Psych 215L: Language Acquisition

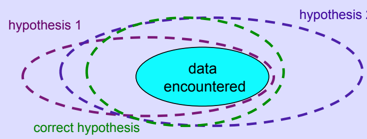
Lecture 17
Poverty of the Stimulus: Syntactic Islands

An induction problem by any other name…

One of the most controversial claims in linguistics is that children face an induction problem:

"Poverty of the Stimulus" (Chomsky 1980, Crain 1991, Lightfoot 1989, Valian 2009)
"Logical Problem of Language Acquisition" (Baker 1981, Hornstein & Lightfoot 1981)
"Plato's Problem" (Chomsky 1988, Dresher 2003)

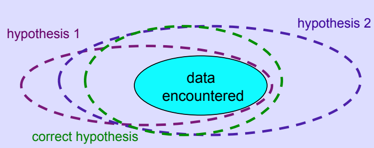
Basic claim:
The data encountered are compatible with multiple hypotheses.



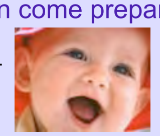
The induction problem

Extended claim:
Given this, the data are insufficient for identifying the correct hypothesis as quickly as children do (Legate & Yang 2002) – or at all.

Big question: How do children do it, then?



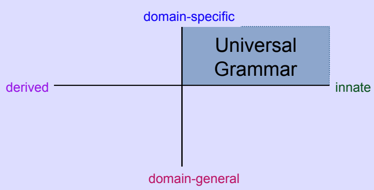
One answer: Children come prepared



- Children are not unbiased learners.
- But if children come equipped with helpful learning biases, then what is the nature of these necessary biases?
  - Are they innate or derived from the input somehow?
  - Are they domain-specific or domain-general?
  - Are they about the hypothesis space or about the learning mechanism?

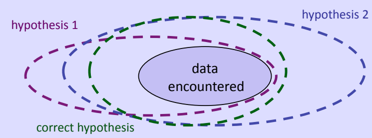
Informing Theory: Arguments from Acquisition

Universal Grammar consists of the necessary learning biases that are both innate and domain-specific (Chomsky 1965, Chomsky 1975).



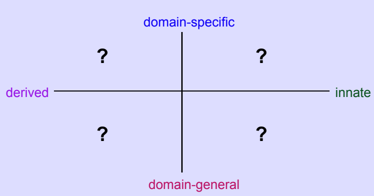
Informing Theory: Arguments from Acquisition

One explicit motivation for Universal Grammar is that it explains how children solve the induction problem inherent in language acquisition.



Informing Theory: Arguments from Acquisition

Open question: For any given piece of linguistic knowledge, what biases are necessary to learn it from child-directed data? Are any of them necessarily both innate and domain-specific?



The Plan

(1) Look at syntactic islands: phenomena central to UG-based syntactic theories.

(2) Explicitly define the target knowledge state, based on adult acceptability judgments.

(3) Identify the kind of data children and adults have in their input, using realistic samples of child-directed and adult-directed input.

(4) Implement a computational learner that is able to reach the target knowledge state, given realistic data distributions, and see what kind of learning biases it requires.

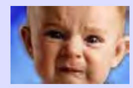
Syntactic Islands

Dependencies can exist between two non-adjacent items, and these do not appear to be constrained by length (Chomsky 1965, Ross 1967).

What does Jack think __?
What does Jack think that Lily said __?
What does Jack think that Lily said that Sarah heard __?
What does Jack think that Lily said that Sarah heard that Jareth stole __?

Syntactic Islands

However, if the gap position appears inside certain structures (called "syntactic islands" by Ross (1967)), the dependency seems to be ungrammatical.



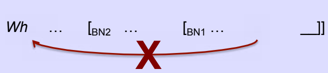
*What did you make [the claim that Jack bought __]? *[Complex NP]*
*What do you think [the joke about __] offended Jack? *[Subject]*
*What do you wonder [whether Jack bought __]? *[Whether]*
*What do you worry [if Jack buys __]? *[Adjunct]*
*What did you meet [the scientist who invented __]?
*What did [that Jack wrote __] offend the editor?
*What did Jack buy [a book and __]?
*Which did Jack borrow [__ book]?

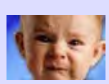
Syntactic Islands

- **Predominant theory in generative syntax:**
  syntactic islands require innate, domain-specific learning biases

Example: Subjacency **(Chomsky 1973, Huang 1982, Lasnik & Saito 1984)**
A dependency cannot cross two or more bounding nodes.

Bounding nodes: language-specific (CP, IP, and/or NP)

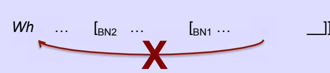


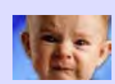


Syntactic Islands

- **Predominant theory in generative syntax:**
  syntactic islands require innate, domain-specific learning biases

Subjacency learning biases:
(1) Innate, domain-specific knowledge of hypothesis space: Exclude hypotheses that allow dependencies crossing 2+ bounding nodes.



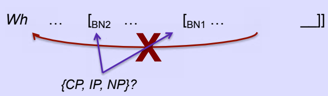


Syntactic Islands

- **Predominant theory in generative syntax:** syntactic islands require innate, domain-specific learning biases

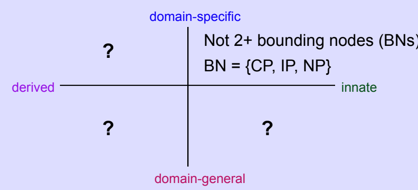
Subjacency learning biases:
(1) Innate, domain-specific knowledge of hypothesis space: Exclude hypotheses that allow dependencies crossing 2+ bounding nodes.

(2) Innate, domain-specific knowledge of hypothesis space: Hypothesis space consists of bounding nodes for all languages, and the child must identify the ones applicable to his language.



Syntactic Islands

- **Predominant theory in generative syntax:** syntactic islands require innate, domain-specific learning biases



The target state: Adult knowledge of syntactic islands

Sprouse et al. (2012) collected magnitude estimation judgments for four different islands:

Complex NP islands
*What did you make [the claim that Jack bought __]?

Subject islands
*What do you think [the joke about __] offended Jack?

Whether islands
*What do you wonder [whether Jack bought __]?

Adjunct islands
*What do you worry [if Jack buys __]?

The target state: Adult knowledge of syntactic islands

Sprouse et al. (2012)'s factorial definition controls for two salient properties of island-crossing dependencies:
- length of dependency (short vs. long)
- presence of an island structure (non-island vs. island)

The target state: Adult knowledge of syntactic islands

Sprouse et al. (2012)'s factorial definition controls for two salient properties of island-crossing dependencies:
- length of dependency (short vs. long)
- presence of an island structure (non-island vs. island)

Complex NP islands

Who __ claimed that Lily forgot the necklace?	matrix | non-island
What did the teacher claim that Lily forgot __?	embedded | non-island
Who __ made the claim that Lily forgot the necklace?	matrix | island
*What did the teacher make the claim that Lily forgot __?	embedded | island

The target state: Adult knowledge of syntactic islands

Sprouse et al. (2012)'s factorial definition controls for two salient properties of island-crossing dependencies:
- length of dependency (short vs. long)
- presence of an island structure (non-island vs. island)

Subject islands

Who __ thinks the necklace is expensive?	matrix | non-island
What does Jack think __ is expensive?	embedded | non-island
Who __ thinks the necklace for Lily is expensive?	matrix | island
*Who does Jack think the necklace for __ is expensive?	embedded | island

The target state: Adult knowledge of syntactic islands

Sprouse et al. (2012)'s factorial definition controls for two salient properties of island-crossing dependencies:
- length of dependency (short vs. long)
- presence of an island structure (non-island vs. island)

Whether islands

Who __ thinks that Jack stole the necklace?	matrix | non-island
What does the teacher think that Jack stole __ ?	embedded | non-island
Who __ wonders whether Jack stole the necklace?	matrix | island
*What does the teacher wonder whether Jack stole __ ?	embedded | island

The target state: Adult knowledge of syntactic islands

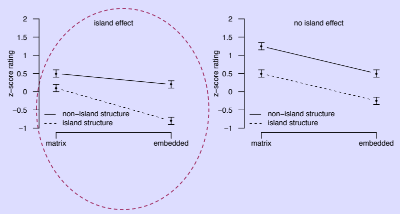
Sprouse et al. (2012)'s factorial definition controls for two salient properties of island-crossing dependencies:
- length of dependency (short vs. long)
- presence of an island structure (non-island vs. island)

Adjunct islands

Who __ thinks that Lily forgot the necklace?	matrix | non-island
What does the teacher think that Lily forgot __ ?	embedded | non-island
Who __ worries if Lily forgot the necklace?	matrix | island
*What does the teacher worry if Lily forgot __ ?	embedded | island

The target state: Adult knowledge of syntactic islands

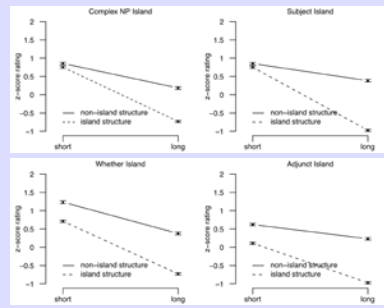
Syntactic island = superadditive interaction of the two factors (additional unacceptability that arises when the two factors are combined, above and beyond the independent contribution of each factor).



The target state: Adult knowledge of syntactic islands

Sprouse et al. (2012)'s data on the four island types (173 subjects)

Superadditivity present for all islands tested
=
Knowledge that dependencies cannot cross these island structures is part of the adult knowledge state



The input: Assessing the induction problem

Data from five corpora of child-directed speech (Brown-Adam, Brown-Eve, Brown-Sarah, Suppes, Valian) from CHILDES (MacWhinney 2000): speech to 25 children between the ages of one and five years old.

Total words: 813,036

Utterances containing a *wh*-dependency: 31,247

Sprouse et al. (2012) stimuli types: *ungrammatical*

	MATRIX + NON-ISLAND	EMBEDDED + NON-ISLAND	MATRIX + ISLAND	*EMBEDDED + ISLAND*
Complex NP	7	295	0	0
Subject	7	29	0	0
Whether	7	295	0	0
Adjunct	7	295	15	0

The input: Assessing the induction problem

These kinds of utterances are fairly rare in general - the most frequent appears about 0.9% of the time (**295** of **31,247**.)

Sprouse et al. (2012) stimuli types: *ungrammatical*

	MATRIX + NON-ISLAND	EMBEDDED + NON-ISLAND	MATRIX + ISLAND	*EMBEDDED + ISLAND*
Complex NP	7	295	0	0
Subject	7	29	0	0
Whether	7	295	0	0
Adjunct	7	295	15	0

The input: Assessing the induction problem

Being grammatical doesn't necessarily mean an utterance will appear in the input at all.

Sprouse et al. (2012) stimuli types:

ungrammatical

	MATRIX + NON-ISLAND	EMBEDDED + NON-ISLAND	MATRIX + ISLAND	*EMBEDDED + ISLAND*
Complex NP	7	295	0	0
Subject	7	29	0	0
Whether	7	295	0	0
Adjunct	7	295	15	0

The input: Assessing the induction problem

Unless the child is sensitive to very small frequencies, it's difficult to tell the difference between grammatical and ungrammatical dependencies sometimes…

Sprouse et al. (2012) stimuli types:

ungrammatical

	MATRIX + NON-ISLAND	EMBEDDED + NON-ISLAND	MATRIX + ISLAND	*EMBEDDED + ISLAND*
Complex NP	7	295	0	0
Subject	7	29	0	0
Whether	7	295	0	0
Adjunct	7	295	15	0

The input: Assessing the induction problem

…and impossible to tell no matter what the rest of the time. This looks like an induction problem for the language learner if we're looking for direct evidence in the input.

Sprouse et al. (2012) stimuli types:

ungrammatical

	MATRIX + NON-ISLAND	EMBEDDED + NON-ISLAND	MATRIX + ISLAND	*EMBEDDED + ISLAND*
Complex NP	7	295	0	0
Subject	7	29	0	0
Whether	7	295	0	0
Adjunct	7	295	15	0

Building a computational learner: Proposed learning biases

Learning Bias: Children track the occurrence of structures that can be derived from phrase structure trees - container nodes.

[$_{CP}$ Who did [$_{IP}$ she [$_{VP}$ like __]]]?
IP VP

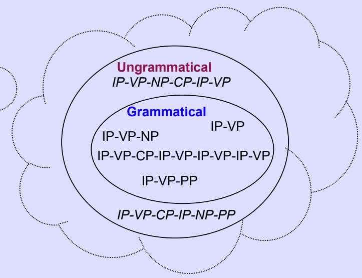
Container node sequence: IP-VP

[$_{CP}$ Who did [$_{IP}$ she [$_{VP}$ think [$_{CP}$ [$_{IP}$ [$_{NP}$ the gift] [$_{VP}$ was [$_{PP}$ from __]]]]]]]?
IP VP CP IP VP PP

Container node sequence: IP-VP-CP-IP-VP-PP

Building a computational learner: Proposed learning biases

Learning Bias: Children's hypotheses are about what container node sequences are grammatical for dependencies in the language.



What does the target knowledge look like?

Sprouse et al. (2012) stimuli:

Complex NP islands		Subject islands
IP	matrix | non-island	IP
IP-VP-CP-IP-VP	embedded | non-island	IP-VP-CP-IP
IP	matrix | island	IP
*IP-VP-NP-CP-IP-VP	embedded | island	*IP-VP-CP-IP-NP-PP

All the ungrammatical dependencies are distinct from all the grammatical dependencies for these syntactic islands.

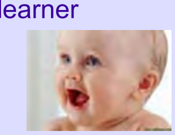
What does the target knowledge look like?

Sprouse et al. (2012) stimuli:

Whether islands		Adjunct islands
IP	matrix | non-island	IP
IP-VP-CP-IP-VP	embedded | non-island	IP-VP-CP-IP-VP
IP	matrix | island	IP
*IP-VP-CP-IP-VP	embedded | island	*IP-VP-CP-IP-VP

Uh oh - the ungrammatical dependencies look identical to some of the grammatical dependencies for these syntactic islands.

Building a computational learner



Learning bias solution:
Have CP container nodes be more specified for the learner: Use the lexical head to subcategorize the CP container node.

CP_{null}, CP_{that}, $CP_{whether}$, CP_{if}, etc.

The learner can then distinguish between these structures:

IP-VP-$CP_{null/that}$-IP-VP
IP-VP-$CP_{whether/if}$-IP-VP

What does the target knowledge look like?

Sprouse et al. (2012) stimuli:

Whether islands		Adjunct islands
IP	matrix | non-island	IP
IP-VP-CP_{that}-IP-VP	embedded | non-island	IP-VP-CP_{that}-IP-VP
IP	matrix | island	IP
*IP-VP-CP-$IP_{whether}$-VP	embedded | island	*IP-VP-CP_{if}-IP-VP

Now the ungrammatical dependencies are distinct from all the grammatical dependencies for these syntactic islands, too.

Building a computational learner: Proposed learning biases

Learning Bias: Implicitly assign a probability to a container node sequence by tracking trigrams of container nodes. A sequence's probability is the smoothed product of its trigrams.

$[_{CP}$ Who did $[_{IP}$ she $[_{VP}$ like __]]]?
IP VP
start-IP-VP-end =
start-IP-VP
IP-VP-end

Probability(IP-VP) = p(start-IP-VP-end)
= p(start-IP-VP) * p(IP-VP-end)

Building a computational learner: Proposed learning biases

Learning Bias: Implicitly assign a probability to a container node sequence by tracking trigrams of container nodes. A sequence's probability is the smoothed product of its trigrams.

$[_{CP}$ Who did $[_{IP}$ she $[_{VP}$ think $[_{CP}$ $[_{IP}$ $[_{NP}$ the gift] $[_{VP}$ was $[_{PP}$ from __]]]]]]]?
IP VP CP_{null} IP VP PP
start-IP-VP-CP_{null}-IP-VP-PP-end =
start-IP-VP
IP-VP-CP_{null}
VP-CP_{null}-IP
CP_{null}-IP-VP
IP-VP-PP
VP-PP-end

Probability(IP-VP-CP_{null}-IP-VP-PP) = p(start-IP-VP-CP_{null}-IP-VP-PP-end)
= p(start-IP-VP) * p(IP-VP-CP_{null})*p(VP-CP_{null}-IP)*p(CP_{null}-IP-VP)
*p(IP-VP-PP)*p(VP-PP-end)

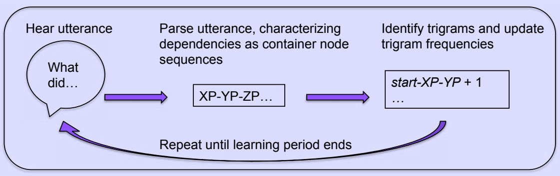
Building a computational learner: Proposed learning biases

Learning Bias: Implicitly assign a probability to a container node sequence by tracking trigrams of container nodes. A sequence's probability is the smoothed product of its trigrams.

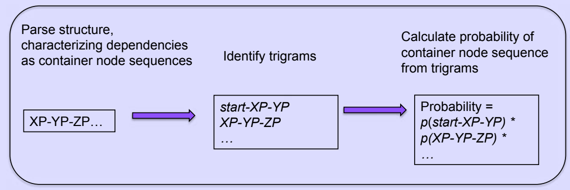
What this does:
- longer dependencies are less probable than shorter dependencies, all other things being equal
- individual trigram frequency matters: short dependencies made of infrequent trigrams will be less probably than longer dependencies made of frequent trigrams

Effect: the frequencies observed in the input temper the detrimental effect of dependency length.

Learning process



Generating grammaticality preferences



Building a computational learner: Empirical grounding

Child-directed speech from CHILDES: If we want to model child learners.

Adult-directed speech (Treebank-3-Switchboard corpus: Marcus et al. 1999) and text (Treebank-3-Brown corpus: Marcus et al. 1999): If we want to model adult learners, since we have adult data.

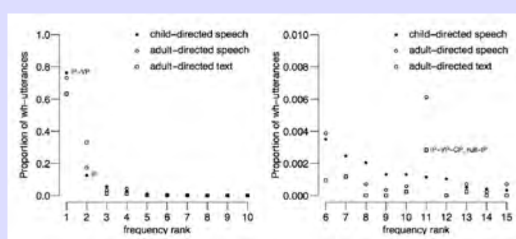
Table 3: Basic composition of the child-directed and adult-directed input corpora.

	Child-directed: speech	Adult-directed: speech	Adult-directed: text
total utterances	101838	74576	24243
total *wh*-dependencies	20923	8508	4230

Building a computational learner: Empirical grounding

Child-directed speech from CHILDES: If we want to model child learners.

Adult-directed speech (Treebank-3-Switchboard corpus: Marcus et al. 1999) and text (Treebank-3-Brown corpus: Marcus et al. 1999): If we want to model adult learners, since we have adult data.



Building a computational learner: Empirical grounding

Child-directed speech from CHILDES:

What kind of dependencies are present?

76.7%	IP-VP	*What did you see __?*
12.8%	IP	*What __ happened?*
5.6%	IP-VP-IP-VP	*What did she want to do __?*
2.5%	IP-VP-PP	*What did she read from __?*
1.1%	IP-VP-CP_{null}-IP-VP	*What did she think he said __?*
…		

Building a computational learner: Empirical grounding

Hart & Risley 1995: Children hear approximately 1 million utterances in their first three years.

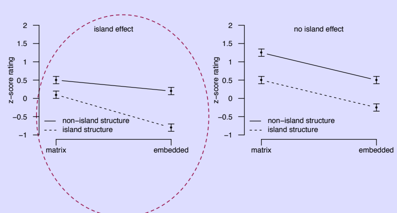
Assumption: learning period for modeled learners is 3 years (ex: between 2 and 5 years old for modeling children's acquisition)

Estimating proportion of *wh*-dependencies in the input, based on child-directed speech sample: total learning period is 200,000 *wh*-dependency data points, drawn from distribution observed in speech and/or text samples.

Success metrics

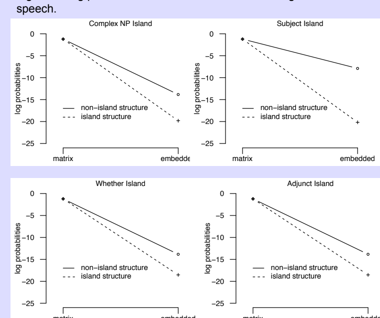
Compare learned grammaticality preferences to Sprouse et al. (2012) judgment data.

Then, for each island, we plot the predicted grammaticality preferences from the modeled learner on an interaction plot, using log probability of the dependency on the y-axis. Non-parallel lines indicate knowledge of islands.



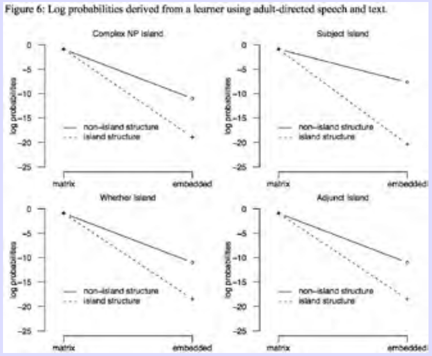
Learning results

Figure 5: Log probabilities derived from a learner using child-directed speech.



Learning results

Figure 6: Log probabilities derived from a learner using adult-directed speech and text.



Learning results

“All of the island violations are at least 4 times less acceptable than the grammatical control conditions, and often more than 10 times less acceptable.”

Why it works

“Crucially, for each of the island-spanning dependencies, there is at least one extremely low probability container node trigram in the container node sequence of the dependency. These trigrams are assigned low probabilities because these trigram sequences are never observed in the input – it is only the smoothing parameter that prevents these probabilities from being 0. Note that some trigrams are low probability due to being rarely encountered in the input (e.g., CP_{that}-IP-VP in child-directed speech) – but, crucially, this is still more than never. Even though CP_{that} rarely appears, it *does* appear, and so it is assigned a probability that is substantially non-zero.”

Learning biases: Discussion

Phrase structure

“One of the most basic components of the proposed learning algorithm is that it operates over input that has been parsed into phrase structure trees. It therefore assumes that both syntactic category information and phrase structure information have already been acquired (or are in the process of being acquired). We do not have too much to say about this assumption because basic syntactic phenomena like syntactic categories and phrase structure parsing are required by nearly every syntactic phenomenon…That being said, for recent work investigating the acquisition of syntactic categories from child-directed input, see Mintz (2003) and (2006), and for recent work investigating the acquisition of hierarchical structure given syntactic categories as input, see Klein & Manning (2002).”

Learning biases: Discussion

Tracking frequencies and calculating probabilities

"Another basic component of the proposed algorithm is that the learner has the ability to track the frequency of units in the input, and then calculate the probabilities of those units. This is a relatively uncontroversial assumption, as many learning theories, both in language and other cognitive domains, assume that the learner can track frequencies and calculate probabilities. The ability to track frequencies and calculate probabilities is likely an innate, domain-general ability."

Learning biases: Discussion

Restricting the input to *wh*-dependencies

"The proposed algorithm assumes that only *wh*-dependencies are used as input by the learner, at least for the acquisition of syntactic island effects with *wh*-dependencies. This assumption is not as neutral as it first appears. First, many syntactic theories recognize similarities between *wh*-dependencies and other types of dependencies, such as relative-clause-dependencies (rc- dependencies)…[but] these two dependencies must be tracked separately for the purposes of acquisition…Second, other dependencies, such as the binding dependencies that hold between nouns and pronouns, do not display syntactic island effects at all…suggest that binding dependencies must be tracked separately from *wh*- dependencies."

Learning biases: Discussion

Restricting the input to *wh*-dependencies

"The fact that it is empirically necessary to separate *wh*-dependencies from other dependency types does not explain how it is that the acquisition system knows to separate the input. While it is logically possible to achieve this type of separation without necessarily invoking innate, domain-specific biases, we simply do not have enough information about the learnability of these other dependency types to evaluate the possibilities. What we can say is that the learning strategy proposed here highlights the fact that the any theory of the acquisition of syntactic islands must be able to track *wh*-dependencies separately from *rc-dependencies* and binding dependencies."

Learning biases: Discussion

Tracking sequences of container nodes

"…appears relatively neutral at first glance; after all, syntactic island effects are constraints on dependencies, and therefore the algorithm should track information about the dependencies. However, this assumption is far from neutral, as it is in essence informing the system that long-distance dependencies may have constraints on them and so information about them should be tracked."

Learning biases: Discussion

Tracking sequences of container nodes

"…how it is that the algorithm knows to track container nodes rather than some other piece of information about a dependency….the fact that the parsing of long-distance dependencies is an active process means that the sequence of container nodes is information that is likely available to the language system, but *availability* is distinct from *attention*….This bias is likely domain-specific, as long-distance dependencies (and their constraints) have not been clearly demonstrated in any other domain of cognition. It is, however, an open question whether this bias is also innate, or whether it can be derived from other biases."

Learning biases: Discussion

Tracking trigrams

Not unigrams

"A unigram model will successfully learn Whether and Adjunct islands, as there are container nodes in these dependencies that never appear in grammatical dependencies ($CP_{whether}$ and CP_{if}), but will fail to learn Complex NP and Subject islands, as all of the container nodes in these islands are shared with grammatical dependencies."

Complex NP: *IP-VP-NP-CP_{that}-IP-VP
Subject: *IP-VP-CP_{null}-IP-NP-PP
Whether: IP-VP-$CP_{whether}$-IP-VP
Adjunct: IP-VP-CP_{if}-IP-VP

Learning biases: Discussion

Tracking trigrams

Not bigrams

"A similar problem arises for a bigram model: At least for Subject islands, there is no bigram that occurs in a Subject island violation but not in any grammatical dependencies. The most likely candidate for such a bigram is IP-NP…However, sentences such as *What, again, about Jack impresses you?* or *What did you say about the movie scared you?* suggest that a gap can arise inside of NPs, as long as the extraction is of the head noun (what), not of the noun complement of the preposition.."

Complex NP: IP-VP-NP-CP_{that}-IP-VP
Subject: *IP-VP-CP_{null}-IP-NP-PP
Whether: IP-VP-$CP_{whether}$-IP-VP
Adjunct: IP-VP-CP_{if}-IP-VP

Learning biases: Discussion

Tracking trigrams

Not n-grams, where n>3

"…there is no straightforward way to accommodate extraction from matrix subject position, which only results in a single container node (IP). It is possible to accommodate these sequences in a trigram model by assuming symbols for *start* and *end*, resulting in *start-IP-end*. *Start* and *end* symbols may not be part of phrase structure grammars, but they are at least psychologically principled in that the algorithm needs to track the beginning and end of dependencies at some level. However there is no obviously principled way to incorporate an additional symbol in a 4-gram model to capture matrix subject dependencies. This suggests that a trigram model is simpler because the 4-gram model will require an exception for these dependencies. This problem holds for every *n-gram* above 3."

Learning biases: Discussion

Tracking trigrams

"…it is an open question how this bias arises. Learning models based on sequences of three units have been proposed and are consistent with children's observable behavior for other linguistic knowledge… additionally, these learning models are consistent with human behavior for non-linguistic phenomena (Saffran et al. 1996) and also with learning behavior in non-human primates (Saffran et al. 2008). Given this, such a bias is likely domain-general; however, the fact that trigrams are an available option does not explain how it is that the learning algorithm knows to leverage trigrams (as opposed to other n-grams) for syntactic islands."

Learning biases: Discussion

Subcategorizing CPs

"The fact that CPs *can* be subcategorized is relatively straightforward… However, the fact that this type of information is available to the language system does not explain how it is that the learner knows to pursue this particular strategy (or knows where to draw the line between types of container nodes). It may be possible to capture part of this behavior with innate, domain-general preferences for certain types of hypotheses (either more specific hypotheses, such as subcategorize all container nodes, or more general hypotheses, such as subcategorize no container nodes) coupled with a domain-specific proposal about the types of information that could be used to correct mistaken hypotheses."

Learning biases: Discussion

So what do we know now that we didn't know before?

"Some of the basic components of the algorithm will be part of the learning theory for any syntactic phenomenon (e.g., assigning phrase structure and tracking frequencies), but others appear to be specific to syntactic island effects, such as restricting the input to *wh*-dependencies, tracking sequences of container nodes, segmenting container node sequences into trigrams, and subcategorizing CP container nodes by the lexical item that introduces them. These biases are interesting because on the one hand, they are significantly less specific than previous approaches to the acquisition of island effects (which tended to directly encode syntactic constraints in the learning algorithm); on the other hand, they are still specific enough to raise difficult questions about how they could arise in the learner. The explicit modeling procedure here (based on realistic input) suggests that any theory that seeks to learn syntactic islands as a type of grammatical constraint will be forced grapple with the empirical necessity of these specific biases."

Parasitic gaps

This learner can't handle parasitic gaps, which are dependencies that span an island (and so should be ungrammatical) but which are somehow rescued by another dependency in the utterance.

*Which book did you laugh [before reading __]?
Which book did you judge $\_\__{true}$ [before reading $\_\__{parasitic}$]?

Adjunct island

*What did [the attempt to repair __] ultimately damage the car?
What did [the attempt to repair $\_\__{parasitic}$] ultimately damage $\_\__{true}$?

Complex NP island

Parasitic gaps

Why not? The current learner would judge the parasitic gap as ungrammatical since it is inside an island, irrespective of what other dependencies are in the utterance.

*Which book did you laugh [before reading __]?
Which book did you judge $\_\__{true}$ [before reading $\_\__{parasitic}$]?

Adjunct island

*What did [the attempt to repair __] ultimately damage the car?
What did [the attempt to repair $\_\__{parasitic}$] ultimately damage $\_\__{true}$?

Complex NP island

This may be able to be addressed in a learner that is able to combine information from multiple dependencies in an utterance (perhaps because the learner has observed multiple dependencies resolved in utterances in the input).

Across-the-board constrcutions

A similar problem occurs for across-the-board constructions.

Which book did you [[read __] and [then review __]]?
dependency for both gaps: IP-VP-VP

*Which book did you [[read the paper] and [then review __]]?
dependency for gap: IP-VP-VP

*Which book did you [[read __] and [then review the paper]]?
dependency for gap: IP-VP-VP

Again, this may be able to be addressed in a learner that is able to combine information from multiple dependencies in an utterance (perhaps because the learner has observed multiple dependencies resolved in utterances in the input).

Some cross-linguitic issues

High probability trigrams that may be ungrammatical

"Rizzi (1982) reports an interesting paradigm in Italian in which it looks as though simply doubling a grammatical sequence of trigrams leads to ungrammaticality…The problem for the current algorithm is that the container node sequence of the ungrammatical sentence in (28) (CP_{wh}-IP-VP-CP_{wh}-IP-VP) consists of the very same trigrams that are in the grammatical sentence in (27) (CP_{wh}-IP-VP, IP-VP-CP_{wh}, and VP-CP_{wh}-IP). Therefore the current algorithm will treat it as grammatical. Whether sentences such as (28) are unacceptable or not is an empirical question."

Some cross-linguitic issues

Constrained variation

"…this model predicts no constraints on the variation of island effects cross-linguistically: Any potential pattern of results (for the four island types investigated) can be derived given the correct input…The problem posed by constrained variation in island effects for the current strategy is straightforward: If there is indeed constrained variation in island effects cross-linguistically, then the current strategy would force us to conclude that the apparent constraint is simply a coincidence…it appears as though WH islands and Subject islands tend to covary (if a language has one, it will have the other; if it lacks one, it will lack the other)…[however] suggest that English relative clause dependencies exhibit Subject island effects but not WH island effects, which casts some doubt on the claim that WH and Subject islands always covary."

Complementizer *that*

That-trace effects

"…so-called *that-trace* effects are unacceptability that occurs when a gap immediately follows the complementizer *that* (32a), but does not arise when *that* is omitted (32b)…The current learning strategy can capture the distinction between these…"

*Who do you think that __ read the book?
Who do you think __ read the book?

Complementizer *that*

That-trace effects

"…However, the current learning strategy will also generate a preference for object gaps without *that* (33b) compared to object gaps with *that* (33a)…"

What do you think that he read __ ?
What do you think he read __ ?

"Interestingly…there is an object *that-trace* effect, but it is much smaller than the subject *that-trace* effect…the model generates an asymmetrical dispreference…when using the adult-directed corpora, which contain more instances of *that* (5.40 versus 2.81). This could be taken to be a developmental prediction of the current algorithm: Children may disprefer object gaps in embedded *that-CP* clauses more than adults, and this dispreference will weaken as they are exposed to additional tokens of *that* in utterances containing dependencies."

Open questions: Summary

Why does the system attempt to learn constraints on dependencies at all?

Why does the system treat *wh-dependencies* as separate from other dependencies like *rc*-dependencies and binding dependencies?

Why does the system track the container nodes of the dependency as opposed to other types of information about the dependencies?

Why does the system segment container node sequences into trigrams as opposed to other possible subsets?

Why does the system define container nodes as maximal projections as opposed to intermediate or smaller projects?

Why does the system subcategorize CP container nodes?

Open questions: Summary

"Although all of these questions can be encoded with explicit biases (as in the proposed algorithm), and many of them can be characterized using the framework in section 1 such that they are not obviously innate and domain-specific (i.e., UG-based) biases, it is not the case that we can confidently rule out the role of innate, domain-specific assumptions in giving rise to these biases. Future research is necessary to determine whether each of these problems raised by the acquisition of syntactic islands can be resolved without any innate, domain-specific biases."