Consistent affix linearization in subanalysis learning

Sebastian Bank
sebastian.bank@uni-leipzig.de
Department of Linguistics
University of Leipzig

American International Morphology Meeting 2
University of California, San Diego
November 8–10, 2013

Introduction
Mission: Subanalysis in morphological learning

Background: Linearization in morphological grammar

Consistent general order

"...the linear descriptive sequence of affixes is reflected in the rule system as the relative ordering of the corresponding rules. That is, the fact that a word has the form stem + affix\textsubscript{1} + affix\textsubscript{2} reflects the fact that the rule attaching affix\textsubscript{1} applied 'before' the rule that attaches affix\textsubscript{2}.”

(Anderson 1992)

Consistency requirements

\[
given \ AB, BA \text{ if } A \prec B \text{ then } *B \prec A \\
hence A\textsubscript{1}B, BA\textsubscript{2}
\]

no symmetry

\[
given \ AB, BC, CA \text{ if } A \prec B \text{ and } B \prec C \text{ then } *C \prec A \\
hence A\textsubscript{1}B, BC, CA\textsubscript{2}
\]

no cycles

…alternatives to homophony: non-subanalysis, reordering

Introduction

Background: Templatic linearization + reordering

Expectations

• morphologies with reordering or cycles are structurally more complex
• typologically overriding the general order should be dispreferred
• sequences like AB, BA or AB, BC, CA should be harder to subanalyze
• diachronic pressure to realign order-inconsistent morphemes

Alternative hypotheses

1. affix linearization reflects pairwise predecessor-successor relations (precedence constraints, e.g. Paster 2005) or bigram constraints (Ryan 2010):
   they can contain \textit{cycles} where the parts never co-occur \textit{en bloc}
   \allowcycles\nontransitivity

2. learners do not generalize over the affix linearization (rote memorization):
   different occurrences of the same affix can be \textit{inconsistent}
   \allowsymmetry

Algorithm-based learning of morphological grammar(s)

Components meaning assignment, segmentation, linearization, blocking, …

Challenges search-space (analytical options), local vs. global optimization
interaction of components, look-ahead, syncretism vs. homonymy

Goals more falsifiable predictions for analysis/framework comparison
determine the distinct logical properties of different formalisms
explore logical space of morphological theories

In this talk
combined meaning assignment and subanalysis learner (Bank & Trommer 2013)
using iterative local optimization (Harmonic Serialism, McCarthy 2010)
that can satisfy three different linearization requirements

\[
A \subset B \subset C
\]
Morpheme order as asymmetric transitive relation

Each morpheme string in a paradigm cell gives a chain (strict total ordering), i.e. a relation that is asymmetric, transitive and total:

\[ (A, B), (A, C), (B, C) \]

 asymmetry \( a < b \Rightarrow b \notin a \)

 transitivity \( a < b \land b < c \Rightarrow a < c \)

 totality \( a < b \lor b < a \)

Morpheme chains are consistent if their union is a strict partial ordering, i.e. a relation that is asymmetric and transitive:

\[ (A, B), (A, C), (B, C), (D, C) \]

 From partial ordering to total ordering (cf. Kahn 1962)

Partial orderings can always be linearly extended into a compatible total ordering.

\[ ABC, BD, E \]

\[ A \lessdot B < C \]

\[ A \lessdot B < D \]

\[ C \lessdot D \]

\[ E \lessdot A, B, C, D \]

Linear order as relation

From partial ordering to linearization template

Templates assign each morpheme to a slot \( i \in \mathbb{N} \) such that slots are totally ordered.

\[ A ightharpoonup 1, B ightharpoonup 2, C ightharpoonup 3, D ightharpoonup 3, E ightharpoonup 3 \]

ordered partition

The ordering between the slots gives a strict weak ordering on the morphemes, i.e. a partial ordering in which also the incomparability is transitive.

\[ \{A\} \lessdot \{B\} \lessdot \{C, D, E\} \]

\[ C \parallel D \land D \parallel E \Rightarrow C \parallel E \]

negative transitivity

Deriving a template from a partial ordering is almost identical to topological sort:

Generate compatible \( \text{item, slot} \) pairs for a set \( A \) with a strict partial ordering \( \lessdot \)

1: \( i \leftarrow 0 \)
2: while \( A \neq \emptyset \) do \( \triangleright \) items to assign
3: \( i \leftarrow i + 1 \) \( \triangleright \) next slot index
4: \( M \leftarrow \{a \in A \mid \forall b \in A : b \notin a\} \)
5: for all \( m \in M \) do \( \triangleright \) in any order
6: \( \text{yield } (m, i) \) \( \triangleright m \to i \)
7: \( A \leftarrow A \setminus M \) \( \triangleright \) remaining items

Input: a set \( C := \{m_1, m_2, \ldots, m_n\} \) of morpheme sequences \( m_s \) of length \( n > 0 \)

Output: an initially empty list \( \text{Slots} := \{S_1, \ldots, S_n\} \) of morphemes in the same slot \( S \)

Position class template generator (left-anchored version)

\( C = \{(b, \text{in}), (b, \text{ist}), (s, \text{in}, d), (s, \text{ei}, d)\} \)

\( R = \{(b, \text{in}), (b, \text{ist}), (s, \text{in}, d), (s, \text{d}), (e, \text{d})\} = R^+ \)

\( M = \{(b, \emptyset), (b, s), (s, \emptyset), (b, \{b\}), (s, \{b\}), (s, d), (s, e, d), (e, s)\} \)

\( A_1 = \{b, \text{in}, s, d, e\} \)

\( A_2 = \{s, \text{in}, d, e\} \)

\( A_3 = \{d\} \)

\( S_1 = \{b, s\} \)

\( S_2 = \{s, \text{in}, e\} \)

\( S_3 = \{d\} \)

\( \text{Slots} = \{(b, s), (s, \text{in}, e), (d)\} \)

German 'to be'

\( s_g \) pl

1 b-in s-in-d

2 b-ist s-e-i-d

3 ist s-in-d

\( R = \{0100000, 0001000, 1001000, 0000100, 0000010\} \)

\( 1\) \( 2\) \( 3\) \( 4\) \( 5\) \( 6\) \( 7\) \( 8\)
Morpheme order as predecessor-successor relation

\[ \{A, B, C\} \iff \{A, B, C, A\} \]

\[ \text{asymmetry: } a < b \Rightarrow b \not< a \]
\[ \text{transitivity: } a < b < c \Rightarrow a < c \]

\[ [\begin{array}{ccc}
0 & 1 & 0 \\
0 & 0 & 1 \\
0 & 0 & 0
\end{array}] \]

If the relation has a cycle, it is not possible to enumerate all elements in consistent order. Topological sort ends up in an infinite loop.

If \( A < B < C \), then it is not possible to enumerate.

AB, BC, CA ↔ \{A, B, C\}, \{C, A, B\}

Maintaining precedence in incremental segmentation

Annotate left-string-right with left = predecessors and right = successors sets.

Discharge (sub)strings only when \( \text{left} \) and \( \text{right} \) are disjoint.

\[ \{\phi \text{aphikta}_a\} \]
\[ \{\phi \text{aphikti}_c\} \]
\[ \{\phi \text{aphikti}_b\} \]
\[ \{\phi \text{hamphikta}_a\} \]

1: discharge phik with left \( \subseteq \emptyset \) and right \( \subseteq \emptyset \)

2: discharge t with left \( \subseteq \{1\} \) and right \( \subseteq \emptyset \)

3: discharge a with left \( \subseteq \emptyset \) and right \( \subseteq \{1, 2\} \)

4: discharge a with left \( \subseteq \{1, 2\} \) and right \( \subseteq \emptyset \)

Consistency alternatives and extensions

- remove markers either from left to right or from right to left?
- identify stem and then remove affixes either inside out or outside in?
- prefer full strings over edge-including substrings over internal substrings (Bank & Trommer to appear)
- all other things being equal, prefer longer strings
- exclude removals that would introduce a cycle in the transitive closure of the union of all the precedence pairs ⇒ templatetic order
Incremental learning with cyclic form removal

- iterative local optimization akin to Harmonic Serialism (greedy algorithm)
- competing segmentation options are bled by form removal
- no exhaustive search through analytical options
- abstracts away from marker hypothesis generation and evaluation

**Input:** a set $P := \{(f, c), \ldots\}$ of cell form $f$ and cell meaning $c$ pairs

- a set $\mathcal{U}$ of possible meanings with a partial order $\leq$

**Output:** an initially empty lexicon $\text{Lex}$

1. $P \leftarrow \{(f, c, i) \mid (f, c) \in P, i = 1\}$ \quad \triangleright \text{add position index}
2. while $3(f, c, i) \in P : f \neq \epsilon$ do
3. \quad $H \leftarrow \{(f, m) \mid \text{form } f \text{ and meaning } m \text{ of a marker hypothesis for } P\}$
4. \quad $\langle f, m \rangle \leftarrow \text{an optimal hypothesis } \epsilon H$
5. $P \leftarrow \text{DISCHARGE}(P, f, m)$
6. $\text{Lex} \leftarrow \text{Lex} \cup \{(f, m)\}$

Evaluating the accuracy of form-meaning pairs

Meaning assignment seeks optimal patterns of paradigmatic distributions (Pertsova 2007)

**Informal and formal paradigm representation**

| $\overset{[\cdot + y]}{[\cdot - x]}$ | $\overset{[\cdot - y]}{a}$ | $\overset{[\cdot + y]}{b}$ | $\overset{[\cdot + x - y]}{a, [\cdot + x - y]}$, $\overset{[\cdot - x - y]}{a}, [\cdot - x - y]}$ | $\overset{[\cdot + y]}{[\cdot - x]}$ | $\overset{[\cdot - y]}{a}$ | $\overset{[\cdot + y]}{b}$ | $\overset{[\cdot - x + y]}{a, [\cdot - x + y]}$, $\overset{[\cdot + y]}{[\cdot - x]}$ | $\overset{[\cdot - y]}{a}$ |
|-------------------------|----------------|----------------|-----------------------------|-------------------------|----------------|----------------|-----------------------------|----------------|----------------|
| Affix hypotheses with accuracy evaluation |
| (form, meaning) | false positives | false negatives | implication | precision | recall |
| a. $(b, [-y])$ | $-$ | $-$ | $\leftarrow$ | 1 | 1 |
| b. $(a, [+y])$ | $-$ | yes | $\leftarrow$ | $\frac{1}{2}$ | $\frac{1}{2}$ |
| c. $(a, [-x])$ | $-$ | yes | $\leftarrow$ | $\frac{2}{3}$ | $\frac{2}{3}$ |
| d. $(a, [\cdot])$ | yes | $-$ | $\rightarrow$ | 3 | 1 |
| e. $(b, [+x])$ | yes | yes | none | $\frac{3}{4}$ | $\frac{3}{4}$ |
| f. $(a, [-y])$ | yes | yes | none | 1 | 1 |

Avoid false positives ~ optimize precision ~ underinsertion

**OverInsertion:** Assign * to every paradigm cell where the hypothesis predicts the marker but the marker does not occur

Ainu verbal agreement (Tamura 2000)

<table>
<thead>
<tr>
<th>1sg</th>
<th>1pl</th>
<th>2sg</th>
<th>2pl</th>
<th>3sg</th>
<th>3pl</th>
<th>-</th>
</tr>
</thead>
<tbody>
<tr>
<td>1sg</td>
<td>eci-</td>
<td>eci-</td>
<td>ku-</td>
<td>ku-</td>
<td>ku-</td>
<td></td>
</tr>
<tr>
<td>1pl</td>
<td>eci-</td>
<td>eci-</td>
<td>ci-</td>
<td>ci-</td>
<td>-as</td>
<td></td>
</tr>
<tr>
<td>2sg</td>
<td>en-</td>
<td>-</td>
<td>e-</td>
<td>e-</td>
<td>e-</td>
<td></td>
</tr>
<tr>
<td>2pl</td>
<td>eci-en-</td>
<td>eci-un-</td>
<td>eci-</td>
<td>eci-</td>
<td>eci-</td>
<td></td>
</tr>
<tr>
<td>3sg</td>
<td>en-</td>
<td>un-</td>
<td>e-</td>
<td>eci-</td>
<td>O-</td>
<td>O-</td>
</tr>
<tr>
<td>3pl</td>
<td>en-</td>
<td>un-</td>
<td>e-</td>
<td>eci-</td>
<td>O-</td>
<td>O-</td>
</tr>
</tbody>
</table>

Meaning assignment for eci- with precision-favoring ranking

<table>
<thead>
<tr>
<th>*OverIns</th>
<th>*UnderIns</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. eci-[+2 +pl]</td>
<td>*OverIns</td>
</tr>
<tr>
<td>b. eci-[+2]</td>
<td>*7 (2sg, 2sg:X, 3:2sg)</td>
</tr>
<tr>
<td>c. eci-[Acc +2]</td>
<td>*5 (2pl, 2pl:X)</td>
</tr>
</tbody>
</table>
Avoid false negatives ~ optimize recall ~ overinsertion

*UNDERINSERTION:* Assign * to every paradigm cell where the marker occurs but the hypothesis does not predict it.

### Ainu verbal agreement

<table>
<thead>
<tr>
<th></th>
<th>1sg</th>
<th>1pl</th>
<th>2sg</th>
<th>2pl</th>
<th>3sg</th>
<th>3pl</th>
</tr>
</thead>
<tbody>
<tr>
<td>1sg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1pl</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2sg</td>
<td>en-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2pl</td>
<td>eci-</td>
<td>un-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3sg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3pl</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Meaning assignment for *eci-* with recall-favoring ranking

<table>
<thead>
<tr>
<th>a.</th>
<th>*</th>
<th>*</th>
<th>*</th>
<th>*</th>
<th>*</th>
<th>*</th>
</tr>
</thead>
<tbody>
<tr>
<td>b.</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>c.</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

### Consistent linear order subanalysis learner (main loop)

**MARKERS()** provides (sub)strings with their set of occurrences, one cell per occurrence.

**HYPOTHESES()** provides consistent marker hypotheses with accuracy-determining info.

**DISCHARGE()** removes material from the paradigm keeping track of precedence relations.

**Input:** a set \( P = \{(f, c), \ldots \} \) of cell form \( f \) and cell meaning \( c \) pairs

**Output:** an initially empty lexicon \( \text{Lex} \)

#### Transitive relation

**Nontransitivity in Chumbivilcas Quechua derivation affixes**

- INCH.AST: ri-schi
- ASS.RECP: schi-na
- RECP: na-ri

#### Costs and effects of transitive consistency

Homophony and extended exponent of the cycle-breaking na-variants more markers than possible without linearization, less general markers, reduced accuracy.

Locality of cycle detection effects generalization asymmetry in learning derives perfect generalizations as early as possible, less accurate residue at later stages.
Ensuring consistency: \textbf{HYPOTHESES}($M, \mathbb{U}, LR, Ord$)

\begin{itemize}
\item[A] $L \cap R = \emptyset$ and asymmetric relation with templatic linearization
\item[B] $R \times L \times Ord^* = \emptyset$ and asymmetric transitive closure
\item[C] no restriction and no generalization on morpheme order
\end{itemize}

Output: an initially empty set $H := \{ (f, C, L, R), m, (TP, FP, FN, TN), \ldots \}$ of form, contexts, predecessors, successors, meaning, true positives, false positives, false negatives, true negatives

\begin{algorithm}
\begin{algorithmic}
\State for all $(f, C, m) \in (M \times \mathbb{U})$ do
\State $L := \bigcup \{L \mid (c, i) \in C, m \subseteq c, (c, i, L, R) \in LR\}$ \Comment{occurrences meaning pairings}
\State $R := \bigcup \{R \mid (c, i) \in C, m \subseteq c, (c, i, L, R) \in LR\}$ \Comment{predecessors of subsumed}
\State if $L \cap R = \emptyset$ and $R \times L \times Ord^* = \emptyset$ then
\State $C' := \{ (c, i') \in C \mid (c, i') \in P, (c, i', L, R, c') \in LR, R_c \cap L = \emptyset, R_c \cap L = \emptyset, (R_c \times L \cup R \times L_c) \in Ord^* = \emptyset\}$ \Comment{consistent and cycle-free}
\State $T := \{ c \mid (c, i) \in C'\}$ \Comment{target cells}
\State $P := \{ c \in \mathbb{U}_c \mid m \subseteq c\}$ \Comment{positives}
\State $N := \{ c \in \mathbb{U}_c \mid m \subseteq c\}$ \Comment{negatives}
\State $H := H \cup \{ (f, C', L, R), m, (P \cap T, P \cap T, N \cap T, N \cap T)\}$
\end{algorithmic}
\end{algorithm}

Resulting paradigm and morpheme precedence relation

\begin{table}[h]
\begin{tabular}{|c|c|c|c|}
\hline
1 & 12 & 2 & 3 \\
\hline
SG & phik-t-a & a1-phik-t-a\textcolor{red}{2} & phik-t-a\textcolor{red}{2} \\
DU & phik-t-i & phik-t-i & a1-phik-t-i \\
PL & phik-k-t-i & phik-k-t-i & a1-phik-t-ini \\
\hline
\end{tabular}
\end{table}

Incremental templatic linearization of Dumi phikni ‘to get up’

\begin{table}[h]
\begin{tabular}{|c|c|c|c|}
\hline
Marker evaluation & *\text{UnderInsertion} \gg *\text{OverInsertion} \gg \text{MaxCoverage} \\
\hline
1 & 12 & 2 & 3 \\
\hline
SG & a & a-a & a \\
DU & i & a-i & a \\
PL & k-i & k-i & i+i \\
\hline
1 & 12 & 2 & 3 \\
\hline
SG & a & a-a & a \\
DU & i & a-i & a \\
PL & i & a-a & a \\
\hline
3 & 12 & 2 & 3 \\
\hline
SG & a & a-a & a \\
DU & i & a-i & a \\
PL & i & a-a & a \\
\hline
\end{tabular}
\end{table}

Implementation and results

Convert precedence relation to morpheme template

\begin{table}[h]
\begin{tabular}{|c|c|c|c|}
\hline
1 & 12 & 2 & 3 \\
\hline
\hline
\end{tabular}
\end{table}

\begin{table}[h]
\begin{tabular}{|c|c|c|c|}
\hline
1 & 12 & 2 & 3 \\
\hline
\hline
\end{tabular}
\end{table}

Extrinsic ordering

\begin{table}[h]
\begin{tabular}{|c|c|c|c|}
\hline
1 & 12 & 2 & 3 \\
\hline
\hline
\end{tabular}
\end{table}

Transitive reduction
Conclusion

- subanalysis for inflectional grammars with extrinsic order like Anderson (1992) can be learned with local optimization and order consistency checking
- learner biases can be used to predict the complexity/markedness of certain linearization patterns like nontransitivity and classify them
- modelling learning of morphological grammar(s) uncovers their logical structure and helps to make their comparison more empirically informed
- how can an interaction between subanalysis and reordering or correspondences between linearization and blocking be learned in a local way?

References