Psych 215L: Language Acquisition

Lecture 16
Complex Systems

Computational Problem: Figure out the order of words (syntax)

Jareth juggles crystals
Subject Verb Object

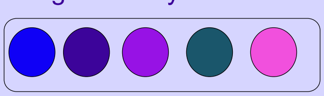
English
Subject Verb Object

German
Subject Verb *Subject* Object *Verb*

Kannada
Subject *Object* Verb Object

Remember:
Children only see the output of the system (the observable word order of Subject Verb Object) and have to reverse engineer the generative process behind it.

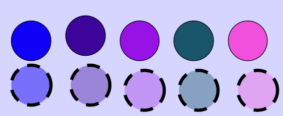
Thinking About Syntactic Variation



Similarities & Differences: Parameters

Chomsky: Different combinations of different basic elements (parameters) would yield the observable languages (similar to the way different combinations of different basic elements in chemistry yield many different-seeming substances).

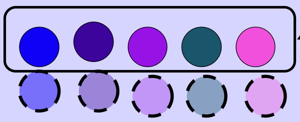
Big Idea: A relatively small number of syntax parameters yields a large number of different languages' syntactic systems.



Similarities & Differences: Parameters

Chomsky: Different combinations of different basic elements (parameters) would yield the observable languages (similar to the way different combinations of different basic elements in chemistry yield many different-seeming substances).

Big Idea: A relatively small number of syntax parameters yields a large number of different languages' syntactic systems.

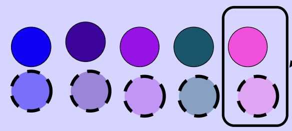


5 different parameters of variation

Similarities & Differences: Parameters

Chomsky: Different combinations of different basic elements (parameters) would yield the observable languages (similar to the way different combinations of different basic elements in chemistry yield many different-seeming substances).

Big Idea: A relatively small number of syntax parameters yields a large number of different languages' syntactic systems.

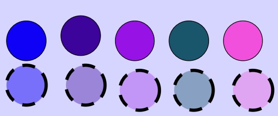


2 different parameter values of one parameter

Similarities & Differences: Parameters

Chomsky: Different combinations of different basic elements (parameters) would yield the observable languages (similar to the way different combinations of different basic elements in chemistry yield many different-seeming substances).

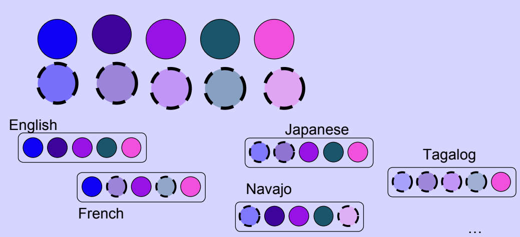
Big Idea: A relatively small number of syntax parameters yields a large number of different languages' syntactic systems.



Total languages that can be represented = $2^5 = 32$

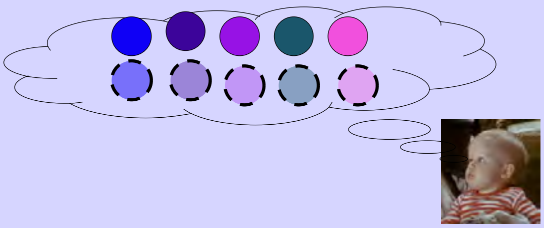
Similarities & Differences: Parameters

Big Idea: A relatively small number of syntax parameters yields a large number of different languages' syntactic systems.



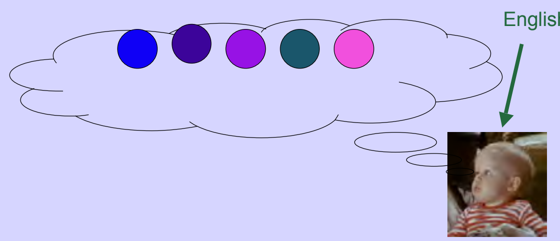
Learning Language Structure

Chomsky: Children are born knowing the parameters of variation (and also potentially what values that can have). This is part of Universal Grammar. Input from the native linguistic environment determines what values these parameters should have.



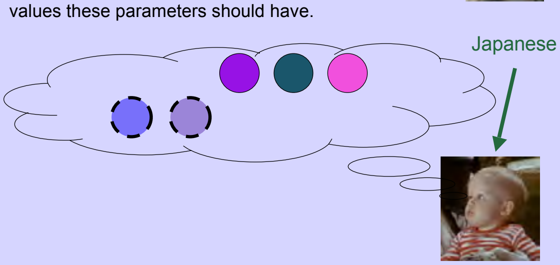
Learning Language Structure

Chomsky: Children are born knowing the parameters of variation (and also potentially what values that can have). This is part of Universal Grammar. Input from the native linguistic environment determines what values these parameters should have.



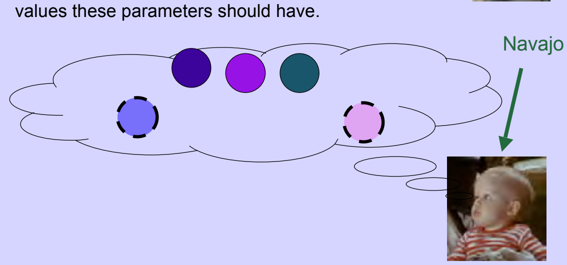
Learning Language Structure

Chomsky: Children are born knowing the parameters of variation (and also potentially what values that can have). This is part of Universal Grammar. Input from the native linguistic environment determines what values these parameters should have.



Learning Language Structure

Chomsky: Children are born knowing the parameters of variation (and also potentially what values that can have). This is part of Universal Grammar. Input from the native linguistic environment determines what values these parameters should have.



Yang (2004): Learning Complex Systems Like Language

Only humans seem able to learn human languages
Something in our biology must allow us to do this.

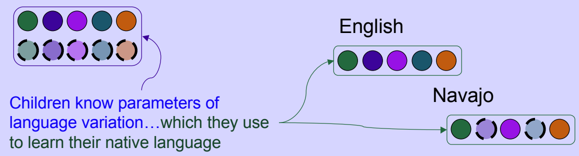
This is what Universal Grammar is: innate biases for learning language that are available to humans because of our biological makeup (specifically, the biology of our brains).

Chomsky

Yang (2004): Learning Complex Systems Like Language

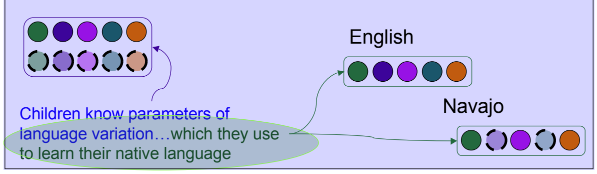
But obviously language is *learned*, so children can't know everything beforehand. How does this fit with the idea of innate biases/knowledge?

Observation: we see constrained variation across languages in their sounds, words, and structure. The knowledge of the ways in which languages vary is children's innate knowledge.



Yang (2004): Learning Complex Systems Like Language

The big point: even if children have innate knowledge of language structure, we still need to understand how they learn what the correct structural properties are for their particular language. One idea is to remember that children are good at tracking statistical information (like transitional probabilities) in the language data they hear.



Yang (2004): Learning Complex Systems

The linguist-psychologist breakdown

Linguists
Characterize "scope and limits of innate principles of Universal Grammar that govern the world's languages".

Noam Chomsky

David Lightfoot

Stephen Crain

Psychologists
Emphasize the "role of experience and the child's domain-general learning ability".

Michael Tomasello

Elizabeth Bates

Brian MacWhinney

Yang (2004): Learning Complex Systems

Statistics for word segmentation (remember Gambell & Yang (2006))

"Modeling shows that the statistical learning (Saffran et al. 1996) does not reliably segment words such as those in child-directed English. Specifically, precision is 41.6%, recall is 23.3%. In other words, about 60% of words postulated by the statistical learner are not English words, and almost 80% of actual English words are not extracted. This is so even under favorable learning conditions".

Unconstrained (simple) statistics: not so good.

If statistical measure is constrained by language-specific knowledge (words have only one main stress), performance increases dramatically: 73.5% precision, 71.2% recall.

Constrained statistics - much better!

Yang (2004): Learning Complex Systems

Combining statistics with Universal Grammar

A big deal:
"Although infants seem to keep track of statistical information, any conclusion drawn from such findings must presuppose that children know *what kind* of statistical information to keep track of."

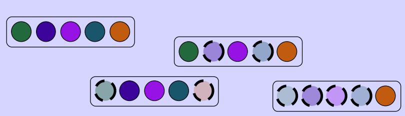
Ex: Transitional Probability

…of rhyming syllables?
…of syllables with nasal consonants?
…of syllables of the form CV (ba, ti)?

Linguistic Knowledge for Learning Structure

Parameters = constraints on language variation. Only certain rules/patterns are possible. This is linguistic knowledge.

A language's grammar
= combination of language rules
= combination of parameter values



Idea: use statistical learning to learn which value (for each parameter) that the native language uses for its grammar. This is a combination of using linguistic knowledge & statistical learning.

Yang (2004): Variational Learning

Idea taken from evolutionary biology:
In a population, individuals compete against each other. The fittest individuals survive while the others die out.

How do we translate this to learning language structure?

Yang (2004): Variational Learning

Idea taken from evolutionary biology:
In a population, individuals compete against each other. The fittest individuals survive while the others die out.

How do we translate this to learning language structure?

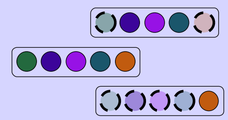
Individual = grammar (combination of parameter values that represents the structural properties of a language)

Fitness = how well a grammar can analyze the data the child encounters

Yang (2004): Variational Learning

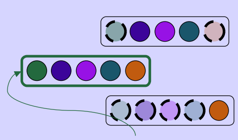
Idea taken from evolutionary biology:
A child's mind consists of a population of grammars that are competing to analyze the data in the child's native language.

Population of Grammars



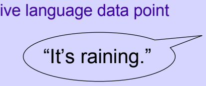
Yang (2004): Variational Learning

Intuition: The most successful (fittest) grammar will be the native language grammar because it can analyze all the data the child encounters. This grammar will "win", once the child encounters enough native language data because none of the other competing grammars can analyze all the data.



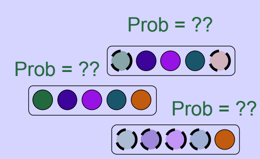
This grammar can analyze the data point while the other two can't.

Native language data point



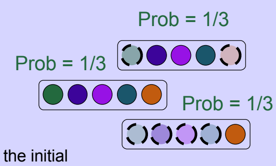
Variational Learning Details

At any point in time, a grammar in the population will have a probability associated with it. This represents the child's belief that this grammar is the correct grammar for the native language.



Variational Learning Details

Before the child has encountered any native language data, all grammars are equally likely. So, initially all grammars have the same probability, which is 1 divided the number of grammars available.

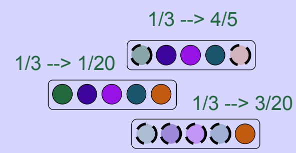


If there are 3 grammars, the initial probability for any given grammar = 1/3

Variational Learning Details

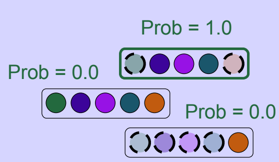
As the child encounters data from the native language, some of the grammars will be more fit because they are better able to account for the structural properties in the data.

Other grammars will be less fit because they cannot account for some of the data encountered. Grammars that are more compatible with the native language data will have their probabilities increased while grammars that are less compatible will have their probabilities decreased over time.



Variational Learning Details

After the child has encountered enough data from the native language, the native language grammar should have a probability near 1.0 while the other grammars have a probability near 0.0.

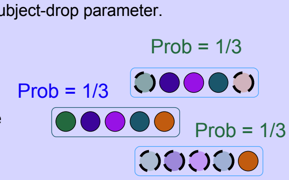


Variational Learning Details

How do we know if a grammar can successfully analyze a data point or not?

Example: Suppose is the subject-drop parameter.

is +subject-drop, which means the language may optionally choose to leave out the subject of the sentence, like in Spanish.



is -subject-drop, which means the language must always have a subject in a sentence, like English.

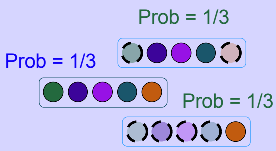
Here, one grammar is +subject-drop while two grammars are -subject-drop.

Variational Learning Details

How do we know if a grammar can successfully analyze a data point or not?

Example data: Vamos = *coming-1st-pl* = "We're coming"

The +subject-drop grammar is able to analyze this data point as the speaker optionally dropping the subject.



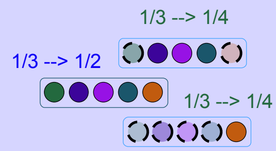
The -subject-drop grammars cannot analyze this data point since they require sentences to have a subject.

Variational Learning Details

How do we know if a grammar can successfully analyze a data point or not?

Example data: Vamos = *coming-1st-pl* = "We're coming"

The +subject-drop grammar would have its probability increased if it tried to analyze the data point.

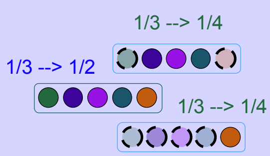


The -subject-drop grammars would have their probabilities decreased if either of them tried to analyze the data point.

Variational Learning Details

Important idea: From the perspective of the subject-drop parameter, certain data will only be compatible with +subject-drop grammars. These data will always reward grammars with +subject-drop and always punish grammars with -subject-drop.

Certain data always reward +subject-drop grammar(s).



Certain data always punish -subject-drop grammar(s).

These are called unambiguous data for the +subject-drop parameter value because they unambiguously indicate which parameter value is correct (here: +subject-drop) for the native language.

The Power of Unambiguous Data

Unambiguous data from the native language can only be analyzed by grammars that use the native language's parameter value.

This makes unambiguous data very influential data for the child to encounter, since it is incompatible with the parameter value that is incorrect for the native language.

Ex: the -subject-drop parameter value is not compatible with sentences that drop the subject. So, these sentences are unambiguous data for the +subject-drop parameter value.

Important to remember: To use the information in these data, the child must know the subject-drop parameter exists.

Yang (2004): Learning Complex Systems

Learning Parametric Systems: Variational Learning

Grammars compete against each other to see which can best analyze the available data.

Added perk: Learning is then gradual (probabilistic).

Problem: Do unambiguous data exist for entire grammars?
This requires data that are incompatible with every other possible parameter of every other possible grammar….

Yang (2004): Learning Complex Systems

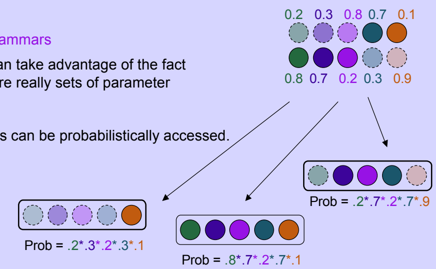
Learning Parametric Systems: Variational Learning

Grammars compete against each other to see which can best analyze the available data.

Parameterized Grammars

This algorithm can take advantage of the fact that grammars are really sets of parameter values.

Parameter values can be probabilistically accessed.



Yang (2004): Learning Complex Systems

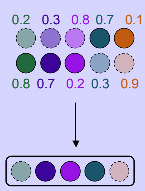
Learning Parametric Systems: Variational Learning

Grammars compete against each other to see which can best analyze the available data.

The Learning Algorithm

For each data point *d* encountered in the input

Choose a grammar probabilistically from available grammars by probabilistically accessing the parameter values.



Yang (2004): Learning Complex Systems

Learning Parametric Systems: Variational Learning

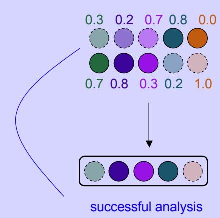
Grammars compete against each other to see which can best analyze the available data.

The Learning Algorithm

For each data point *d* encountered in the input

Choose a grammar probabilistically from available grammars by probabilistically accessing the parameter values.

If this grammar can analyze the data point, increase the probability of all participating parameters values slightly (reward)



Yang (2004): Learning Complex Systems

Learning Parametric Systems: Variational Learning

Grammars compete against each other to see which can best analyze the available data.

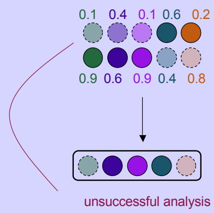
The Learning Algorithm

For each data point *d* encountered in the input

Choose a grammar probabilistically from available grammars by probabilistically accessing the parameter values.

If this grammar can analyze the data point, increase the probability of all participating parameters values slightly (reward)

Else
decrease the probability of all participating parameters values slightly (punish)



Yang (2004): Learning Complex Systems

Learning Parametric Systems: Variational Learning

Grammars compete against each other to see which can best analyze the available data.

Problem ameliorated: unambiguous data much more likely to exist for individual parameter values instead of entire grammars.

Yang (2004): Learning Complex Systems

Variational Learning: Sample Case

Null subjects:

Parameter 1: Pro-drop, rely on unambiguous subject-verb agreement
Ex: Spanish, Italian (+pro-drop) Ex: English (-pro-drop)

√	Yo puedo cantar. I can-1st-sg sing-inf 'I can sing'	√	I can sing
√	Puedo cantar. can-1st-sg sing-inf 'I can sing'	x	* Can sing
√	Hay lluvia. Is-3rd-sg rain "There is rain"	x	* Is rain
		√	There is rain.

Yang (2004): Learning Complex Systems

Variational Learning: Sample Case

Null subjects:

Parameter 1: Topic-drop, drop subject/object if discourse topic
Ex: Chinese (+topic-drop) Ex: English (-topic-drop)

(Topic = Jareth)

√	Mingtian guiji hui xiayu. Tomorrow estimate will rain 'It is tomorrow that *Jareth* believes it will rain'	x	*It is tomorrow that believes will rain.

Yang (2004): Learning Complex Systems

Variational Learning: Sample Case

Null subjects: 2 binary parameters, 4 grammars

+pro-drop, +topic-drop
Warlpiri, American Sign Language

+pro-drop, -topic-drop
Italian, Spanish

-pro-drop, +topic-drop
Chinese

-pro-drop, -topic-drop
English

What happens for an English-learning child?

Yang (2004): Learning Complex Systems

Variational Learning: Sample Case

Null subjects: 2 binary parameters, 4 grammars

+pro-drop, +to... drop
Warlpiri, A... Sign Language

+pro-dro... opic-drop
Italia... nish

-pro-drop, +topic-drop
Chinese

-pro-drop, -topic-drop
English

What happens for an English-learning child?

Pro-drop languages usually depend on rich subject-verb agreement morphology.
English doesn't have that, which is something a child will easily notice.
Knock out +pro-drop grammars.

Yang (2004): Learning Complex Systems

Variational Learning: Sample Case

Null subjects: 2 binary parameters, 4 grammars

+pro-drop, +to... drop
Warlpiri, A... Sign Language

+pro-dro... opic-drop
Italia... nish

-pro-drop, +topic-drop
Chinese

-pro-drop, -topic-drop
English

What happens for an English-learning child?

But this still leaves the +topic-drop option. What data will rule that out?

Answer: Expletive subjects. (Can't topic-drop them.)
"There's a goblin in the castle."
"It's raining outside."

But this only occurs in 1.2% of the data. (fairly rare)

Yang (2004): Learning Complex Systems

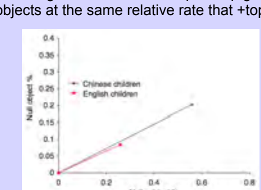
Variational Learning: Sample Case

Null subjects: Prediction if kids take awhile to notice English is -topic-drop

English kids use +topic-drop (Chinese-style) grammar until they encounter enough expletives to notice that English does not optionally drop topics.

Property of Chinese-style grammar: Can drop both subjects and objects

Prediction: When English children use +topic-drop grammar, they will drop subjects and objects at the same relative rate that +topic-drop (Chinese) children do



Same rate:
English children using
Chinese grammar?

Yang (2004): Learning Complex Systems

Variational Learning: General Predictions

The time course of when a parameter is set depends on how frequent the necessary evidence is in child-directed speech.

Parameters set early: more unambiguous data
Parameters set late: less unambiguous data
Parameters set at the same time: equal quantity of unambiguous data

Parameter	Target language	Requisite evidence	Input (%)	Time of acquisition
Wh fronting[a]	English	*Wh*-questions	25	very early [53]
verb raising[b]	French	verb adverb	7	1;8 [54]
obligatory subject	English	expletive subjects	1.2	3;0 [40,41]
verb second[c]	German/Dutch	OVS sentences [7,35]	1.2	3;0–3;2 [56]
scope marking[d]	English	long-distance *wh*-questions	0.2	4;0+ [56]

[a]English moves *Wh*-words in questions; in languages like Chinese, *Wh*-words stay put.
[b]In language like French, the finite verb moves past negation and adverbs ('*Jean voit souvent/pas Marie*'; 'Jean sees often/not Marie'), in contrast to English.
[c]In most Germanic languages, the finite verb takes the second position in the matrix clause, following one and exactly one phrase (of any type).
[d]In German, Hindi and other languages, long-distance *Wh*-questions leave intermediate copies of *Wh*-markers: '*Wer glaubst du wer Recht hat?*'; 'Who think you who right has?' (Who do you think has right?). For children to know that English doesn't use this the option, long-distance *Wh*-questions must be heard in the input. For many children, the German-option persists for quite some time, producing sentences like '*Who do you think who is the box?*' [56].

Additional Evidence for the importance of (un)ambiguity

Hadley, Rispoli, Fitzgerald, & Bahnsen (2010): input informativity (how much ambiguity in the input) is the most consistent predictor for morphosyntactic growth.

Pelham (2011): input ambiguity affects how children acquire pronoun forms ("It appears children may be sensitive to levels of ambiguity such that low ambiguity may aid error-free acquisition, while high ambiguity may blind children to case distinctions, resulting in errors.")

Another case study for variational learning

Explain why children's early output consistently contains "optional infinitives" (OIs) that are ungrammatical in the adult language. They produce these incorrect forms at the same time that they produce correct "finite" forms.

<u>English</u>

Correct:	"Mummy goes to work."
Occasional output:	"Mummy go to work"

Another case study for variational learning

Note: Not just a matter of shortening the word form – sometimes, the incorrect form is actually longer (French, Dutch). Also, the word order sometimes changes (Dutch). This seems likely to be the result of some process happening in the child's mind, rather than simple production error.

<u>French</u>

Input:	"La poupée dort."
	The doll sleep-3rd-sg
Occasional output:	"La poupée dormir"
	The doll sleep-inf

<u>Dutch</u>

Input:	"Ik eet ijs."
	I eat-3rd-sg ice cream
Occasional output:	"Ik ijs eten"
	I ice cream eat-inf

One explanation: Variational Learning Model

Legate & Yang (2007)

Grammar options: +Tense (English) vs. –Tense (Mandarin Chinese)

OI errors results because initial hypothesis is –Tense. This lessens over time when unambiguous +Tense data are observed.

+Tense unambiguous data: Morphological marking

he goes home

Prediction:

Morphologically rich languages like Spanish have a very short OI stage because a large proportion of the input rewards +Tense (and punishes –Tense).

Morphologically poor languages like English have a longer OI stage because only a small proportion of the input rewards the [+Tense] grammar (and punishes –Tense).

One explanation: Variational Learning Model

Legate & Yang (2007)

Languages tested:

English, French, Spanish

Observed behavior seems to match unambiguous input distributions

OI duration:

English (high) > French (moderately high) >> Spanish (very low)

+Tense unambiguous data:

English > French >> Spanish

Possible critique (from Freudenthal et al. 2010)

Too easy because rates of OI are very different. What about Dutch and German, who have OI rates that are moderately high?

Another explanation: MOSAIC model

Freudenthal et al. (2010)

Model of Syntax Acquisition in Children: "MOSAIC is a constructivist model of language learning, with no built-in knowledge of syntactic categories or rules, which is implemented as a working computational model." – Algorithmic level?

"MOSAIC takes as input corpora of child- directed speech and learns to produce as output 'child-like' utterances that become progressively longer as learning proceeds…input corpora are fed through the model multiple times."

Fig. 1. A sample MOSAIC network that has learned an utterance-initial and utterance-final phrase.

Input:
"He will"
"He wants"
"Go home"
"Go away"

Another explanation: MOSAIC model

Freudenthal et al. (2010)

- has a strong utterance-final bias in learning

"MOSAIC does not encode a word or phrase unless everything that follows that phrase has already been encoded in the network."

- has a weak utterance-initial bias in learning

"The utterance-initial bias enables MOSAIC to associate utterance-initial words and short (frequent) phrases with (longer) utterance-final phrases."

- represents declaratives and questions separately

(so no underlying linkage between these forms)

Who could you see? has no relation to You could see him.

Another explanation: MOSAIC model

Freudenthal et al. (2010)
Where OI errors come from: Compound finites

English:

He can go home.

Another explanation: MOSAIC model

Freudenthal et al. (2010)
Where OI errors come from: Compound finites

English:

~~He can~~ go home. → "Go home"
utterance-final bias

Another explanation: MOSAIC model

Freudenthal et al. (2010)
Where OI errors come from: Compound finites

English:

He ~~can~~ go home. → "Go home", "He go home"
utterance-final bias
+ weak utterance-initial bias + linking

Another explanation: MOSAIC model

Freudenthal et al. (2010)
Where OI errors come from: Compound finites

English:

He can go home. → "Go home", "He go home"
utterance-final bias
+ weak utterance-initial bias + linking

Dutch (+ changed word order):

Hij wil ijs eten. →
He wants ice cream eat-inf
"He wants to eat ice cream."

Another explanation: MOSAIC model

Freudenthal et al. (2010)
Where OI errors come from: Compound finites

English:

He can go home. → "Go home", "He go home"
utterance-final bias
+ weak utterance-initial bias + linking

Dutch (+ changed word order):

~~*Hij wil*~~ *ijs* *eten.* → "Ijs eten"
He wants ice cream eat-inf
"He wants to eat ice cream."
utterance final bias

Another explanation: MOSAIC model

Freudenthal et al. (2010)
Where OI errors come from: Compound finites

English:

He can go home. → "Go home", "He go home"
utterance-final bias
+ weak utterance-initial bias + linking

Dutch (+ changed word order):

Hij ~~*wil*~~ *ijs* *eten.* → "Ijs eten", "*Hij ijs eten*"
He wants ice cream eat-inf
"He wants to eat ice cream."
utterance final bias
+weak utterance initial bias + linking

Freudenthal et al. (2010) Concluding Thoughts

"…it is clear that both the VLM and MOSAIC do a relatively good job of predicting the cross-linguistic data…if we focus on the results of the second set of analyses, it is clear that there are important lexical effects on the distribution of OI errors in children's speech that are difficult for the VLM to explain…"

"…A more lexically oriented input-driven account could probably deal with this problem relatively easily by simply distinguishing between what the child is learning about copulas and auxiliaries and what the child is learning about lexical verbs, and predicting high levels of OI errors on lexical verbs and lower levels of OI errors on copulas and auxiliaries. Interestingly, this is exactly the pattern of results reported in two recent lexically oriented analyses of early child English (Wilson, 2003; Pine, Conti-Ramsden, Joseph, Lieven & Serratrice, 2008)."