Relevant Ambiguity vs Irrelevant Ambiguity

February 21, 2018

Learners

Learners

A learner is an agent who, given:

Learners

A learner is an agent who, given:

- a sample of the language (set of sentences) generated by the grammar G

$$L(G) = \{S_1 S_2 S_3 \ldots\}$$

Learners

A learner is an agent who, given:

- a sample of the language (set of sentences) generated by the grammar G

$$L(G) = \{S_1 S_2 S_3 \ldots\}$$

- is trying to deduce the grammmar

$$G$$

that generated them.

Learners

A learner is an agent who, given:

- a sample of the language (set of sentences) generated by the grammar G

$$L(G) = \{S_1 S_2 S_3 \ldots\}$$

- is trying to deduce the grammmar

$$G$$

that generated them.
They're trying to work backwards, and map from the sentences to the grammar.

$$\{S_1 S_2 S_3 \ldots\} \rightarrow G$$

Hypotheses

[1] the hypothesis is a function of the vector of per-parameter weights

Hypotheses

At any given point during the learning process, the learner has a hypothesis as to what G is. Depending on the learner, this can include

[1] the hypothesis is a function of the vector of per-parameter weights

Hypotheses

At any given point during the learning process, the learner has a hypothesis as to what G is. Depending on the learner, this can include

- a single grammar, as in the Trigger Learning Algorithm or Yang's Variational Learner [1]:

$$\begin{bmatrix} 0 & 1 & 1 \end{bmatrix}$$

[1] the hypothesis is a function of the vector of per-parameter weights

Hypotheses

At any given point during the learning process, the learner has a hypothesis as to what G is. Depending on the learner, this can include

- a single grammar, as in the Trigger Learning Algorithm or Yang's Variational Learner [1]:

$$\begin{bmatrix} 0 & 1 & 1 \end{bmatrix}$$

- a set of competing grammars, as in Clark's Genetic Algorithm:

$$\left\{ \begin{bmatrix} 0 & 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \right\}$$

[1] the hypothesis is a function of the vector of per-parameter weights

Hypotheses

At any given point during the learning process, the learner has a hypothesis as to what G is. Depending on the learner, this can include

- a single grammar, as in the Trigger Learning Algorithm or Yang's Variational Learner [1]:

$$\begin{bmatrix} 0 & 1 & 1 \end{bmatrix}$$

- a set of competing grammars, as in Clark's Genetic Algorithm:

$$\left\{ \begin{bmatrix} 0 & 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \right\}$$

- a vector of per-parameter confidence levels, as in Katherine's learner:

$$\begin{bmatrix} 0.05 & 0.6 & 0.98 \end{bmatrix}$$

[1] the hypothesis is a function of the vector of per-parameter weights

Sentences

Sentences

- Each sentence the learner receives from $L(G)$ is a potential piece of evidence they can use to update their hypothesis.

Sentences

- Each sentence the learner receives from $L(G)$ is a potential piece of evidence they can use to update their hypothesis.
- If we characterize the learner as forming a hypothesis for each parameter value separately, we can classify a sentence s as evidence for the setting of parameter P_i in the following three ways.

1. strong (relevant) evidence that $P_i = 0$ or $P_i = 1$

1. strong (relevant) evidence that $P_i = 0$ or $P_i = 1$

- The only grammars in the domain that ever license sentence s are those that have have $P_5 = 1$, for example.

1. strong (relevant) evidence that $P_i = 0$ or $P_i = 1$

- The only grammars in the domain that ever license sentence s are those that have have $P_5 = 1$, for example.
- Observing s in the input data is a globally valid trigger for $P_5 = 1$. We would never observe s in a language with $P_5 = 0$.

2. ambiguous (relevant) evidence as to the value of P_i

2. ambiguous (relevant) evidence as to the value of P_i

- s exists in languages where $P_5 = 0$ and in languages where $P_5 = 1$.

2. ambiguous (relevant) evidence as to the value of P_i

- s exists in languages where $P_5 = 0$ and in languages where $P_5 = 1$.
- The fact that we've observed s is not useful information on its own, it's not a global trigger.

2. ambiguous (relevant) evidence as to the value of P_i

- s exists in languages where $P_5 = 0$ and in languages where $P_5 = 1$.
- The fact that we've observed s is not useful information on its own, it's not a global trigger.
- But maybe we can still learn something relevant to the setting of P_5 by inspecting the contents of s.

3. irrelevant as evidence of P_i 's value

3. irrelevant as evidence of P_i 's value

- For every grammar G that licenses s, there is a corresponding grammar G' that also licenses s.

3. irrelevant as evidence of P_i 's value

- For every grammar G that licenses s, there is a corresponding grammar G' that also licenses s.
- In G, $P_5 = 0$
- In G', $P_5 = 1$

3. irrelevant as evidence of P_i 's value

- For every grammar G that licenses s, there is a corresponding grammar G' that also licenses s.
- In G, $P_5 = 0$
- In G', $P_5 = 1$
- All the other parameters are exactly the same.
- G and G' are minimal pairs on P_5.

3. irrelevant as evidence of P_i 's value

- For every grammar G that licenses s, there is a corresponding grammar G' that also licenses s.
- In G, $P_5 = 0$
- In G', $P_5 = 1$
- All the other parameters are exactly the same.
- G and G' are minimal pairs on P_5.
- Like the ambiguous case, the fact that we've observed s is not useful information on its own,

3. irrelevant as evidence of P_i 's value

- For every grammar G that licenses s, there is a corresponding grammar G' that also licenses s.
- In G, $P_5 = 0$
- In G', $P_5 = 1$
- All the other parameters are exactly the same.
- G and G' are minimal pairs on P_5.
- Like the ambiguous case, the fact that we've observed s is not useful information on its own,
- But we can also conclude that because *it never matters* what P_5 is set to, we should not try to learn about P_5 from s.

3. irrelevant as evidence of P_i 's value

- Whatever syntactic phenomena P_5 describes is simply not expressed at all in s (can we actually draw this strong conclusion from the domain-level data?).

An Algorithm for generating irrelevance strings

To determine what kind of evidence s is with regard to P_3:

[2] another grammar in G^s that's exactly the same as g except for P_3.

An Algorithm for generating irrelevance strings

To determine what kind of evidence s is with regard to P_3:

1. G^s = all grammars that license s.

[2] another grammar in G^s that's exactly the same as g except for P_3.

An Algorithm for generating irrelevance strings

To determine what kind of evidence s is with regard to P_3:

1. G^s = all grammars that license s.
2. P_3^s = the set of all values of P_3 in G^s

[2] another grammar in G^s that's exactly the same as g except for P_3.

An Algorithm for generating irrelevance strings

To determine what kind of evidence s is with regard to P_3:

1. G^s = all grammars that license s.
2. P_3^s = the set of all values of P_3 in G^s
3. If $P_3^s = \{0\}$ or $P_3^s = \{1\}$ (only one value showed up),
   - then s is a globally valid trigger for $P_3 = 0$ or $P_3 = 1$, respectively. Emit a 0 or 1.

[2] another grammar in G^s that's exactly the same as g except for P_3.

An Algorithm for generating irrelevance strings

To determine what kind of evidence s is with regard to P_3:

1. G^s = all grammars that license s.
2. P_3^s = the set of all values of P_3 in G^s
3. If $P_3^s = \{0\}$ or $P_3^s = \{1\}$ (only one value showed up),
   - then s is a globally valid trigger for $P_3 = 0$ or $P_3 = 1$, respectively. Emit a 0 or 1.
4. Otherwise we can now assume s is either ambiguously relevant * or irrelevant ~. Assume s irrelevant, unless step 5 finds finds otherwise.

[2] another grammar in G^s that's exactly the same as g except for P_3.

An Algorithm for generating irrelevance strings

To determine what kind of evidence s is with regard to P_3:

1. G^s = all grammars that license s.
2. P_3^s = the set of all values of P_3 in G^s
3. If $P_3^s = \{0\}$ or $P_3^s = \{1\}$ (only one value showed up),
   - then s is a globally valid trigger for $P_3 = 0$ or $P_3 = 1$, respectively. Emit a 0 or 1.
4. Otherwise we can now assume s is either ambiguously relevant * or irrelevant ~. Assume s irrelevant, unless step 5 finds finds otherwise.
5. For each g in G^s

[2] another grammar in G^s that's exactly the same as g except for P_3.

An Algorithm for generating irrelevance strings

To determine what kind of evidence s is with regard to P_3:

1. G^s = all grammars that license s.
2. P_3^s = the set of all values of P_3 in G^s
3. If $P_3^s = \{0\}$ or $P_3^s = \{1\}$ (only one value showed up),
   - then s is a globally valid trigger for $P_3 = 0$ or $P_3 = 1$, respectively. Emit a 0 or 1.
4. Otherwise we can now assume s is either ambiguously relevant * or irrelevant ~. Assume s irrelevant, unless step 5 finds finds otherwise.
5. For each g in G^s
   - look for the minimal pair [2] of g on P_3. We can do this by toggling bit P_3 in g, and checking if $\mathit{toggled}(g) \in G^s$.

[2] another grammar in G^s that's exactly the same as g except for P_3.

An Algorithm for generating irrelevance strings

To determine what kind of evidence s is with regard to P_3:

1. G^s = all grammars that license s.
2. P_3^s = the set of all values of P_3 in G^s
3. If $P_3^s = \{0\}$ or $P_3^s = \{1\}$ (only one value showed up),
   - then s is a globally valid trigger for $P_3 = 0$ or $P_3 = 1$, respectively. Emit a 0 or 1.
4. Otherwise we can now assume s is either ambiguously relevant * or irrelevant ~. Assume s irrelevant, unless step 5 finds finds otherwise.
5. For each g in G^s
   - look for the minimal pair [2] of g on P_3. We can do this by toggling bit P_3 in g, and checking if $toggled(g) \in G^s$.
   - if one exists, then toggling P_3 changed nothing wrt s.

[2] another grammar in G^s that's exactly the same as g except for P_3.

An Algorithm for generating irrelevance strings

To determine what kind of evidence s is with regard to P_3:

1. G^s = all grammars that license s.
2. P_3^s = the set of all values of P_3 in G^s
3. If $P_3^s = \{0\}$ or $P_3^s = \{1\}$ (only one value showed up),
   - then s is a globally valid trigger for $P_3 = 0$ or $P_3 = 1$, respectively. Emit a 0 or 1.
4. Otherwise we can now assume s is either ambiguously relevant * or irrelevant ~. Assume s irrelevant, unless step 5 finds finds otherwise.
5. For each g in G^s
   - look for the minimal pair [2] of g on P_3. We can do this by toggling bit P_3 in g, and checking if $\mathit{toggled}(g) \in G^s$.
   - if one exists, then toggling P_3 changed nothing wrt s.
   - else if it doesn't exist, but that's because it's not one of the legal 3072 colag languages, we can't make a claim (?)

[2] another grammar in G^s that's exactly the same as g except for P_3.

An Algorithm for generating irrelevance strings

To determine what kind of evidence s is with regard to P_3:

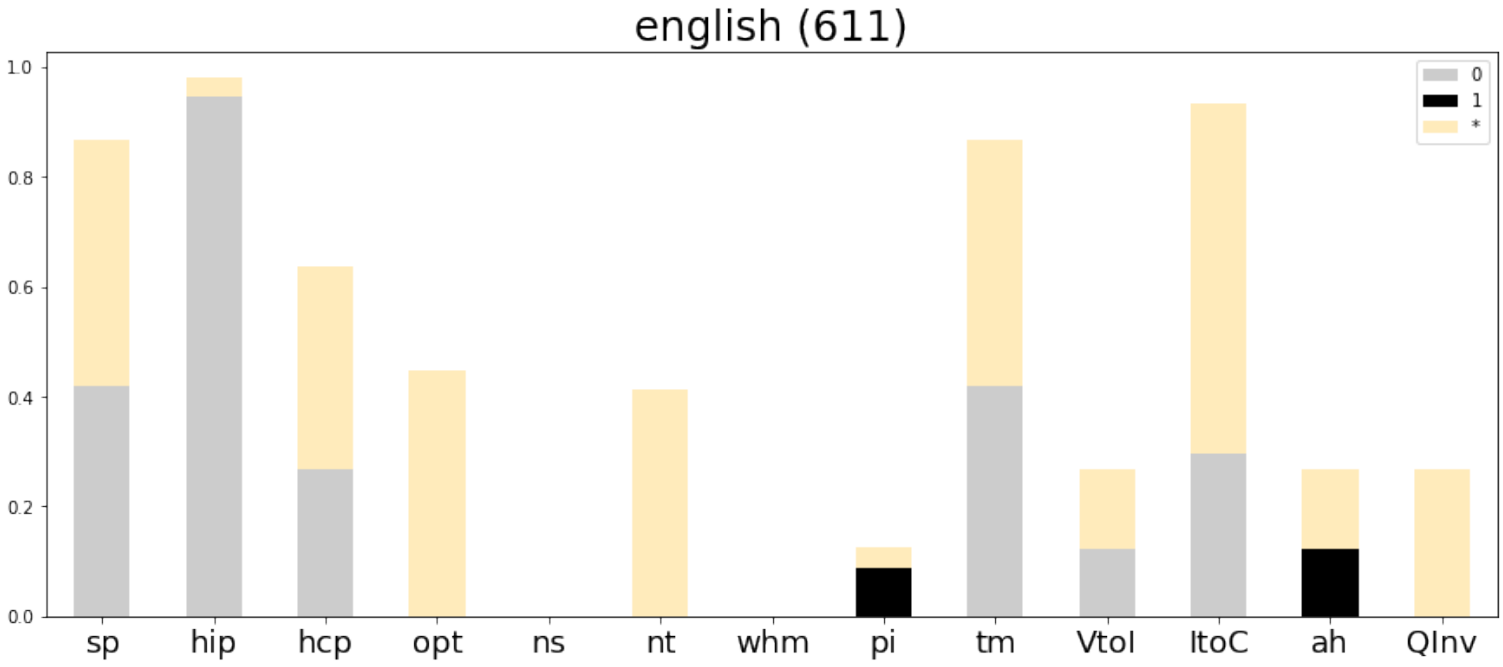
1. G^s = all grammars that license s.
2. P_3^s = the set of all values of P_3 in G^s
3. If $P_3^s = \{0\}$ or $P_3^s = \{1\}$ (only one value showed up),
   - then s is a globally valid trigger for $P_3 = 0$ or $P_3 = 1$, respectively. Emit a 0 or 1.
4. Otherwise we can now assume s is either ambiguously relevant * or irrelevant ~. Assume s irrelevant, unless step 5 finds finds otherwise.
5. For each g in G^s
   - look for the minimal pair [2] of g on P_3. We can do this by toggling bit P_3 in g, and checking if $toggled(g) \in G^s$.
   - if one exists, then toggling P_3 changed nothing wrt s.
   - else if it doesn't exist, but that's because it's not one of the legal 3072 colag languages, we can't make a claim (?)
   - else if it doesn't exist, we've found an example where P_3 actually has some effect on the appearance of s in a language. It's not irrelevant, just ambiguous. Emit a *.

[2] another grammar in G^s that's exactly the same as g except for P_3.

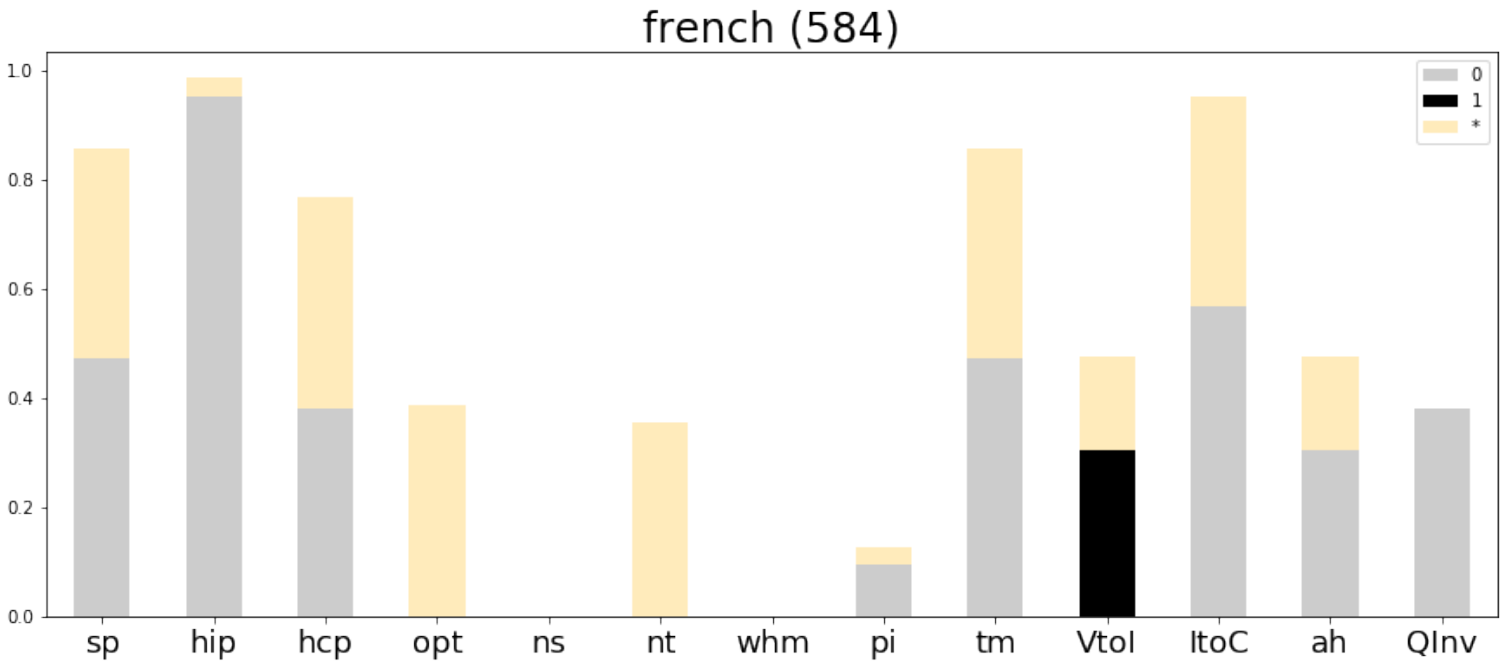
Question

```
if (minimal_pair not in generators and
    minimal_pair not in disallowed):
    relstr[param] = '*'
    break
```

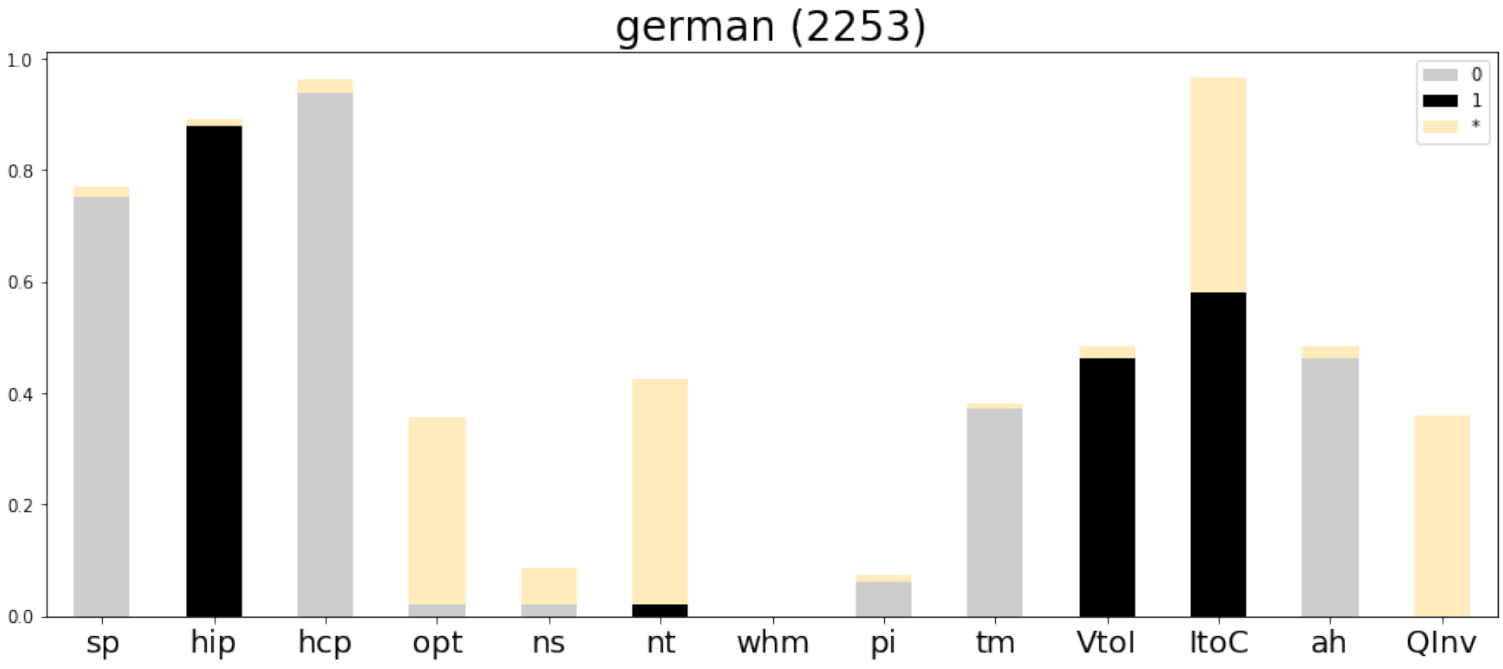
Some languages



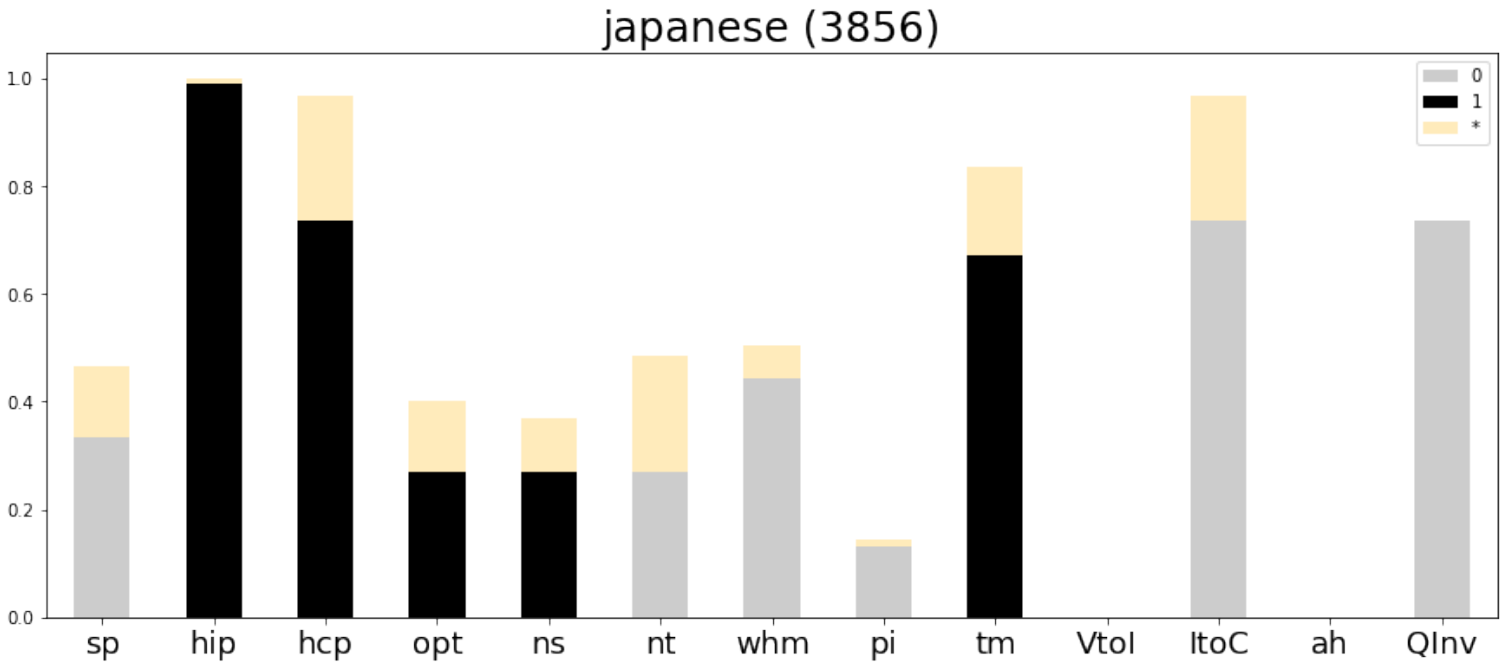
Some languages



Some languages



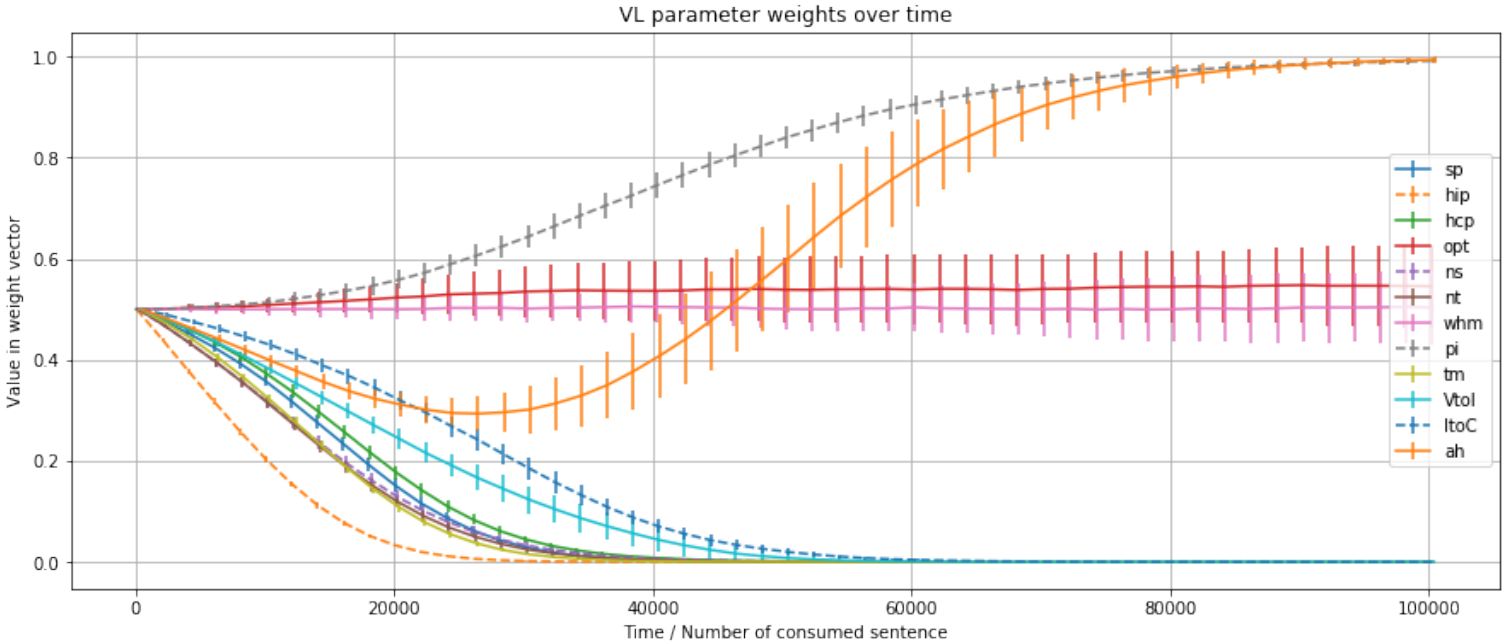
Some languages

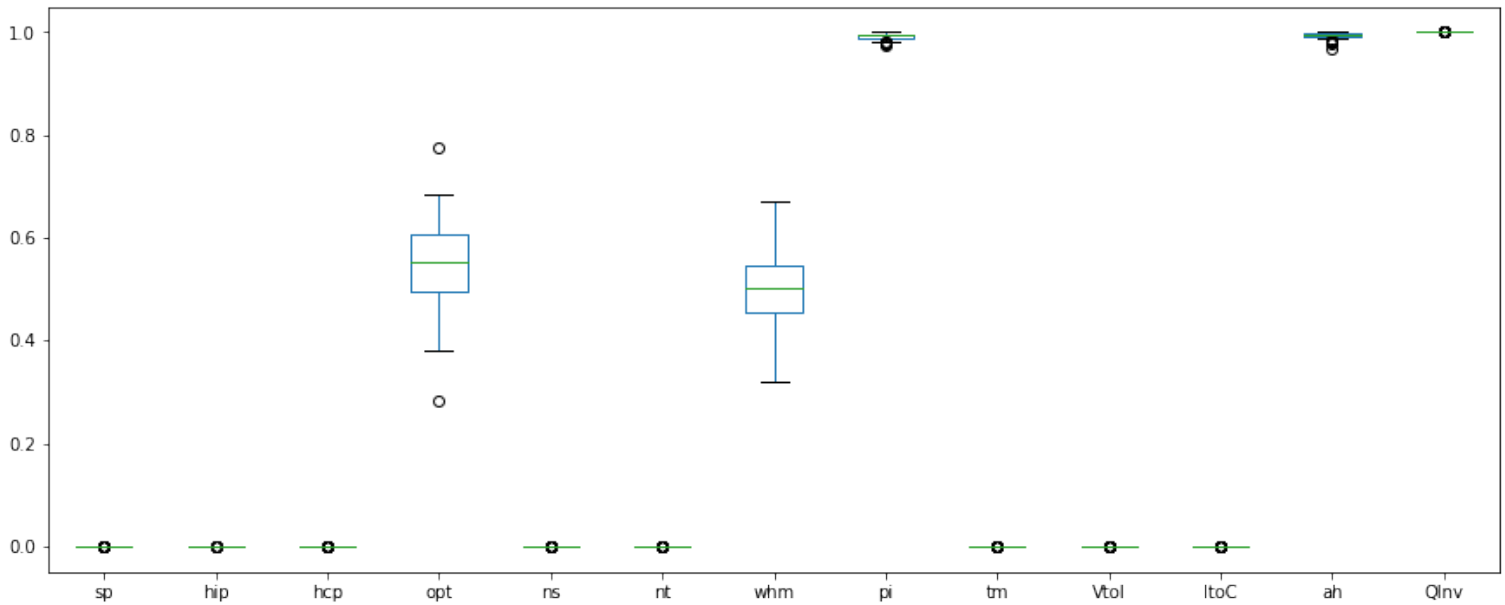


Discarding useless data

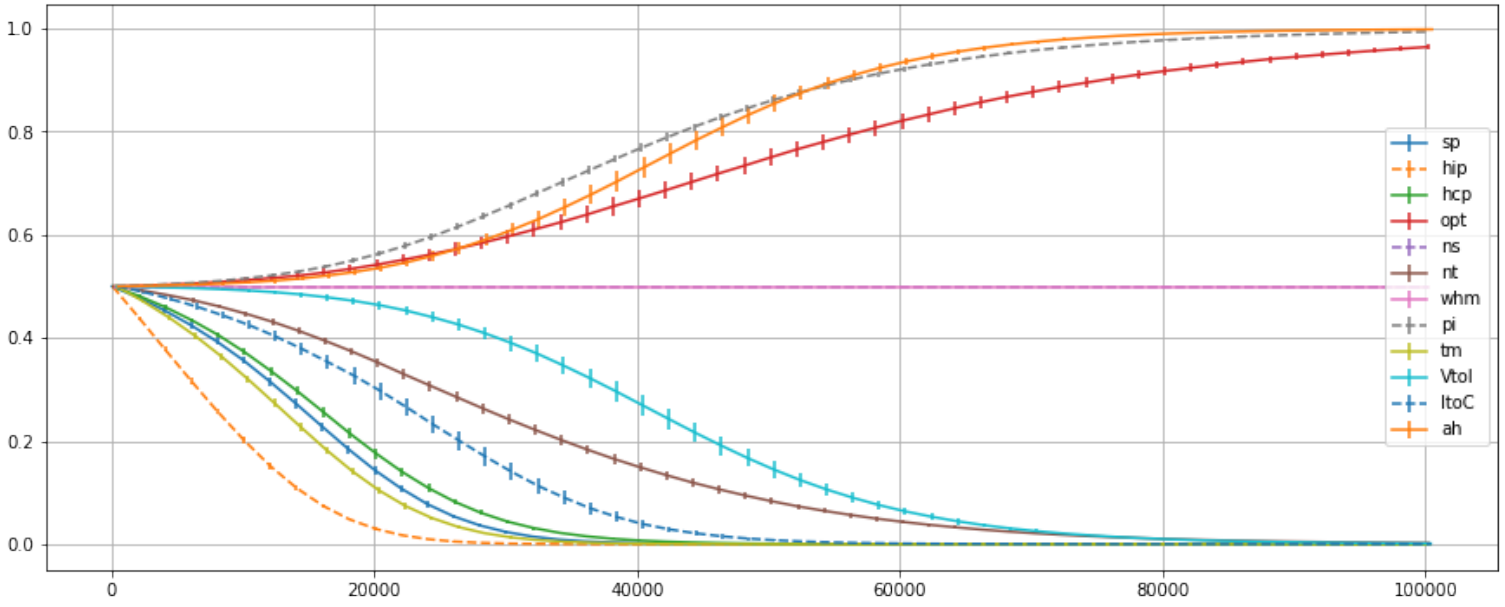
- What happens if a learner uses information about irrelevance to discard sentences?

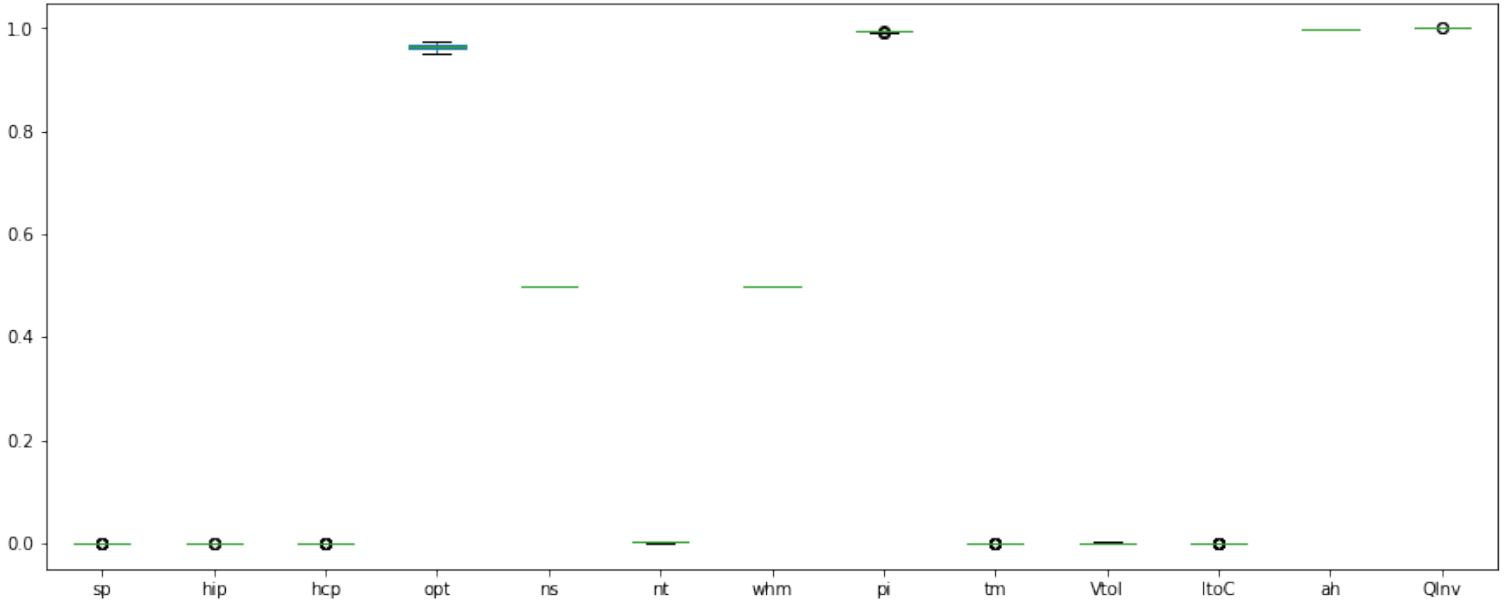
Yang's Reward-only VL Learning English (611)





Reward-relevant-only VL Learning English (611)

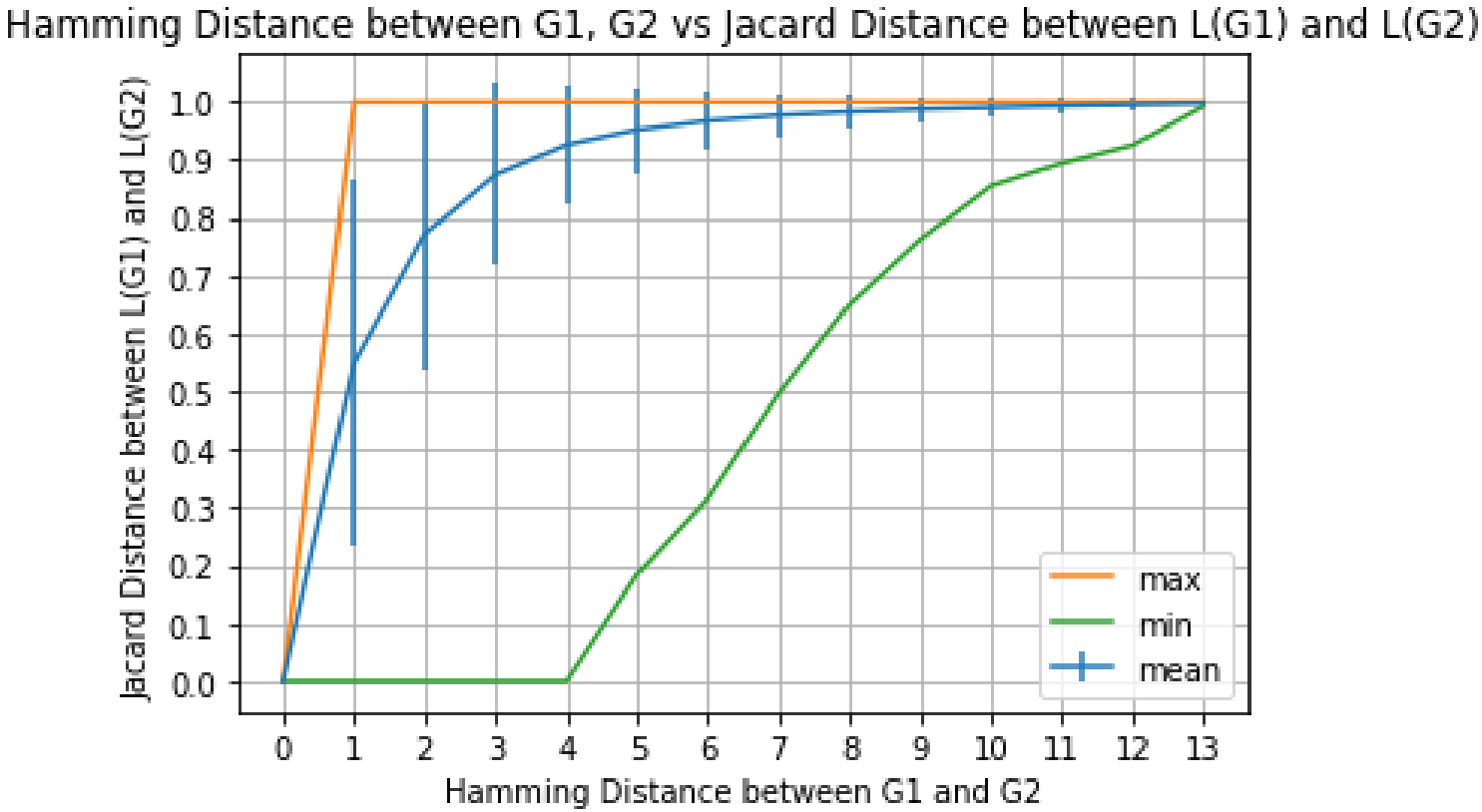




Reward-relevant-only VL Learning English (611)

- Optional Topic converges
- Null Subject fails to converge
- Affix-hopping moves in a single direction
- VtoI, ItoC and Null Topic take longer to learn

Smoothness: Parameters vs Sentences



- hamming distance – number of bits that differ between two bit-strings

$$jacard(L(G1), L(G2)) = \frac{L(G1) \cap L(G2)}{L(G1) \cup L(G2)} = \frac{\#\text{ sentences in common}}{\#\text{ sentences in total}}$$

Smoothness: Parameters vs Trigger-types