COSC421: Neural Models of Language

Lecture 10: Introduction to empiricist models of syntax

Alistair Knott

Dept. of Computer Science, University of Otago

Recap

Lecture 6: 'where language is implemented in the brain'.

- Phonology: auditory cortex, (pre-)motor cortex, STS, IP
- Word meanings: temporal cortex, (pre-)motor cortex, PFC
- Syntax: Broca's area and anterior temporal areas

Lecture 7: how infants learn single words.

- Statistical learning of phonological input-output mappings
- Statistical learning of word forms (SRN)
- Cross-situational learning; role of phonological STM

Lectures 8 & 9: what syntax is (according to Chomsky).
Today: non-Chomskyan models of syntax & syntactic development.

Outline of today's lecture

1. Learning syntax: early developmental stages
2. The nativist-empiricist debate
3. Empiricist models of syntax

Outline of today's lecture

Learning syntax: early developmental stages

Early syntactic development has some fairly well agreed stages.

1. Single-word utterances (**holophrases**).

Utterances in service of specific *goals*.

- The goal can be 'declarative' (e.g. *car!*)
- The goal can be 'imperative' (e.g. *doggy! more!*)

It's only when children have learned the mapping between meanings and words that such utterances become effective.

- Note: children not only learn that words have *meanings*: they also learn that they have *effects*.

Learning syntax: early developmental stages

2. Simple two-word utterances.

Word combinations: unstructured collections of words.

- E.g. *my . . . cup! cup . . . my!*

Pivot schemas: two word units structured around a single word

- E.g. *my cup! my cake!* [*my X*]

Tomasello: pivot schemas support some generalisation, but are mainly based on surface word ordering conventions.

Learning syntax: early developmental stages

3. **Item-based syntactic constructions**

At 18 months, children begin to understand simple transitive sentences.

Around 24 months: the earliest 'syntactic constructions' are produced.

- Children begin to produce transitive sentences.
- Children begin to use syntactic **function words** (e.g. *the*, *of*) and inflections (e.g. *like***s**).

The interesting thing about early constructions is that they tend to be *tied to specific words*.

- **Open** *it with this.*
- *He* **hit** *me this.*

Learning syntax: early developmental stages

4. Progressively more complex syntactic constructions.

At this point, utterances are complex enough that you need a proper syntactic theory to chart development.

That's where things start to get contentious.

Outline of today's lecture

The nativist-empiricist debate

There's a huge debate between **nativists** and **empiricists** in developmental linguistics.

- Nativists believe that infants are born with 'knowledge' of the universal properties of language.
  All they have to learn from their environment are the **parameter settings** which define their particular language.
- Empiricists believe that infants use *general-purpose learning mechanisms* to acquire language.
  They learn language 'from scratch'.

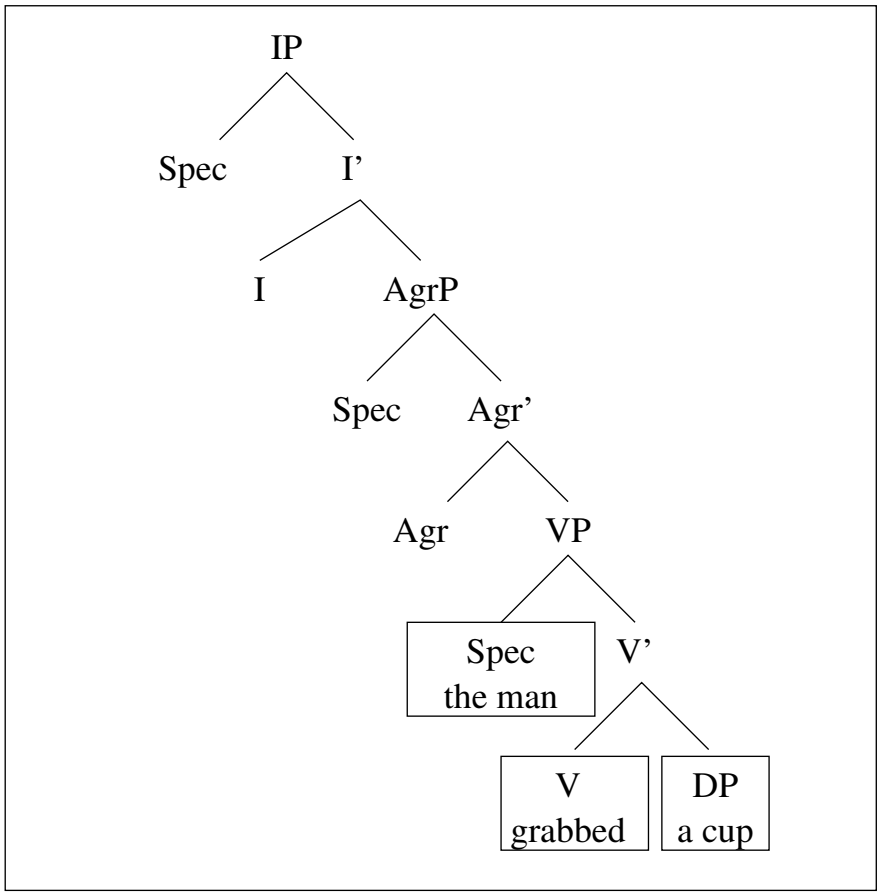
An example of a nativist model: Minimalism

Recall:

- The Minimalist model of ‘The man grabbed a cup’ holds that the same **LF** structure underlies this sentence in every language.
- This LF structure contains multiple positions for the agent, patient and inflected verb. (Because these items ‘move’ during derivation.)
- Children are born knowing how to derive the LF representation.
- What they have to learn is the *mapping from LF to PF*.
  I.e. whether to ‘read out’ items before or after movement.

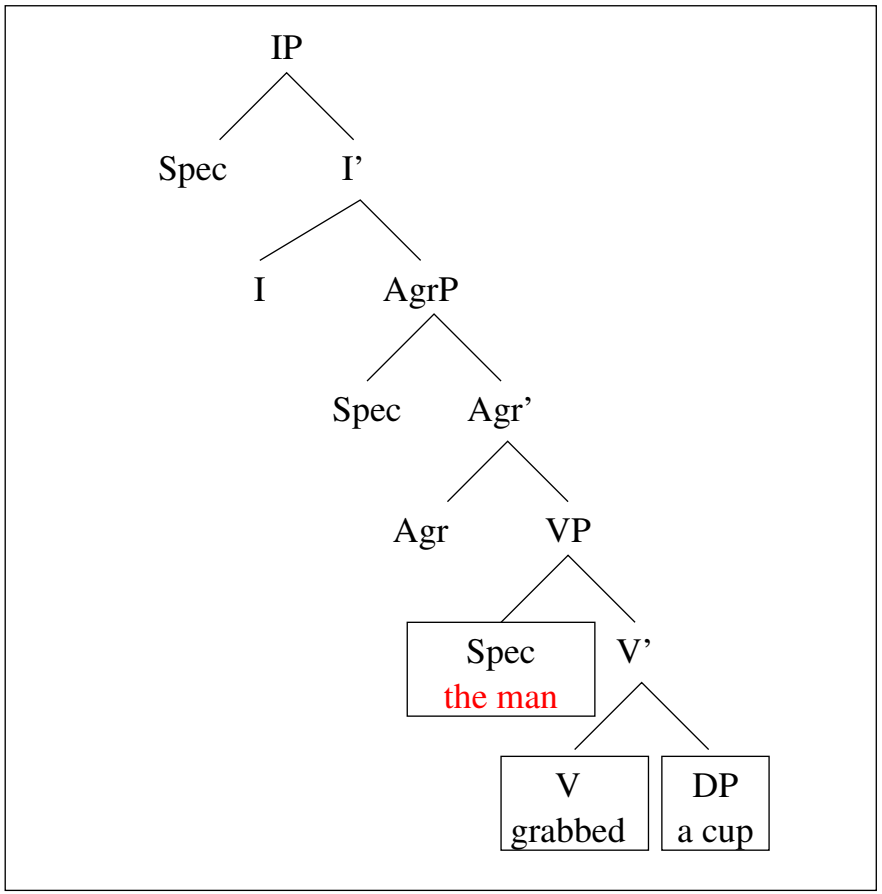
An example of a nativist model: Minimalism

Here's the LF derivation of our example sentence.



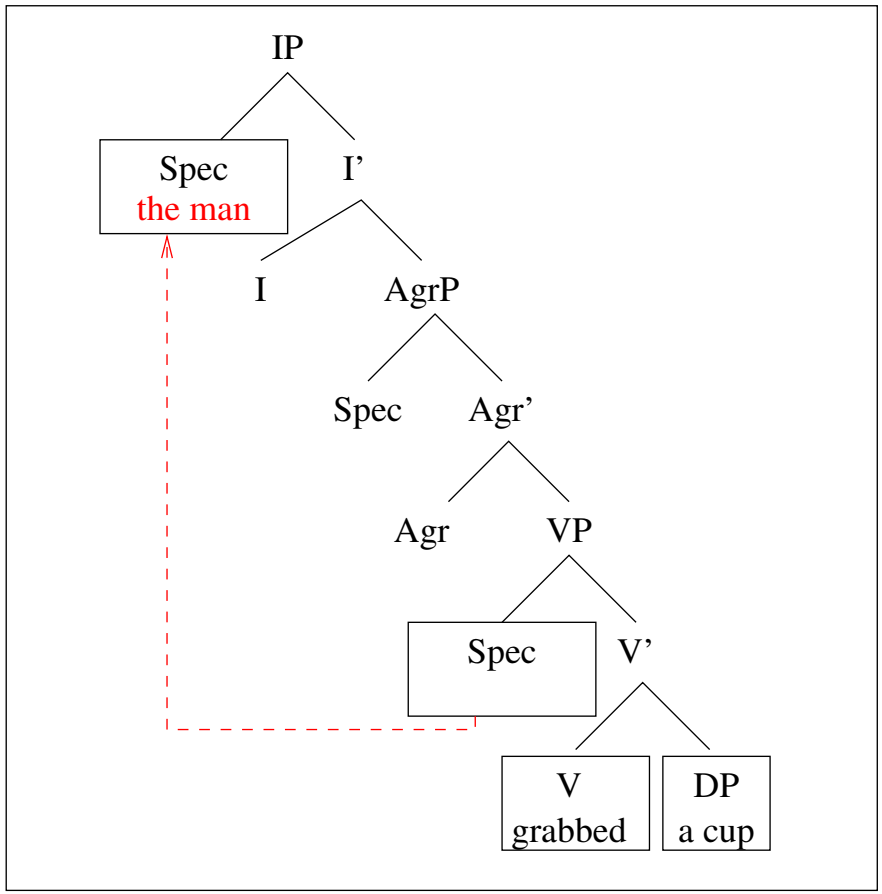
An example of a nativist model: Minimalism

Here's the LF derivation of our example sentence.



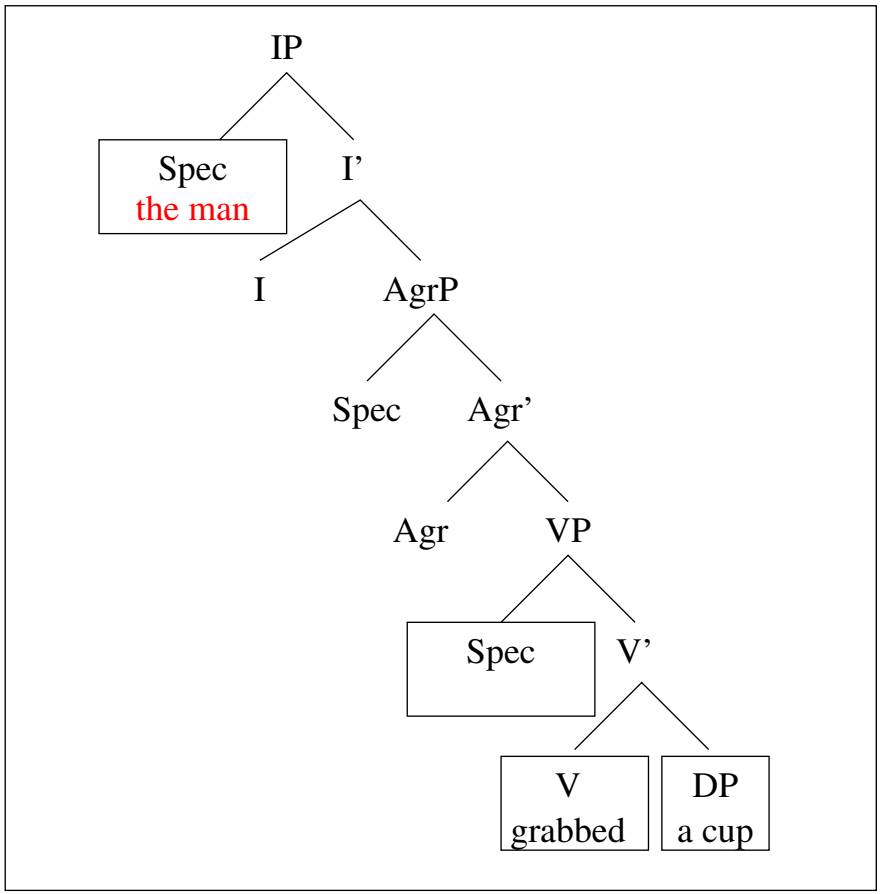
An example of a nativist model: Minimalism

Here's the LF derivation of our example sentence.



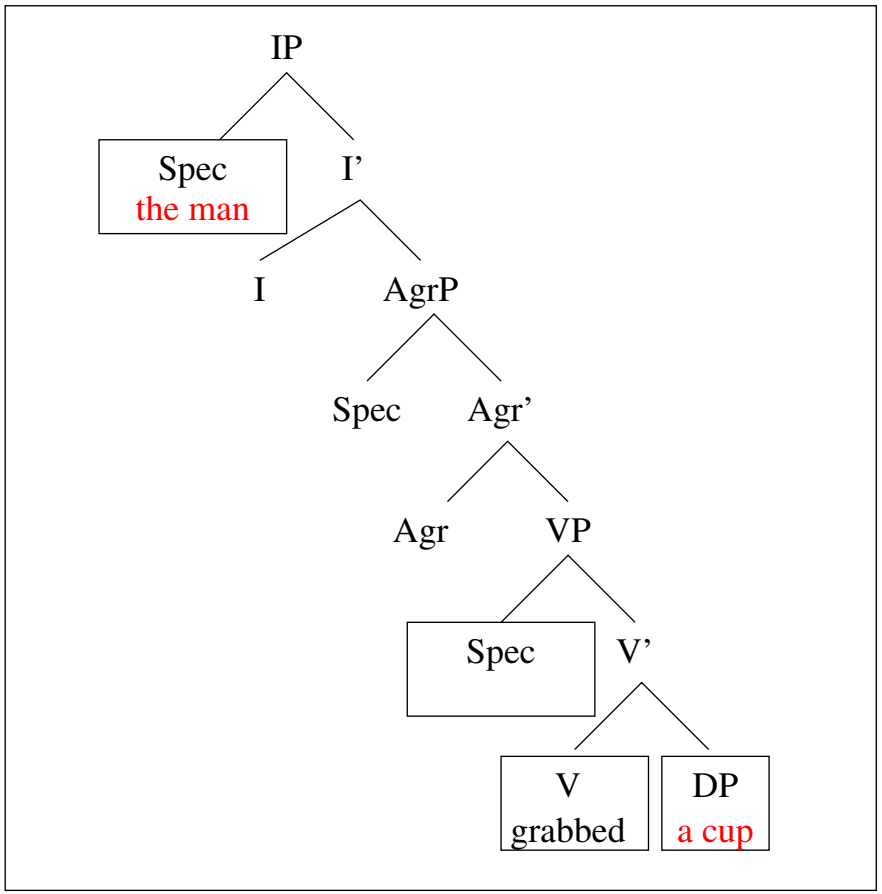
An example of a nativist model: Minimalism

Here's the LF derivation of our example sentence.



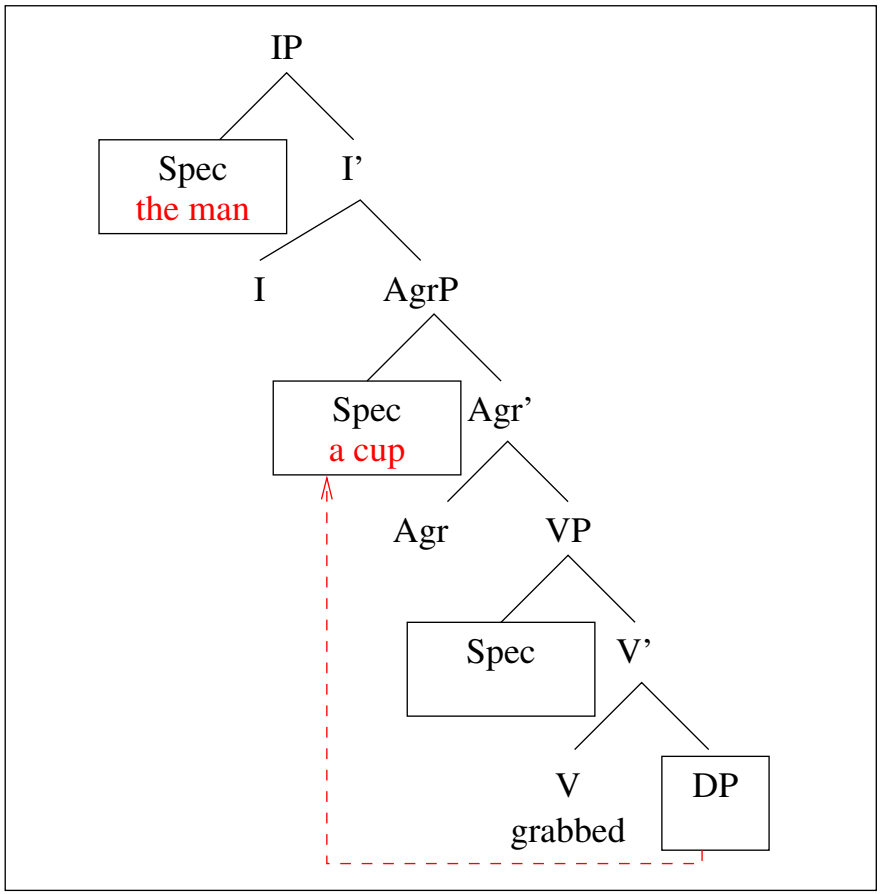
An example of a nativist model: Minimalism

Here's the LF derivation of our example sentence.



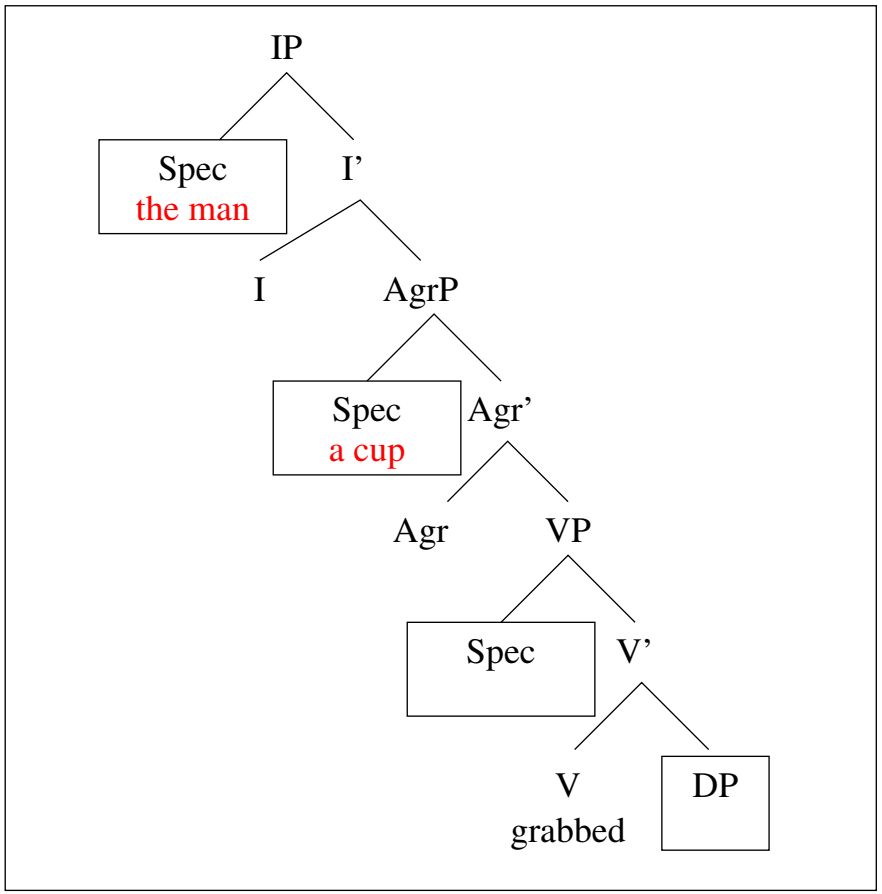
An example of a nativist model: Minimalism

Here's the LF derivation of our example sentence.



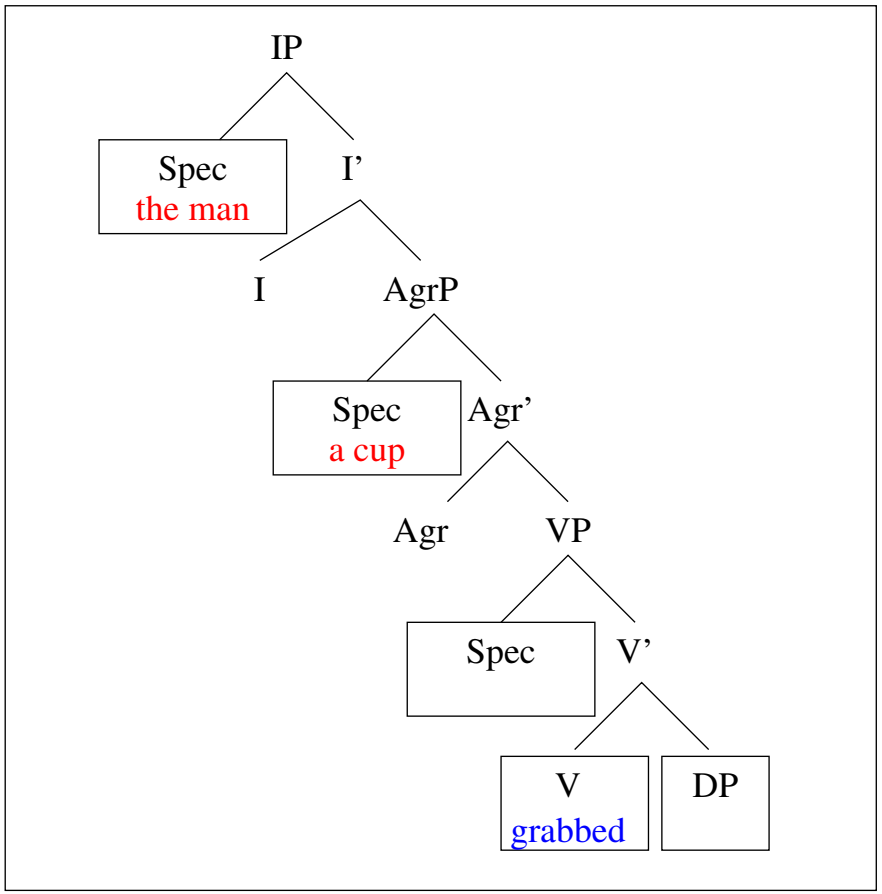
An example of a nativist model: Minimalism

Here's the LF derivation of our example sentence.



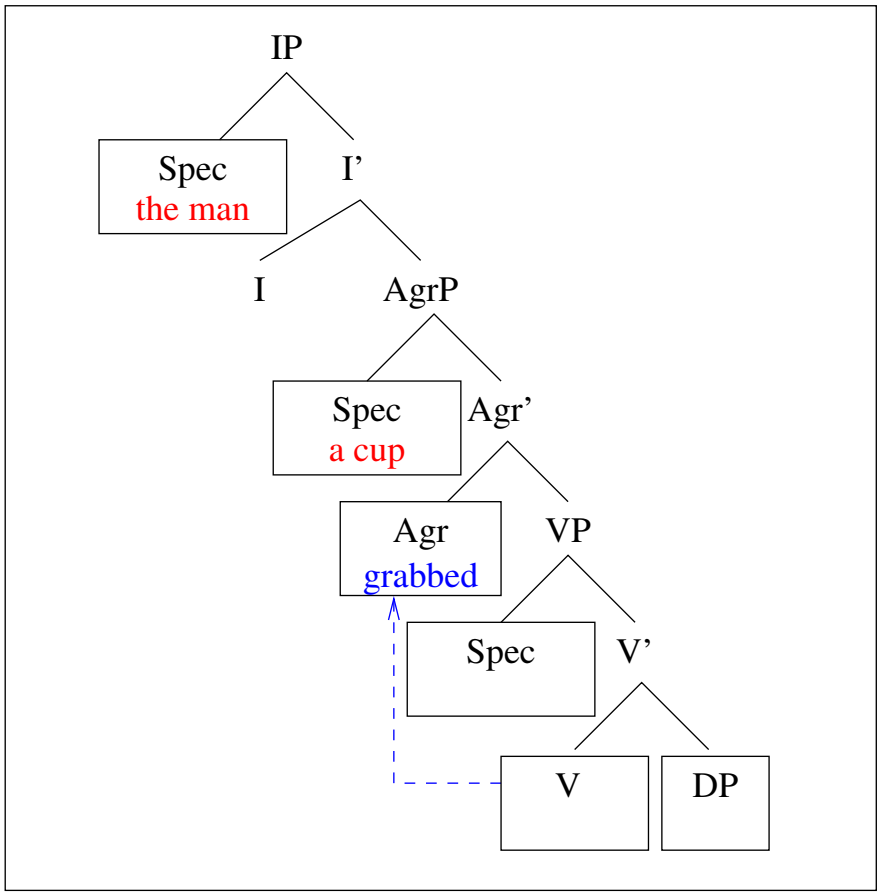
An example of a nativist model: Minimalism

Here's the LF derivation of our example sentence.



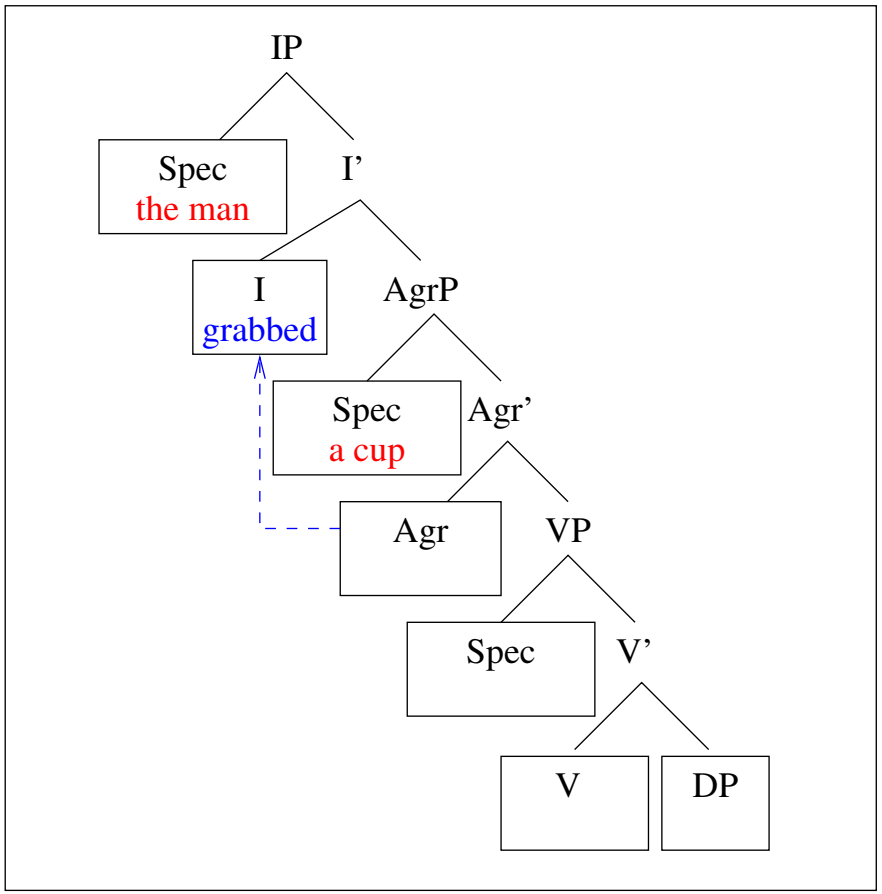
An example of a nativist model: Minimalism

Here's the LF derivation of our example sentence.



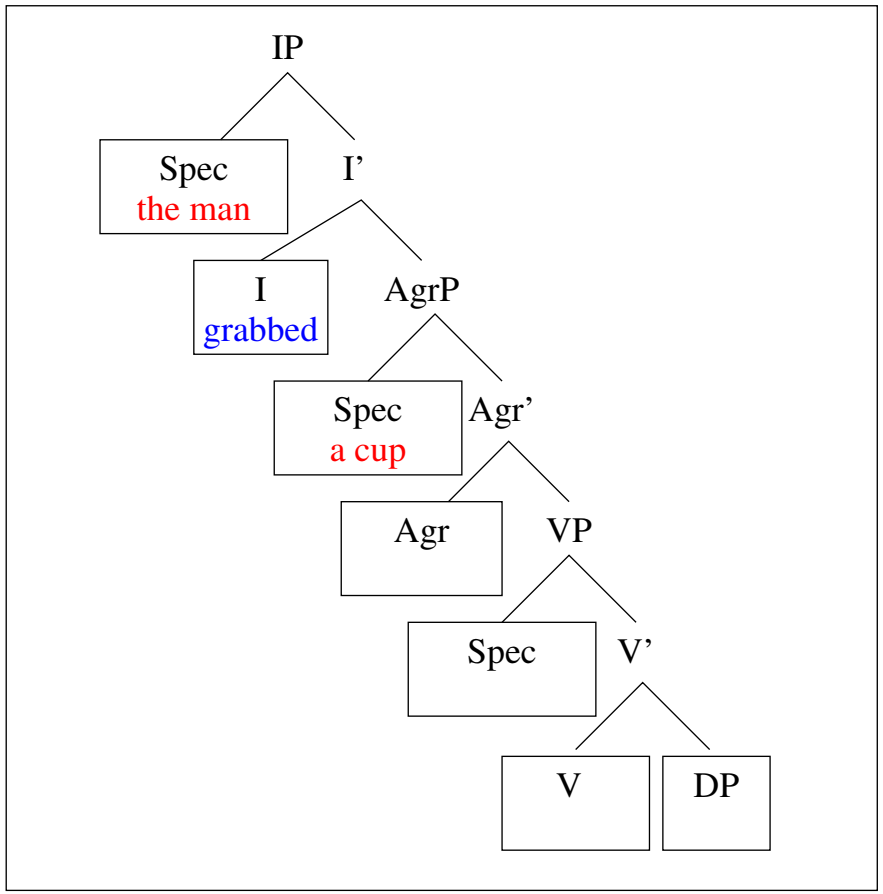
An example of a nativist model: Minimalism

Here's the LF derivation of our example sentence.



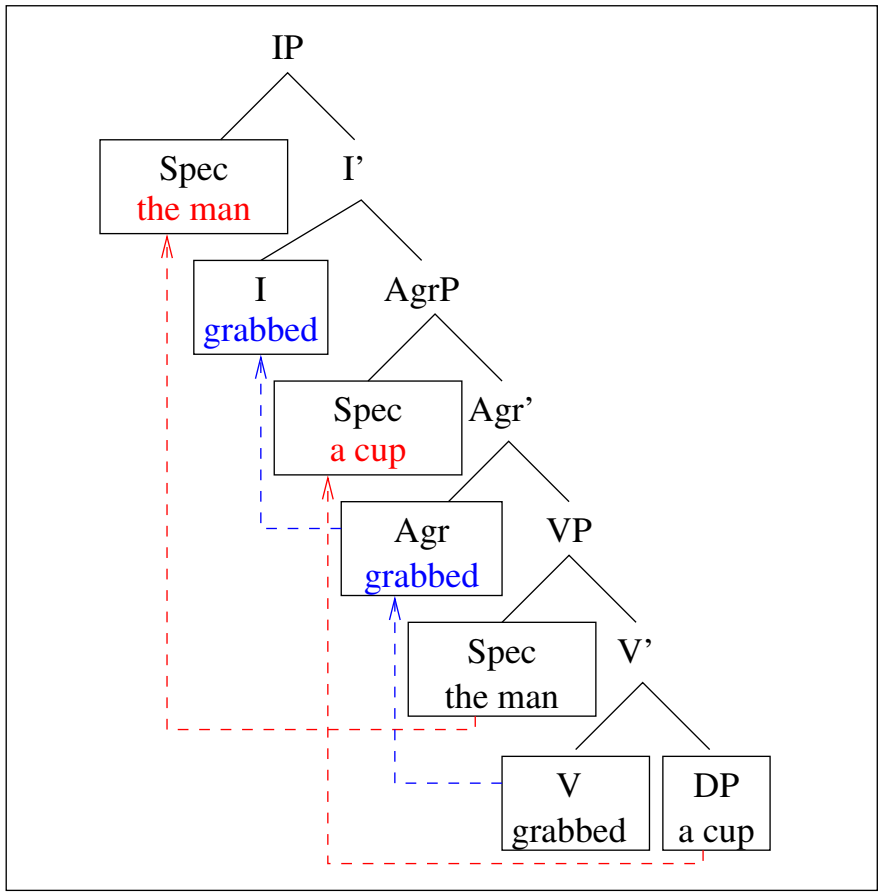
An example of a nativist model: Minimalism

Here's the LF derivation of our example sentence.



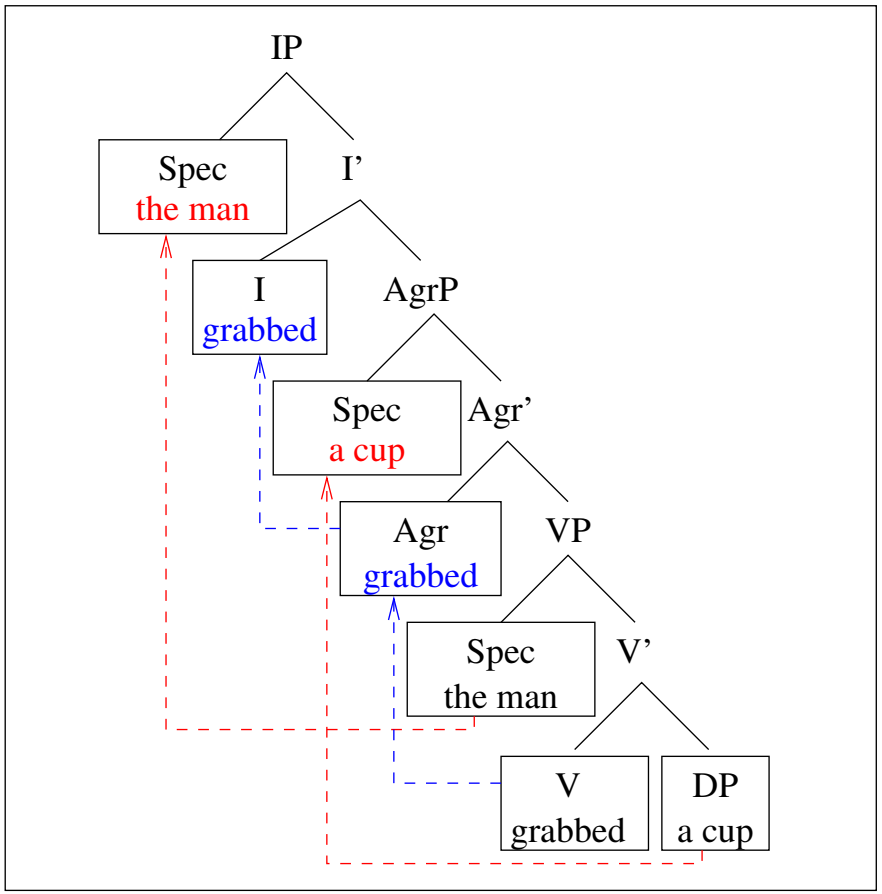
An example of a nativist model: Minimalism

Here's the LF derivation of our example sentence.



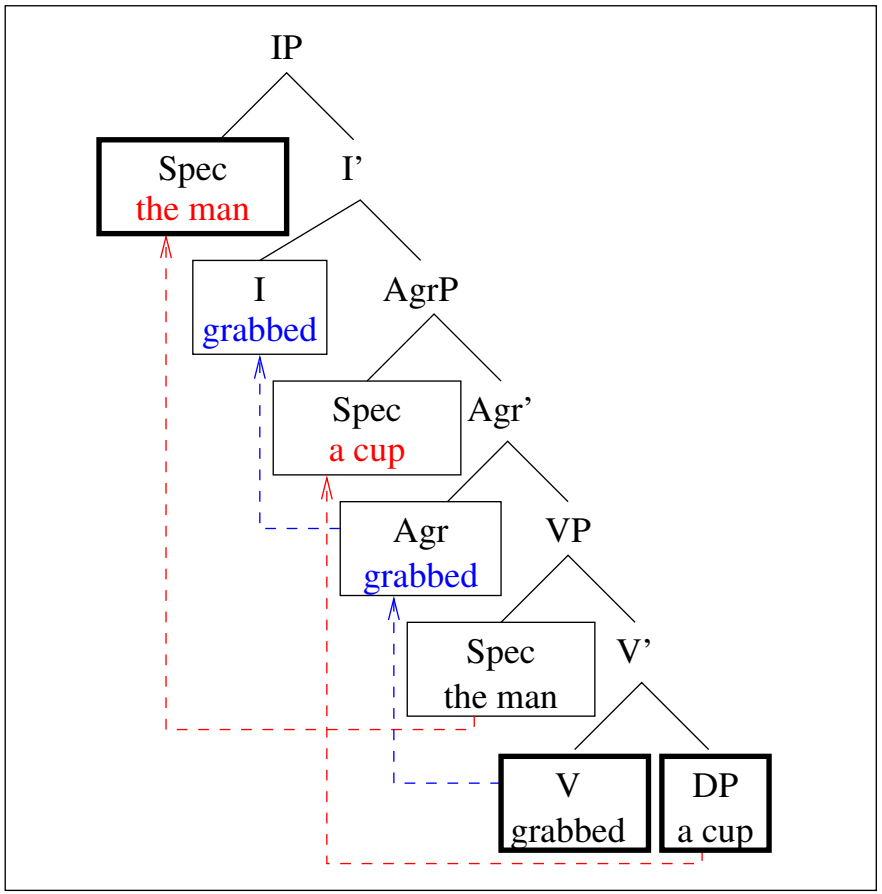
An example of a nativist model: Minimalism

Both agent and patient can be read out before or after movement.
The inflected verb can be read out in three positions.
All the child has to learn is 'when to read out each item'.



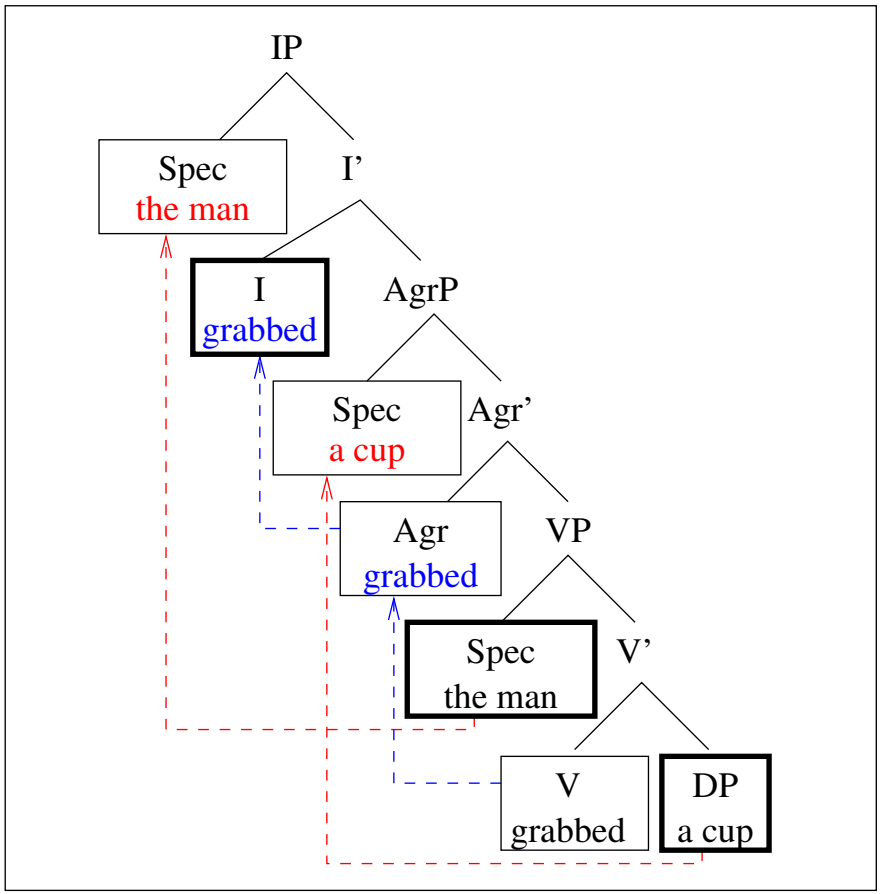
An example of a nativist model: Minimalism

In English, we read out as follows:



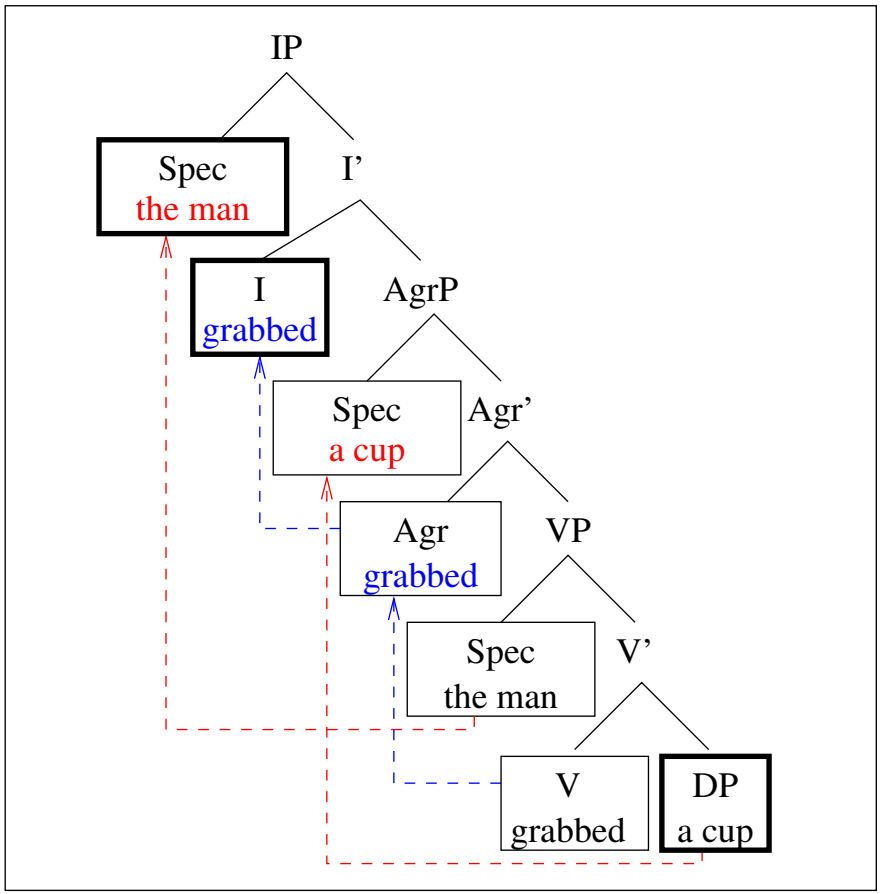
An example of a nativist model: Minimalism

In Maori, we read out as follows:



An example of a nativist model: Minimalism

In French/Italian, we read out as follows:



Some arguments for a nativist position

1. ‘Poverty of the stimulus’ arguments. (Chomsky, 1980)
   - ‘There’s not enough information in language exposure data to learn a language.’
   - ‘Language is just too complex to be able to learn from data.’ (Minimalist syntax is certainly complex!)

2. Arguments from pidgins and creoles (Bickerton, 1981)
   - **Pidgins** are languages which are ‘invented’ when two language communities meet, and need to communicate.
     They are not true natural languages.
   - Children who grow up in communities speaking a pidgin language develop a **creole**.
     Creoles have all of the syntactic complexity of ‘established’ natural languages.

Outline of today's lecture

Empiricist models of syntax

Empiricist linguists argue that children have very powerful general-purpose learning mechanisms.

- These are sufficient to acquire a language without (much) innate language-specific machinery.

The training data: **utterances** occurring in **communicative contexts**.

- There are regularities *within utterances*.
- There are regularities *linking utterances and contexts*.

Children have **pattern-finding mechanisms** which pick up these regularities.

What pattern-finding mechanisms are involved?

What pattern-finding mechanisms are involved?

1. A mechanism which finds regularities in sequential data.
   - Consider the following sequence: *John went to the. . .*
     What word comes next?
   - We've already seen that infants can pick up regularities in a stream of phonemes (Saffran *et al.*).
     ga bi ro **to ba di** *ga bi ro* **to ba di**

What pattern-finding mechanisms are involved?

1. A mechanism which finds regularities in sequential data.
   - Consider the following sequence: *John went to the. . .*
     What word comes next?
   - We've already seen that infants can pick up regularities in a stream of phonemes (Saffran *et al.*).
     ga bi ro **to ba di** *ga bi ro* **to ba di**

2. A mechanism which finds mappings between pairs of *complex patterns*.
   - We've already hypothesised such a mechanism in our accounts of the mirror system.
   - It's also attested in our ability to perform analogical reasoning.

What sorts of pattern are found?

What sorts of pattern are found?

1. Patterns are *statistical tendencies*, rather than universal rules.

A traditional grammar divides sentences discretely into ‘well-formed’ and ‘ill-formed’.
Empiricist language models often assign *probabilities* to sentences.

John went to the **pub** *John went to the* **?? cup**

Traditional grammar works with ‘cleaned-up’ sentences, with pauses, false starts, repetitions etc removed.

- Chomsky distinguished between syntactic **competence** and **performance**. He saw grammar as modelling competence.

Empiricist grammars tend to be trained on ‘real’ language data.

What sorts of pattern are found?

2. Patterns are often patterns in *surface language*.

In generative grammar, most of the rules are about deriving LF.

- There are no rules about the 'surface form' of a sentence.

However, in language, there appear to be lots of regularities which can *only* be expressed as surface regularities.

- The classic example is **idioms**.

Idioms

An idiom is an arbitrary sequence of words which collectively have an arbitrary semantic interpretation.

E.g. **by and large** (meaning 'typically').

- The meaning of this phrase doesn't come from the meanings of its individual words.
- It doesn't conform to any general syntactic rules.

Idioms often have 'slots', which can be filled by syntactically regular constituents.

- **Far be it from** *NP* **to** *VP*.

Idioms are often syntactically regular, even though their meaning is not compositional.

- *NP* **kicked the bucket**.

Idioms in the nativist-empiricist debate

Empiricist linguists argue that idioms are very common in language.

They argue that there's a *continuum* of idiomaticity.

- At one end, there are 'pure' idioms (e.g. *by and large*).
- In the middle there are idioms containing 'slots', and grammatically regular idioms.
- At the other end there are statistical tendencies.
  E.g. *went to the pub*, *give up*, *pull over*...

Empiricist models are well-suited for capturing idioms.

- Idioms are *statistical* regularities in *surface language*, mapped to arbitrary semantic/pragmatic patterns.

Minimalist models **have real difficulties with idioms**.

Idioms in a Minimalist model

If idioms are *continuous*, they can simply be treated as **multi-word lexical items**.

- E.g. *Winnie-the-Pooh*, *by-and-large*. . .

The difficulty is with *non-continuous idioms*, and with idioms which retain some degree of syntactic regularity.

- **Take** *NP* **to task** (= criticise NP)
- *NP* **let the cat out of the bag**
  The cat was let out of the bag *by NP*
  NP always lets the cat out of the bag

There's nothing in Minimalism which can explain these constructions.

Empiricist models of language acquisition

Empiricist models of language acquisition have an easier time explaining the different stages of syntactic development.

- Infants begin by detecting simple statistical regularities in surface language, and map these to semantic or pragmatic representations.
  - Holophrases
- Then they identify progressively more abstract regularities.
  - Pivot schemas
  - Item-based constructions
  - Fully abstract syntactic rules.

Minimalism is an account of ‘mature’ language competence; it’s not clear how this emerges during development.

Construction grammar

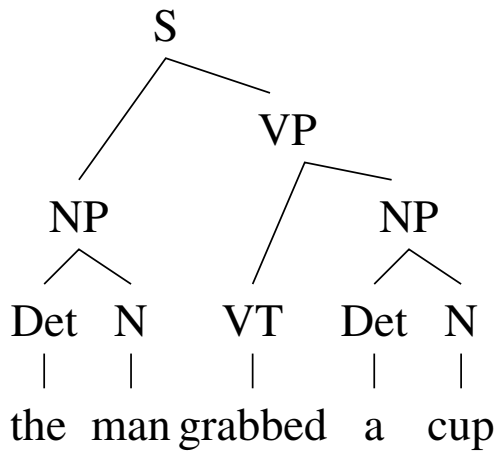
Construction grammar (Goldberg, 1995) is a theory of grammar embodying the empiricist model of language just given.

- Language is a collection of constructions.
- Each construction is an association between a pattern in language and a semantic (conceptual) pattern.
- The language patterns range from fully concrete to fully abstract.
  Concrete: patterns of particular words (e.g. *I want it*)
  Abstract: patterns of phrases (e.g. *NP VP*)
  In between: *I wanna VP*
- Good for representing structures like *I dunno*.

Construction grammar

Technically, how does construction grammar model patterns of phrases?
Constructivist models are based on the simple context-free grammars I introduced in Lecture 8.

S –––> NP, VP
NP –––> Det, N
VP –––> VT, NP
Det –––> the
Det –––> a
N –––> man
N –––> cup
VT –––> grabbed



They use **features** on phrases to model syntactic dependencies. (E.g. subject-verb agreement.)

Simple recurrent networks

Obviously, empiricists need to propose models of the **learning architectures** which infants are using to learn patterns in language.

- One of the key models is the **simple recurrent network** (**SRN**; Elman, 1990).

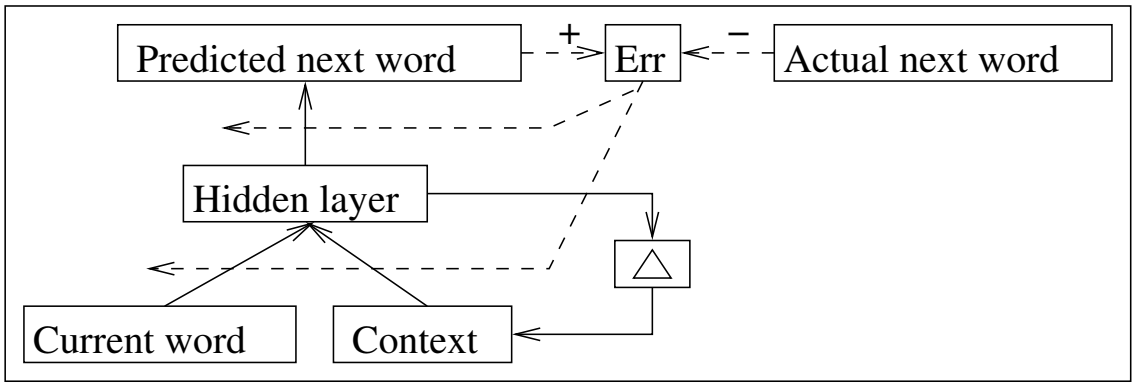
We've already seen SRNs in a model of word forms.

- There, the SRN takes as input a sequence of phonemes, and learns to predict the next phoneme.

We can also use a SRN in a model of syntax.

- Here, the SRN takes a sequence of whole words, and learns to predict the next word.

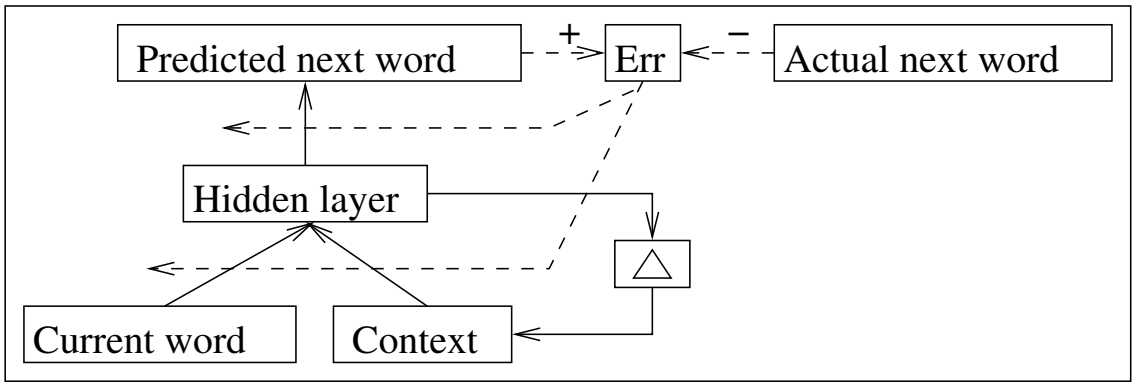
Simple recurrent networks



A SRN maintains a **context** representation, which is a copy of its hidden layer at the previous timestep.

- The context layer holds a history of recent inputs.
- After training, the context units can be interpreted as holding a representation of the most common sequences in the training data.

Simple recurrent networks



A trained SRN can't (normally) predict *exactly* which word will come next.

- It can distinguish between those words which are *likely* to come next, and those which are *unlikely*.
- It learns a (very simple) model of *syntax*.

Simple recurrent networks

Elman trained his SRN on simple sentences generated by a simple phrase-structure grammar.

For example:

girl/boy eat sandwich/cookie
girl/boy smash plate/glass

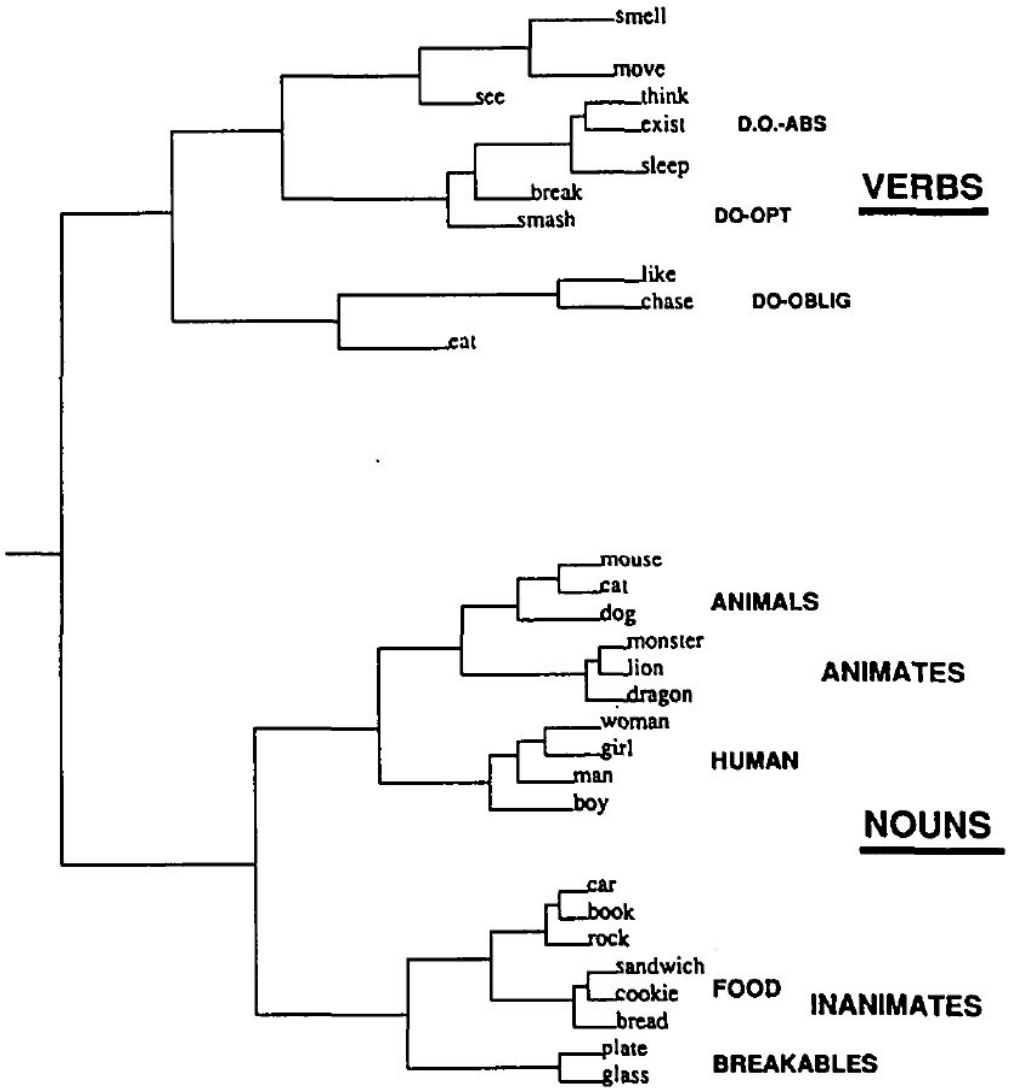
Simple recurrent networks

Interestingly, after training, words from the same syntactic (and even semantic) categories generate similar patterns of activation in the hidden layer of an SRN.

- This is because words from the same syntactic/semantic categories tend to occur in the same (surface) contexts.

Overleaf is a diagram showing how the activities hidden-unit word representations cluster after training.

Simple recurrent networks



SRNs and empiricist models of syntax learning

Before Elman's SRN, people assumed that grammatical word classes (e.g. 'noun', 'transitive verb' etc) probably had to be innate.
Elman showed that these classes could be induced by an SRN, simply through exposure to sentences from a natural language.

Summary

There's a big debate between nativist and empiricist syntacticians.

- Arguments for nativism:
  - Construction grammars are too simple: they miss important cross-linguistic generalisations.
  - 'Real syntax' is too complicated to be learnable from scratch.
  - Creoles show that children naturally invent complex natural languages.
- Arguments for empiricism:
  - The Minimalist model is far too complicated: its complex machinery can't be justfied from linguistic evidence.
  - The correct model of syntax must make reference to surface linguistic forms.
  - Children acquire syntax by progressively abstracting away from surface forms.

Summary

What the two theories have in common:

- At least some syntactic structures are defined using *phrases*, which abstract away from the details of specific words.
- Sentences often have hierarchical phrase structure.
- Actually, construction grammars often draw heavily on the idea of X-bar schemas.