphotactics specifies the ordering of affixes in complex morphology.
- To see this, it will be more instructive to look beyond English.

Form paradigms are more complex in other languages.

⟨ready, tekobejegen⟩	(I)
⟨derived, tekobejegenm⟩	(II)
⟨one, ntekobejegen⟩	(III)
⟨one, ntekobejegenm⟩	(IV)

A partial form paradigm of nominal possession by the first person for Potawatomi TKOBJEGEN, 'string'. These are quasi-underlying forms. (Morpho-)phonological process have not applied, nor are we concerned with allomorphic variation for particular morphs.

WP for sign-based CGs: a focus on morphotactics

Form paradigms are more complex in other languages.

Form paradigms are more complex in other languages.

⟨ready, tekobejegen⟩	(I)
⟨derived, tekobejegenm⟩	(II)
⟨one, ntekobejegen⟩	(III)
⟨one, ntekobejegenm⟩	(IV)

A partial form paradigm of nominal possession by the first person for Potawatomi TKOBJEGEN, 'string'. These are quasi-underlying forms. (Morpho-)phonological process have not applied, nor are we concerned with allomorphic variation for particular morphs.

- This form paradigm for a Potawatomi noun meaning 'string' (FCPC, 2014) is a nice middle ground between the simplicity of English and a more complex systems, at least given this limited subparadigm.
- In Potawatomi, possession of a noun often involves two morphs.
- One of them, represented by the *n* in lines (III) and (IV), is an exponent of first person.
- The second, the *m* suffix, in lines (II) and (IV) is an example of multiple exponence indicating possession. In other words it is a "redundant morpheme".

WP for sign-based CGs: a focus on morphotactics

Form paradigms are more complex in other languages.

Form paradigms are more complex in other languages.

⟨ready, tekobejegen⟩	(I)
⟨derived, tekobejegenm⟩	(II)
⟨one, ntekobejegen⟩	(III)
⟨one, ntekobejegenm⟩	(IV)

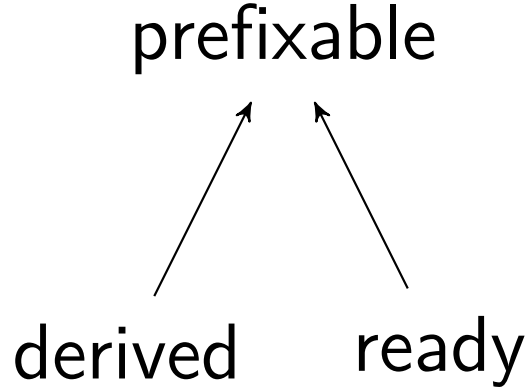
A partial form paradigm of nominal possession by the first person for Potawatomi TKOBJEGEN, 'string'. These are quasi-underlying forms. (Morpho-)phonological process have not applied, nor are we concerned with allomorphic variation for particular morphs.

- This *m* morph is obligatory for some words in the language, when they are possessed. In others it *cannot* occur.
- In some cases, such as we find here, it is optional.
- In lines (I), (III) and (IV), we see forms that appear freely in syntax.
- In line (II), the form does not occur freely.
- We would like to make a generalization in our process that prefixes the *n* that it may apply to either something in state 'ready' or state 'derived'.
- Yet, as I said above, processes may only apply to a form entry with a particular state.
- Hierarchies will help us here, just as they did in the paradigm mappings previously discussed.

Hierarchies of states are used by processes.

derived $\leq$ prefixable

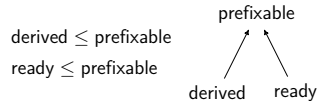
ready $\leq$ prefixable



WP for sign-based CGs: a focus on morphotactics

Hierarchies of states are used by processes.

Hierarchies of states are used by processes.

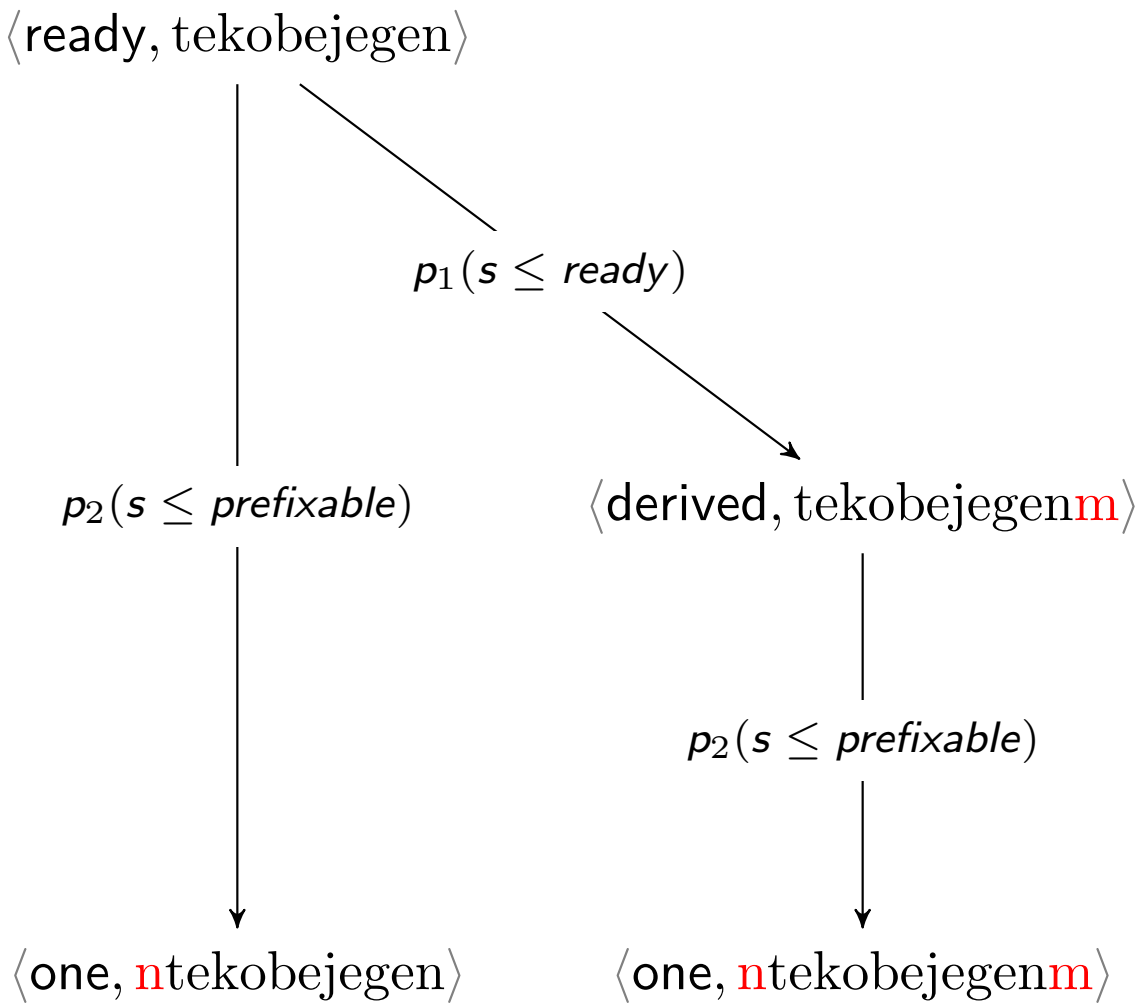


- Before we wished to say that for ‘past’ or ‘pstpart’, ‘allpast’, was a substate of both superstates – it inherited mapping behaviors of both.
- Here we want to say that two different states are substates of a sing superstate. This will capture that entries with such states are both capable of having the prefix process apply.

The processes refer to the state hierarchy.

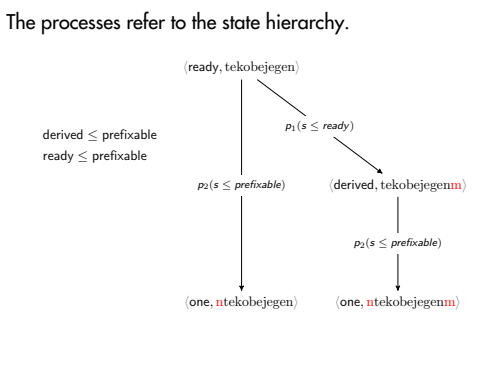
derived $\leq$ prefixable

ready $\leq$ prefixable



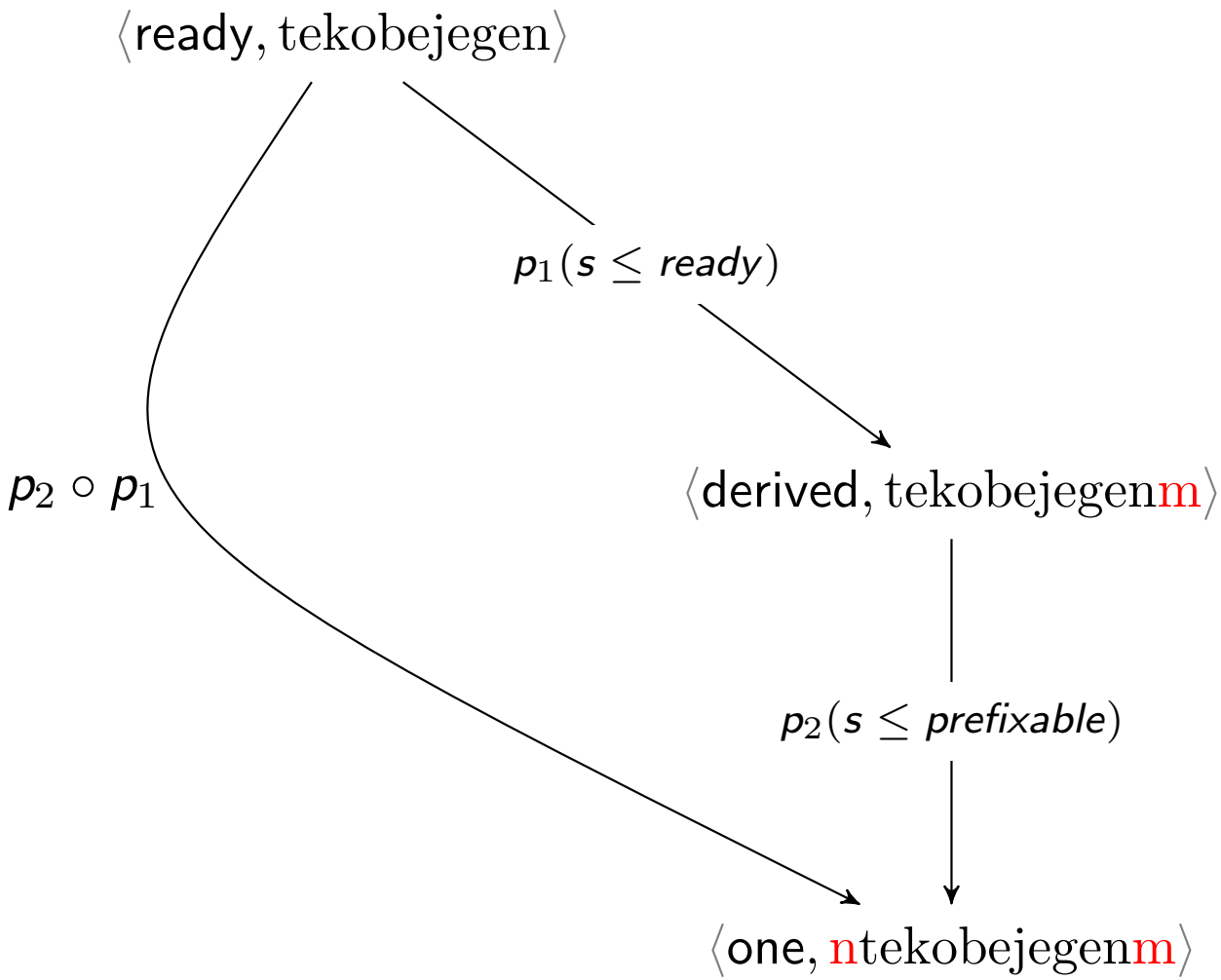
WP for sign-based CGs: a focus on morphotactics

The processes refer to the state hierarchy.



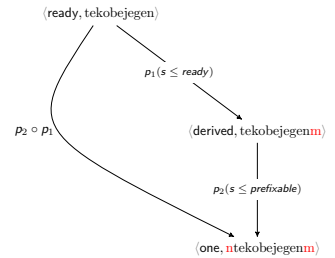
- In this slide, we can see the hierarchy in action.
- The $s \leq$ means that the process may apply when the state of the input is less than or equal to the specified state.
- The process p_1, for suffixing the m, is only able to apply to entries in the 'ready' state.
- The process p_2, is able to apply to any entry that is in a state less than or equal to 'prefixable'.

The processes can compose.



WP for sign-based CGs: a focus on morphotactics

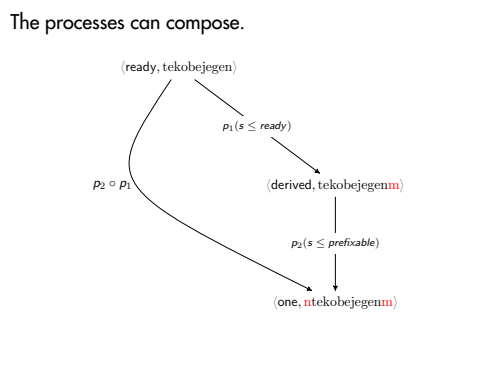
The processes can compose.



- Just a quick note on non-free form paradigm members like 'derived'.
- Some may doubt that such forms are part of our lexical knowledge.
- Note that given the graph in this slide, we can take the transitive closure of the p_1, p_2 path.

WP for sign-based CGs: a focus on morphotactics

The processes can compose.



- We are not committed to stored forms here, but relationships between forms.
- Perhaps this composed path would more accurately reflect plausible form storage, while the other is more reflective of the analogical inferences that are made across paradigms.
- This is a nice attribute of the theory as there is some evidence that competing pathways are a psycholinguistic reality. See Hay (2001) for one discussion of the issue.

States cannot always be simple.

⟨simple, miku⟩	(V)
⟨past, mikurka⟩	(VI)
⟨perf, mikushka⟩	(VII)
⟨{simple, one}, mikuni⟩	(VIII)
⟨{past, one}, mikurkani⟩	(IX)
⟨{perf, one}, mikushkani⟩	(X)

A partial form paradigm of verbal inflection for Ecuadorian Quechua MIKUNA, 'eat'.

WP for sign-based CGs: a focus on morphotactics

States cannot always be simple.

States cannot always be simple.

⟨simple, miku⟩	(V)
⟨past, mikurka⟩	(VI)
⟨perf, mikushka⟩	(VII)
⟨{simple, one}, mikuni⟩	(VIII)
⟨{past, one}, mikurkani⟩	(IX)
⟨{perf, one}, mikushkani⟩	(X)

A partial form paradigm of verbal inflection for Ecuadorian Quechua MIKUNA, 'eat'.

- The Potawatomi example remains fairly simple.
- In order to categorize forms in a way that facilitates mappings to lexical entries, it is necessary that state names become more complex than the singleton state names that we've already seen.

WP for sign-based CGs: a focus on morphotactics

States cannot always be simple.

States cannot always be simple.

⟨simple, miku⟩	(V)
⟨past, mikurka⟩	(VI)
⟨perf, mikushka⟩	(VII)
⟨{simple, one}, mikuni⟩	(VIII)
⟨{past, one}, mikurkani⟩	(IX)
⟨{perf, one}, mikushkani⟩	(X)

A partial form paradigm of verbal inflection for Ecuadorian Quechua MIKUNA, 'eat'.

- In Ecuadorian Quechua, forms are built up in an agglutinative fashion.
- Each paradigm member, with each successive addition, requires a new state name both for paradigm mappings and for possible form paradigm internal scope-like effects, where certain suffixes may be incompatible.
- In this slide the past tense is associated with 'rka' and the perfect suffix with 'shka'. These suffixes do not co-occur (Catta Q., 1994).
- Following them suffixes for person are added.

WP for sign-based CGs: a focus on morphotactics

States cannot always be simple.

States cannot always be simple.

⟨simple, miku⟩	(V)
⟨past, mikurka⟩	(VI)
⟨perf, mikushka⟩	(VII)
⟨{simple, one}, mikuni⟩	(VIII)
⟨{past, one}, mikurkani⟩	(IX)
⟨{perf, one}, mikushkani⟩	(X)

A partial form paradigm of verbal inflection for Ecuadorian Quechua MIKUNA, 'eat'.

- In order to aid in capturing the generality of the pattern, a new singleton state name is not provided for each individual form paradigm member – though what I am doing is equivalent to providing an atomic name, in some sense.
- The formalism allows for the addition of 'one' to the preexisting state name set when the first person suffix is added.
- Formally, all of the previously seen state "names", such as 'allpast', etc., are singleton set-like objects. I generally omit the brackets for singletons.
- It seems inevitable for such agglutinating languages that increase in *something* will be correlated to increase in recognizable morphs. Neither compositionality nor an increase in semantic complexity is necessarily implied by this system.

A final example demonstrates overabundance in plural marking.

$$\langle \text{simple}, \text{allqu} \rangle \quad \text{(XI)}$$

$$\langle \{\text{plural1}, \text{simple}\}, \text{allqus} \rangle \quad \text{(XII)}$$

$$\langle \{\text{plural2}, \text{simple}\}, \text{allqukuna} \rangle \quad \text{(XIII)}$$

$$\langle \{\text{plural2}, \text{plural1}, \text{simple}\}, \text{allquskuna} \rangle \quad \text{(XIV)}$$

$$\langle \{\text{plural1}, \text{plural2}, \text{simple}\}, \text{allqukunas} \rangle \quad \text{(XV)}$$

A partial form paradigm of plural inflection for Bolivian Quechua ALLQU, 'dog'.

WP for sign-based CGs: a focus on morphotactics

A final example demonstrates overabundance in plural marking.

A final example demonstrates overabundance in plural marking.

⟨simple, allqu⟩	(XI)
⟨{plural1, simple}, allqus⟩	(XII)
⟨{plural2, simple}, allqukuna⟩	(XIII)
⟨{plural2, plural1, simple}, allquskuna⟩	(XIV)
⟨{plural1, plural2, simple}, allqukunas⟩	(XV)

A partial form paradigm of plural inflection for Bolivian Quechua ALLQU, 'dog'.

- A difficult problem in morphological theory is handling cases where morphological exponents can occur in more than one position.
- Cochabamba Bolivian Quechua borrowed the Spanish plural and it co-exists with the native plural *kuna* (Lastra, 1968).
- The two suffixes can be used together as well, in the pattern displayed.
- The interesting thing about this is that *s* is not only able to co-occur with *kuna* but vary in position.
- The issue is ensuring that one stays relatively general in capturing the pattern, while not allowing the pattern to run away, with *kuna* following whenever there is an *s* and *s* following whenever there is a *kuna*.

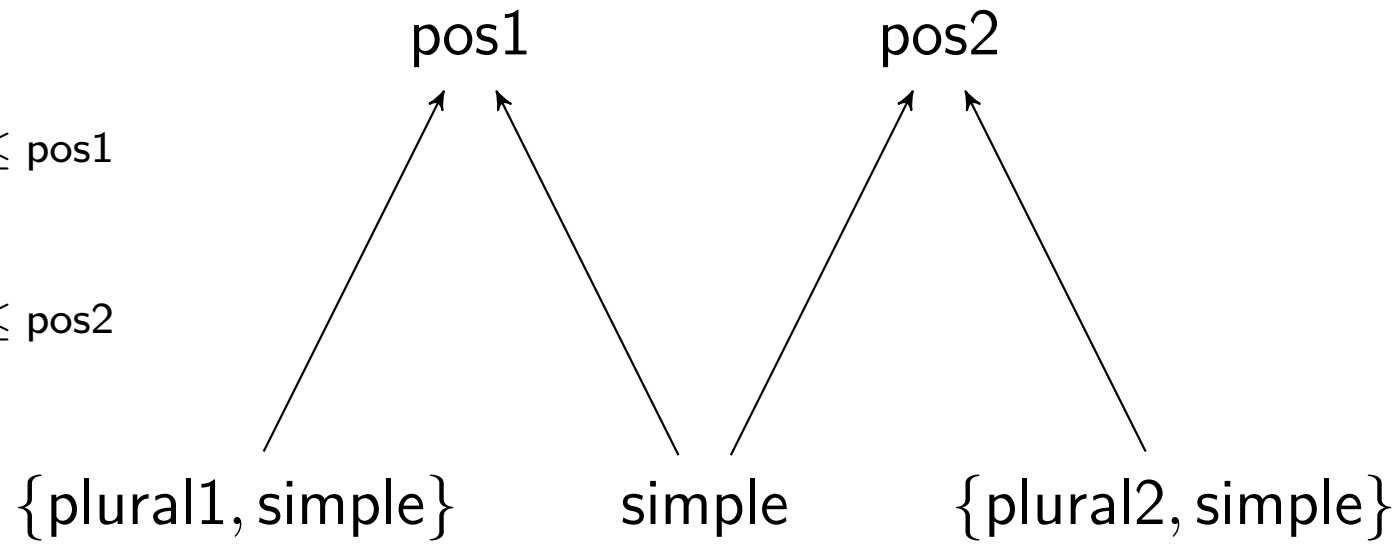
Utilizing a state hierarchy, the ordering pattern is captured.

$\text{simple} \leq \text{pos1}$

$\{\text{plural1}, \text{simple}\} \leq \text{pos1}$

$\text{simple} \leq \text{pos2}$

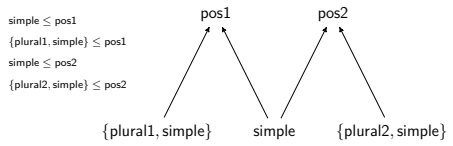
$\{\text{plural2}, \text{simple}\} \leq \text{pos2}$



WP for sign-based CGs: a focus on morphotactics

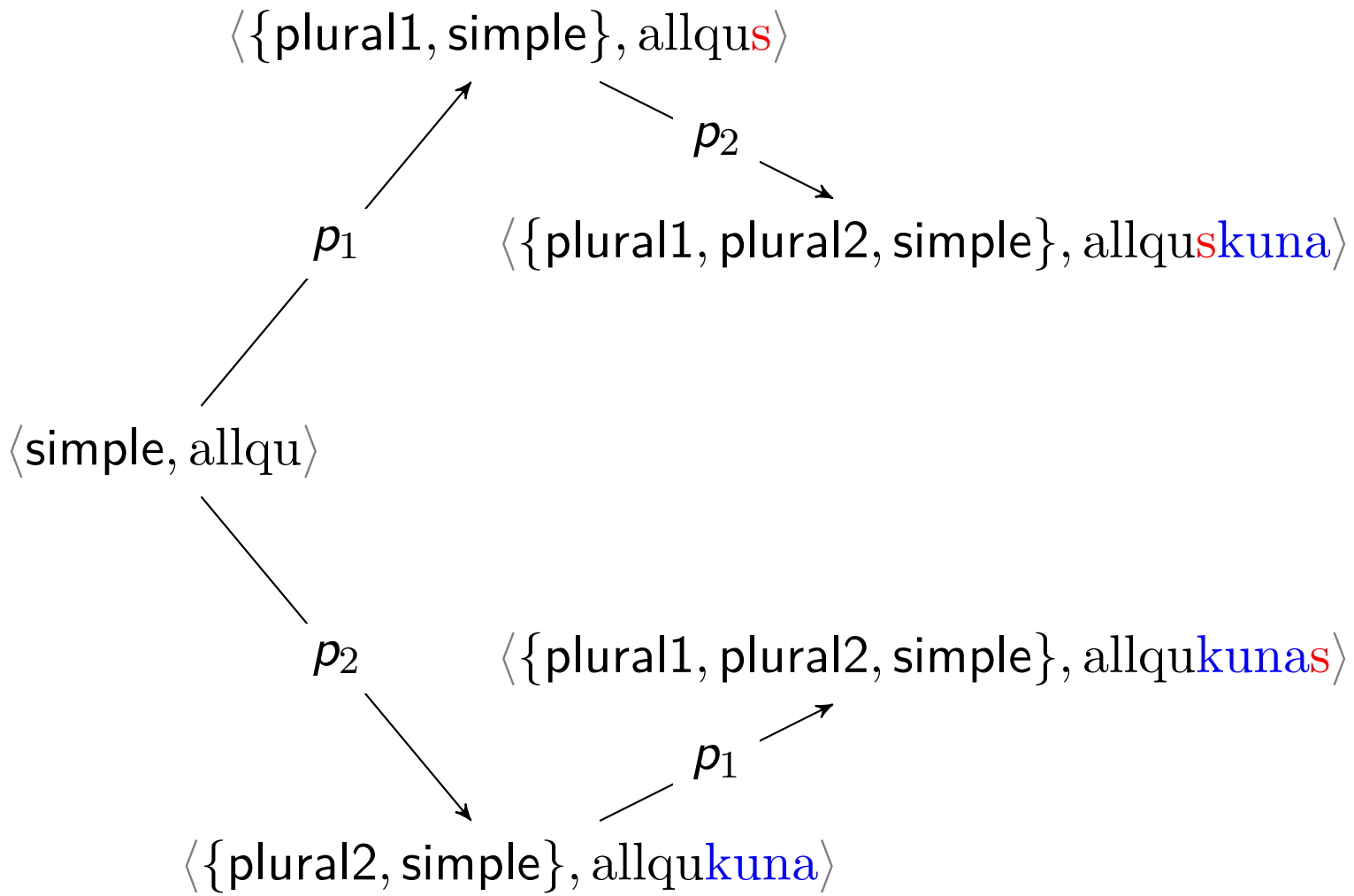
Utilizing a state hierarchy, the ordering pattern is captured.

Utilizing a state hierarchy, the ordering pattern is captured.



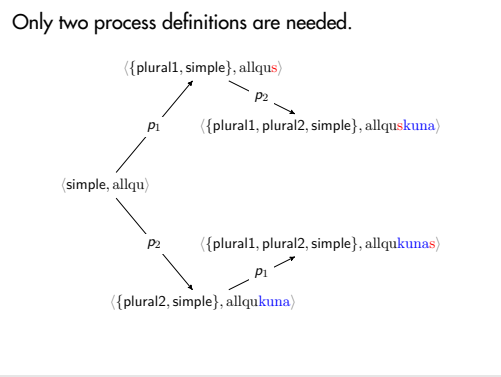
- The solution that I use here involves two position categories.
- 'pos1' allows for affixing *kuna*.
- 'pos2' allows for affixing *s*.

Only two process definitions are needed.



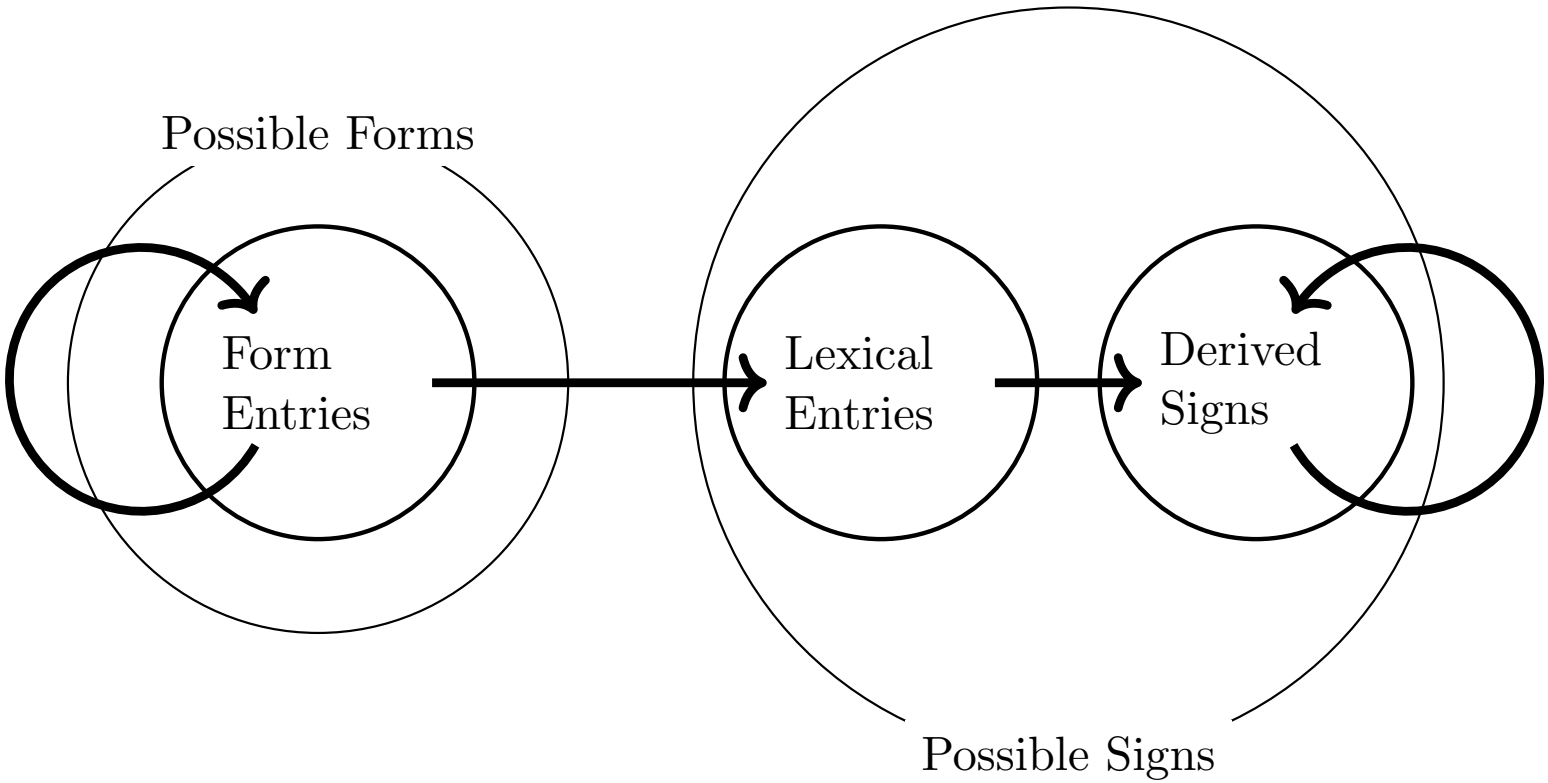
WP for sign-based CGs: a focus on morphotactics

Only two process definitions are needed.



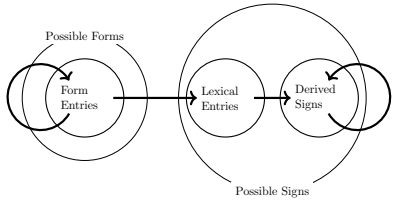
- Process p_1 appends an *s* to something with a state less than or equal to 'pos2'.
- Process p_2 appends a *kuna* to something with a state less than or equal to 'pos1'.
- Process p_1 always adds 'plural1' to the current state name.
- Process p_2 always adds 'plural2' to the current state name.
- State '{plural1, plural2, simple}' combined is not less than or equal to either state 'pos1' or 'pos2', so no further plural suffixing can continue. I do not make use of the natural subset order over these complex category designations.

We've gone a long way toward separating possible lexical entries from non-lexical entries.



WP for sign-based CGs: a focus on morphotactics

We've gone a long way toward separating possible lexical entries from non-lexical entries.

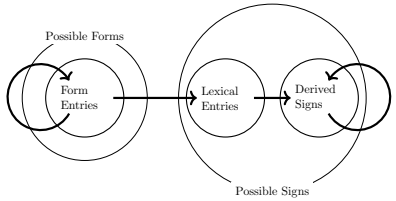


└We've gone a long way toward separating possible lexical entries from non-lexical entries.

- Although there are many reasons to explore the syntax morphology interface and many reasons to explore morphological patterns in their own right, a benefit for a sign based categorial grammar, such as LCG, is that we can reveal patterns in lexical entries, which were previously only axioms.
- We have two areas of information that we are considering.
- There are possible signs, some of which do not conform to a specific grammar and some which are unlikely to correspond to any grammar.
- Members of these that are grammatical can be determined in two ways.
- Lexical entries are mappings from a form paradigm to a paradigm in the sign codomain.

WP for sign-based CGs: a focus on morphotactics

We've gone a long way toward separating possible lexical entries from non-lexical entries.



└We've gone a long way toward separating possible lexical entries from non-lexical entries.

- LCG can use these lexical entries to derive additional signs.
- LCG can also derive grammatical signs from the derived signs.
- The form entries, which feed this, are built from other form entries via processes.
- Though the lexicon continues to contain purely stipulated content in this theory, we can capture a great many generalizations by adding a morphological component to LCG.

References I

- Ajdukiewicz, Kasimir (1935). "Die syntaktische Konnexität". In: *Studia Philosophica* 1, pp. 1–27.
- Bar-Hillel, Yehoshua (1953). "A quasi-arithmetical notation for syntactic description". In: *Language* 29.1, pp. 47–58.
- Bobaljik, Jonathan David (2002). "Syncretism without paradigms. Remarks on Williams 1981, 1994". In: *Yearbook of Morphology 2001*. Ed. by Geert Booij and Jaap Van Marle, pp. 53–85.
- Catta Q., Javier (1994). *Gramática del quichua ecuatoriano*. Editorial Abya Yala.
- Chomsky, Noam (1957). *Syntactic structures*. Mouton, The Hague.
- Curry, Haskell B (1961). "Some logical aspects of grammatical structure". In: *Structure of language and its mathematical aspects*. Vol. 12. American Mathematical Society, pp. 56–68.

References II

- de Saussure, Ferdinand (2011). *Course in general linguistics*. Columbia University Press.
- FCPC, (Forest County Potawatomi Community) (2014). *Ézhe-bmadzimgek gdebodwéwadmi-zheshmomenan. Potawatomi Dictionary*. Comp. by Monica Macaulay, Lindsay Marean, Laura Welcher, and Kimberly Wensaut. Forest County Potawatomi Community. ISBN: 978-0-578-14283-8.
- Harley, Heidi (2008). "When is syncretism more than a syncretism? Impoverishment, metasyncretism and underspecification". In: *Phi theory. Phi-features across modules and interfaces*. Ed. by Daniel Harbour, David Adger, and Susana Béjar. Oxford University Press, pp. 251–294.
- Harris, Zellig S (1951). *Methods in Structural Linguistics*. University of Chicago Press.

References III

- Hay, Jennifer (2001). "Lexical frequency in morphology: is everything relative?" In: *Linguistics* 39.6, pp. 1041–1070.
- Koenig, Jean-Pierre and Daniel Jurafsky (1994). "Type underspecification and On-line Type Construction in the Lexicon". In: *West coast conference on formal linguistics*. Vol. 13. CSLI Publications/SLA, pp. 270–285.
- Lambek, Joachim (1958). "The mathematics of sentence structure". In: *American Mathematical Monthly* 65.3, pp. 154–170.
- Lastra, Yolanda (1968). *Cochabamba Quechua Syntax*. The Hague: Mouton.
- Martin-Löf, Per (1984). *Intuitionistic type theory*. Naples: Bibliopolis.

References IV

- Mihaliček, Vedrana and Carl Pollard (2012). "Distinguishing phenogrammar from tectogrammar simplifies the analysis of interrogatives". In: *Formal Grammar*. Ed. by Philippe de Groote and Mark-Jan Nederhof. Springer, pp. 130–145.
- Pollard, Carl and Ivan Sag (1994). *Head-driven phrase structure grammar*. University of Chicago Press.
- Robins, Robert H. (1959). "In defence of WP". In: *Transactions of the Philological Society* 58.1, pp. 116–144.
- Stump, Gregory (2016). *Inflectional Paradigms. Content and Form at the Syntax-morphology Interface*. Cambridge University Press.

References V

- Thornton, Anna M (2011). "Overabundance (multiple forms realizing the same cell): a non-canonical phenomenon in Italian verb morphology.". In: *Morphological autonomy: Perspectives from Romance inflectional morphology*. Ed. by Martin Maiden, John Charles Smith, Maria Goldbach, and Marc-Olivier Hinzelin. Oxford University Press, pp. 358–381.
- Williams, Edwin (1981). "On the notions "Lexically related" and "Head of a word"""'. In: *Linguistic Inquiry* 12.2, pp. 245–274.
- – (1994). "Remarks on lexical knowledge". In: *Lingua*, pp. 7–34.