Language in Populations: The Interaction Between Learning & Change

Lisa Pearl
University of Maryland
April 2, 2007
Northwestern Institute on Complex Systems

Language Characteristics
(1) A trait individuals within a population have
(2) Transmitted via learning, rather than solely genetic
Language

Language Characteristics
(1) a trait individuals within a population have
(2) transmitted via learning, rather than solely genetic
(3) some parts mutable only during learning period

Language

Language Characteristics
(1) a trait individuals within a population have
(2) transmitted via learning, rather than solely genetic
(3) some parts mutable only during learning period
(4) linguistic composition of population can change
Language Change in a Population

Two opposing linguistic structures (e.g. Object Verb and Verb Object order) can be used probabilistically by individuals in a population.

Change in the probability of usage within the population proceeds at a certain rate.

Certain changes proposed to be the result of imperfect learning of precisely the right amount at the individual level (Lightfoot, 1991).

Imperfect Learning = Language Change

Individuals: the learner’s final probability distribution is different from the adult’s by a certain amount.

These individuals: source of data for future individuals.
Future individuals: converge on a probability distribution that is different.

Population-level: the population as a whole shifts at a certain rate, based on the amount individual learners differ from the rest of the population.

Modeling Correct Linguistic Behavior

If we instantiate a certain learning model for individuals of a population and the population changes at the correct rate, we conclude:

(1) individuals misconverged precisely the right amount
(2) the learning model that allows this amount of misconvergence is correct

Language Learning in Individuals: The Tricky Part

There is often a non-transparent relationship between the observable form of the data and the underlying system that produced it.
There is often a non-transparent relationship between the observable form of the data and the underlying system that produced it.

Syntactic System
- Observable form: word order
- Interference: movement rules

Subject  Object  Verb

Verb-Second (V2) movement

"It is unlikely that any example … would show the effect of only a single parameter value; rather, each example is the result of the interaction of several different principles and parameters" - Clark (1994)
Language Learning: Extracting Systematicity

“It is unlikely that any example … would show the effect of only a single parameter value; rather, each example is the result of the interaction of several different principles and parameters” - Clark (1994)

Potential solution: the learner focuses in on an informative subset of the data.

Potential issue: data sparseness

Road Map

Individual Learning Framework Overview

Population Modeling: Syntactic Language Change

Important Feature: grounded in empirical data
- real data distributions
- searching realistic data space for evidence of underlying system

Individual Learning Framework: 3 Components

1) Hypothesis space
   - A \( P_A = 0.5 \)
   - B \( P_B = 0.5 \)

2) Data intake
   - A \( P_A = ?? \)
   - B \( P_B = ?? \)

3) Update procedure
Benefits of Learning Framework

Components:
(1) hypothesis space  (2) data intake  (3) update procedure

Can combine discrete representations (hypothesis space) with probabilistic components (update procedure): get gradualness and variation found in real language learning.

Individual Model: Data Intake Filtering

Intuition 1: Use all available data to uncover a full range of systematicity, and allow probabilistic model enough data to converge.

Intuition 2: Use more "informative" data or more "accessible" data only.

Case Study: Model Specifics

Hypothesis Space: word order parameters
Object Verb (OV) vs. Verb Object (VO) order

Case Study: Model Specifics

Hypothesis Space: word order parameters
Object Verb (OV) vs. Verb Object (VO) order
Case Study: Model Specifics

Hypothesis Space: word order parameters
Object Verb (OV) vs. Verb Object (VO) order

Update Procedure: adapted Bayesian updating
shifts probabilities between opposing hypotheses
amount shifted depends on layout of hypothesis space

Difficult Feature: adult target state is a probability distribution
target state is usually one hypothesis or the other
converging on the right probability is harder

Learning & Change: The Big Questions

1. Is it feasible to filter?
   *Is there a data sparseness problem?*

2. Is it sufficient to filter?
   *Can we get the right population behavior if we filter?*

3. Is it necessary to filter?
   *Must we filter to get the right population behavior?*

Road Map

Individual Learning Framework Overview

Population Modeling: Syntactic Language Change
Old English: description & proposed individual filters
Old English data & feasibility of filtering
Modeled learners and populations
Estimating ground truth
Sufficiency & necessity of filtering
Old English
Learning: Old English OV vs. VO order

Target State: probability distribution between OV and VO hypotheses (YCOE Corpus, 2003; PPCME2 Corpus, 2000; similar models: Yang, 2002; Pintzuk, 2002; Kroch & Taylor, 1997; Bock & Kroch, 1989)

Filter 1: Use data perceived as unambiguous (Dresher, 1999; Lightfoot, 1999; Fodor, 1998)

Filter 2: Use structurally "simple" data - matrix clause or "degree-0" data (Lightfoot, 1991)

Jack told his mother that the giant was easy to fool.

Problems
Potential problem for feasibility: data sparseness
degree-0 unambiguous data set is significantly smaller than entire input set

Learners must use this data set to misconverge the exact right amount at each point in time so that the population changes at the correct rate

Road Map
Individual Learning Framework Overview
Population Modeling: Syntactic Language Change
Old English: description & proposed individual filters
Old English data & feasibility of filtering
Modeled learners and populations
Estimating ground truth
Sufficiency & necessity of filtering
Old English OV and VO

**OV-biased**: between 1000 and 1150 A.D.

he Subj Gode Obj þ ancode TensedVerb
He thanked God
(Beowulf, 625, ~1100 A.D.)

VO-biased: by 1200 A.D.

& [mid his stefne]pp he Subj awec D TensedVerb deade Obj [to life]pp
And with his stem, he awakened the-dead to life
(James the Greater, 30.31, ~1150 A.D.)

Old English OV and VO

**OV-biased**: between 1000 and 1150 A.D.

he Subj God Subj God TensedVerb
He thanked God
(Beowulf, 625, ~1100 A.D.)

VO-biased: by 1200 A.D.

Ambiguous Data

Subject TensedVerb Object is ambiguous
(most common data type)

OV, +V2
he Subj clensaTensedVerb fSbj [ba sawle þes radendan]Obj [to-sawle]TensedVerb
they purified the souls [the advising]-Gen

VO, -V2
he Subj clensaTensedVerb [ba sawle þes radendan]Obj [to-sawle]TensedVerb
they purified the souls [the advising]-Gen

‘They purified the souls of the advising ones.’
(Alcuin’s De Virtutibus et Vitiis, 83.59, ~1150 A.D.)
Perceived Unambiguous Data: Examples

Unambiguous OV

He [he] may pray (to) him

(Ælfric’s Letter to Wulsige, 87.107, ~1075 A.D.)

Unambiguous VO

[Paulus] then lifted [his] head up.

(Blickling Homilies, 167.35, between 900 and 1000 A.D.)

Perceived Unambiguous Data: Making “Unambiguous” Feasible

Definitions of data perceived as unambiguous are heuristic and/or involve only partial knowledge of the adult linguistic system (Lightfoot 1999, Dresher 1999, Fodor 1998)

OV:

[...][... Object TensedVerb ...]

VO:

[...][... Object-TensedVerb Object ...]

This allows the learner to identify some data points as unambiguous (even if they’re actually not for someone with full knowledge of the adult linguistic system)

The Effect of Filtering

Unambiguous degree-0 data distribution may differ from adult distribution used to generate data

...so individuals can misconverge.

Road Map

Individual Learning Framework Overview

Population Modeling: Syntactic Language Change

- Old English: description & proposed individual filters
- Old English data & feasibility of filtering
- Modeled learners and populations
- Estimating ground truth
- Sufficiency & necessity of filtering
The Model: Individual-Level

Individual learner tracks \( p_{VO} \) = probability of using VO
probability of using OV = 1 - \( p_{VO} \)

Old English: \( 0.0 \leq p_{VO} \leq 1.0 \)
Ex: \( 0.3 = 30\% \) use of VO, \( 70\% \) use of OV

Initial \( p_{VO} = 0.5 \) (unbiased)

\[
\text{Max}(\text{Prob}(p_{VO} | u)) = \text{Max}(\text{Prob}(u | p_{VO}) \times \text{Prob}(p_{VO}) \div \text{Prob}(u) \\
\text{Max}(\text{Prob}(p_{VO} | u)) = p_{VO} \times \left[ r \times (1-p_{VO}) \right] \times \left[ p_{VO} \times (1-(1-p_{VO})) \right] \div (n+1) \)
\]

Replace 1 in numerator and denominator with \( c = p_{VOprev} \) if VO, \( c = (1-p_{VOprev}) \) if OV
3.0 \# m \# 5.0

Individual-Level Learning Algorithm

1. Set initial \( p_{VO} \) to 0.5.
2. Encounter data point from an "average" member of the population.
3. If the data point is degree-0 and unambiguous, use update procedure to shift beliefs in hypotheses.
4. Repeat (2-3) until the fluctuation period is over, as determined by \( n \).

\[
\text{If OV data point} \\
P_{VO} = (p_{VOprev} \times n) / (n+c) \]

\[
\text{If VO data point} \\
P_{VO} = (p_{VOprev} / n+c) / (n+c)
\]
Individual-Level Learning Algorithm

(1) Set initial $p_{VO}$ to 0.5.

(2) Encounter data point from an “average” member of the population.

(3) If the data point is degree-0 and unambiguous, use update functions to shift hypothesis probabilities.

(4) Repeat (2-3) until the fluctuation period is over, as determined by $n$.

Biased Data Intake Distributions

$p_{VO}$ shifts away from 0.5 when there is more of one data type in the intake than the other (advantage (Yang, 2000) of one data type).
Biased Data Intake Distributions

$p_{VO}$ shifts away from 0.5 when there is more of one data type in the intake than the other (advantage (Yang, 2000) of one data type)

<table>
<thead>
<tr>
<th></th>
<th>$OV$ Advantage in Unamb $D_0$</th>
<th>$OV$ Advantage in Unamb $D_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000 A.D.</td>
<td>19.5%</td>
<td>41.7%</td>
</tr>
<tr>
<td>1000-1150 A.D.</td>
<td>2.8%</td>
<td>28.7%</td>
</tr>
<tr>
<td>1200 A.D.</td>
<td>-2.7%</td>
<td>-45.2%</td>
</tr>
</tbody>
</table>

Population-Level Algorithm

1. Set the age range of the population from 0 to 60 years old and create 18,000 population members.

2. Initialize the members of the population to the average $p_{VO}$ at 1000 A.D. Set the time to 1000 A.D.

3. Move forward 2 years.

4. Members age 59-60 die off. The rest of the population ages 2 years.

(1) Set the age range of the population from 0 to 60 years old and create 18,000 population members.

(2) Initialize the members of the population to the average $p_{VO}$ at 1000 A.D. Set the time to 1000 A.D.

(3) Move forward 2 years.

(4) Members age 59-60 die off. The rest of the population ages 2 years.
Population-Level Algorithm

1. Set the age range of the population from 0 to 60 years old and create 18,000 population members.

2. Initialize the members of the population to the average p_VO at 1000 A.D. Set the time to 1000 A.D.

3. Move forward 2 years.

4. Members age 59-60 die off. The rest of the population ages 2 years.

5. New members are born. These new members use the individual acquisition algorithm to set their p_VO.

6. Repeat steps (3-5) until the year 1200 A.D.

Road Map

Individual Learning Framework Overview

Population Modeling: Syntactic Language Change

- Old English: description & proposed individual filters
- Old English data & feasibility of filtering
- Modeled learners and populations
- Estimating ground truth
- Sufficiency & necessity of filtering

Estimating Historical p_VO

Historical data used to initialize population at 1000 A.D., calibrate population between 1000 and 1150 A.D., and check target state at 1200 A.D.

Historical data distributions: some data are ambiguous

p_VO: underlying distribution is not ambiguous

<table>
<thead>
<tr>
<th></th>
<th>Degree-0 % Ambiguous</th>
<th>Degree-1 % Ambiguous</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000 A.D.</td>
<td>76%</td>
<td>28%</td>
</tr>
<tr>
<td>1000 - 1150 A.D.</td>
<td>80%</td>
<td>25%</td>
</tr>
<tr>
<td>1200 A.D.</td>
<td>71%</td>
<td>10%</td>
</tr>
</tbody>
</table>

Observations:
(1) Degree-1 data less ambiguous than degree-0 data.
(2) Advantage is magnified in degree-1.

Assumption: degree-1 distribution less distorted from underlying distribution.
Estimating Historical \( p_{VO} \)

Use the difference in distortion between the **degree-0** and **degree-1** unambiguous data distributions to estimate the difference in distortion between the **degree-1** distribution and the underlying unambiguous data distribution in a speaker’s mind.

\[
\begin{align*}
\log \frac{d_0}{d_1} &= \log \frac{(1 - u_1d_1')}{u_2d_1'}
\end{align*}
\]

\[
\begin{align*}
\log \frac{d_0}{d_1} &= \log \frac{(1 - u_1d_1')}{u_2d_1'}
\end{align*}
\]

**Known quantities:**

- Unambiguous and ambiguous data in \( d_0 \) and \( d_1 \)
- Normalize \( d_1 \) to \( d_0 \) distribution: estimate how much \( d_1 \) unambiguous data was “lost” in \( d_0 \)
Known quantities:
Unambiguous and ambiguous data in d0 and d1

Normalize d1 to d0 distribution: estimate how much d1 unambiguous data was “lost” in d0

Calculate OV to VO "loss ratio"
Assumption: \( \approx \)

- **OV Unamb**
- **Amb**
- **VO Unamb**

**D0**

- **OV Unamb**
- **Amb**
- **VO Unamb**

**D1**

- **OV Unamb**
- **Amb**
- **VO Unamb**

**U**

**Estimating Historical \( p_{VO} \)**

**Known quantities:**
- Unambiguous and ambiguous data in \( d_0 \) and \( d_1 \)

**Normalize \( d_1 \) to \( d_0 \) distribution:** estimate how much \( d_1 \) unambiguous data was "lost" in \( d_0 \)

**Calculate OV to VO "loss ratio"**

**Assume \( d_1 \)-to-\( d_0 \) "loss ratio" is same as underlying-to-\( d_1 \) "loss ratio"**

**Use "loss ratio" to estimate how much underlying unambiguous data was "lost" in \( d_1 \)"
Estimating Historical $p_{VO}$

**Known quantities:**
- Unambiguous and ambiguous data in $d_0$ and $d_1$
- Normalize $d_1$ to $d_0$ distribution: estimate how much $d_1$ unambiguous data was "lost" in $d_0$
- Calculate $p_{VO}$ from estimated underlying unambiguous data distribution
- Use "loss ratio" to estimate how much underlying unambiguous data was "lost" in $d_1$

**Calculate $p_{VO}$**
- $p_{VO} = \frac{\text{Underlying unamb VO} \times \text{D1 to D0 loss ratio}}{\text{Underlying unamb VO} \#}$

Assume $d_1$ to $d_0$ "loss ratio" is same as underlying-to-$d_1$ "loss ratio".
Estimating Historical $p_{\text{VO}}$

<table>
<thead>
<tr>
<th></th>
<th>(Initialization) 1000 A.D.</th>
<th>(Calibration) 1000-1150 A.D.</th>
<th>(Termination) 1200 A.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average $p_{\text{VO}}$</td>
<td>0.234</td>
<td>0.310</td>
<td>0.747</td>
</tr>
</tbody>
</table>

Road Map

Individual Learning Framework Overview

Population Modeling: Syntactic Language Change
- Old English: description & proposed individual filters
- Old English data & feasibility of filtering
- Modeled learners and populations
- Estimating ground truth
- Sufficiency & necessity of filtering

Remaining Questions to Answer

(1) **sufficiency**: Can an Old English population whose learners filter their intake down to the degree-0 **unambiguous data** shift at the correct rate?

(2) **necessity**: If the proposed individual filtering during learning is sufficient to cause an Old English population to change at the correct rate, is it in fact necessary? Are the filters responsible?

Sufficiency of Filters: Correct Distribution Biases

![Graph showing Avg $p_{\text{VO}}$ in Population Over Year A.D.](image)
Necessity of Filters: Remove Unambiguous Filter
Learner can use ambiguous data. Strategy: assume surface order is actual order. (Fodor, 1998)
Example: Subject TensedVerb Object = VO

<table>
<thead>
<tr>
<th></th>
<th>Degree-0 OV Advantage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000 A.D.</td>
<td>-21.0%</td>
</tr>
<tr>
<td>1000 - 1150 A.D.</td>
<td>-26.9%</td>
</tr>
<tr>
<td>1200 A.D.</td>
<td>-21.8%</td>
</tr>
</tbody>
</table>

VO order has advantage, even at 1000 A.D.

Necessity of Filters: Removing Degree-0 Filter
Learner can use unambiguous data in both degree-0 and degree-1 clauses.
Necessity of Filters:
Removing Degree-0 Filter
Learner can use unambiguous data in both degree-0 and degree-1 clauses.

<table>
<thead>
<tr>
<th></th>
<th>OV Advantage in Unamb D0</th>
<th>OV Advantage in Unamb D1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000 A.D.</td>
<td>19.5%</td>
<td>41.7%</td>
</tr>
<tr>
<td>1000-1150 A.D.</td>
<td>2.8%</td>
<td>28.7%</td>
</tr>
<tr>
<td>1200 A.D.</td>
<td>-2.7%</td>
<td>-45.2%</td>
</tr>
</tbody>
</table>

Degree-1 data is strongly OV-biased. What is the threshold of permissible % of degree-1 data so the population can still be strongly VO-biased by 1200 A.D.? How does this compare to the amount available to children?

Permissible Threshold: <4% degree-1 data in individual intake.

Necessity of Filters:
Removing Degree-0 Filter
Learner can use unambiguous data in both degree-0 and degree-1 clauses.

<table>
<thead>
<tr>
<th></th>
<th>OV Advantage in Unamb D0</th>
<th>OV Advantage in Unamb D1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000 A.D.</td>
<td>19.5%</td>
<td>41.7%</td>
</tr>
<tr>
<td>1000-1150 A.D.</td>
<td>2.8%</td>
<td>28.7%</td>
</tr>
<tr>
<td>1200 A.D.</td>
<td>-2.7%</td>
<td>-45.2%</td>
</tr>
</tbody>
</table>

Estimated amount available to children (from corpora): ~16%

Conclusion: Filter required so that 16% degree-1 data does not cause Old English population to be too OV-biased.
Necessity of Filters: Removing Degree-0 Filter

Permissible threshold: <4%

Estimated amount available to children (from corpora): ~16%

Conclusion: Filter required so that 16% degree-1 data does not cause Old English population to be too OV-biased

Necessity of Filters: Removing Both Filters

Dropping Unambiguous Data Filter: too much VO
   (change is too fast)
Dropping Degree-0 Filter: too much OV
   (change is too slow)

Drop both?

Necessity of Filters: Removing Both Filters

Dropping Unambiguous Data Filter: too much VO
   (change is too fast)
Dropping Degree-0 Filter: too much OV
   (change is too slow)

Drop both?

<table>
<thead>
<tr>
<th></th>
<th>OV Advantage in D0</th>
<th>OV Advantage in D1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000 A.D.</td>
<td>-21.0%</td>
<td>28.1%</td>
</tr>
</tbody>
</table>

Requires 43% of the intake to be degree-1 data just to get the intake to be OV-biased at 1000 A.D.

Old English Language Change Summary

Language change modeling results: existence proof for feasibility, sufficiency, and necessity of data intake filtering during individual learning

Individual-Level Filters:
1. unambiguous data
2. degree-0 data

There is an interaction of language change modeling and language learning theory. Each can be used to constrain the other.
Open Questions

(1) If we add complexity to the population model, do we still need these individual-level learning filters?

Weight data points in individual intake using various factors:

(a) spatial location of speaker with respect to learner
(b) social status of speaker
(c) speaker’s relation to learner (family, friend, stranger)
(d) context of data point (social context, linguistic context)

(2) Are these filters necessary if we look at other language changes where individual-level learning is thought to be the main factor driving change at the population-level?

Population Modeling: Take Home Messages

(1) Correct population-level behavior can result from correct individual-level behavior (small misconvergences compounded over time).

(2) Learners can extract the correct systematicity by looking at a subset of the data.

(3) Models of language change can (and should) be empirically grounded, with learners searching through realistic data distributions.
Population Modeling: Take Home Messages

(1) Models of language change can (and should) be empirically grounded.
   Individual-level: learning period, data distribution, linguistic representation, probabilistic learning
   Population-level: population size, population growth rate, time period of change, rate of change

(2) Learners can extract the correct system by looking at a subset of the data.

(3) Correct population-level behavior can result from correct individual-level behavior (small misconvergences compounded over time).

Thank You

Amy Weinberg
Bill Idsardi
Colin Phillips
Elizabeth Royston
Raven Alder

Jeff Lidz
Charles Yang
Norbert Hornstein
Philip Resnik
David Poeppel

the Cognitive Neuroscience of Language Lab
at the University of Maryland
Causes of Language Change

Old Norse influence before 1000 A.D.: VO-biased
If sole cause of change, requires exponential influx of Old Norse speakers.

Old French at 1066 A.D.: embedded clauses
predominantly OV-biased (Kibler, 1984)
Matrix clauses often SVO (ambiguous)
OV-bias would have hindered Old English change to VO-biased system.

Evidence of individual probabilistic usage in Old English
Historical records likely not the result of subpopulations of speakers who use only one order

Deriving the Bayesian Update Equations for a Hypothesis Space with 2 Hypotheses

Max(Prob(\(p_{VO} \mid u\))) = Max(\(\frac{\text{Prob}(u \mid p_{VO}) \times \text{Prob}(p_{VO})}{\text{Prob}(u)}\))
Bayes’ Rule, find maximum of a posteriori (MAP) probability
Manning & Schütze (1999)

\[
\frac{\text{d}}{\text{dp}_{VO}} \left( \frac{ \binom{n}{r} \times p_{VO}^r \times (1-p_{VO})^{n-r} }{\text{Prob}(u)} \right) = 0 \\
\frac{\text{d}}{\text{dp}_{VO}} (\text{Prob}(u)) = 0 \quad (\text{P}(u) \text{ is constant with respect to } p_{VO})
\]
\[
p_{VO} = \frac{r+1}{n+1}
\]
Deriving the Bayesian Update Equations for a Hypothesis Space with 2 Hypotheses

\[ p_{VO} = \frac{r+1}{n+1}, \quad r = p_{VOprev} \times n \]

Replace 1 in numerator and denominator with
\[ c = p_{VOprev} \times m \text{ if VO, } c = (1 - p_{VOprev}) \times m \text{ if OV} \]

\[ 3.0 \leq m \leq 5.0 \]

\[ p_{VO} = \frac{p_{VOprev} \times n + c}{n + c} \]

Other Ways to Remove the Unambiguous Filter

Strategies for assessing ambiguous data
(1) assume base-generation
   - attempted and failed
   - system-dependent (syntax)

(2) weight based on level of ambiguity (Pearl & Lidz, in submission)
   - unambiguous = highest weight
   - moderately ambiguous = lower weight
   - fully ambiguous = lowest weight (ignore)

(3) randomly assign to one hypothesis (Yang, 2002)

Perceived Unambiguous Data: OV

Unambiguous OV data

(1) Tensed Verb is immediately post-Object

he\_subj hyne\_obj ge\_bidde\_tense\_verb
He him may-pray
‘He may pray (to) him’
(Ælfric’s Letter to Wulfige, 87.107, ~1075 A.D.)
Perceived Unambiguous Data: OV
Unambiguous OV data
(1) Tensed Verb is immediately post-Object

he₂Sub₃ hyn₂Sub₄ gebidd₁TensedVerb
He him may-pray
‘He may pray (to) him’
(Ælfric’s Letter to Wulfsige, 87.107, ~1075 A.D.)

(2) Verb-Marker is immediately post-Object

we₂Sub₃ sculë₁VerbMarker [ure yfele₁VerbMarker]₅Verb-Marker
we should our evil practices abandon
‘We should abandon our evil practices.’
(Alcuin’s De Virtutibus et Vitiis, 70.52, ~1150 A.D.)

Perceived Unambiguous Data: VO
Unambiguous VO data
(1) Tensed Verb is immediately pre-Object, 2+ phrases precede (due to interaction of V2 movement)

& [mid his stefne]₄PP he₂Sub₃ awede₆TensedVerb deade₇Obj [to life]₉PP
& with his stem he awakened the-dead to life
‘And with his stem, he awakened the dead to life.’
(James the Greater, 30.31, ~1150 A.D.)

(2) Verb-Marker is immediately pre-Object

Paulus₂Sub₃ ahof₄TensedVerb [his heafod]₅Obj
Paulus then lifted his head up.
(Blickling Homilies, 187.35, between 900 and 1000 A.D.)
Verb-Markers

Sub-piece of the verbal complex that is semantically associated with a Verb, used to determine original position of Verb.

Examples: particle (‘up’, ‘out’), a non-tensed complement to tensed Verbs, a closed-class adverbial (‘never’), or a negative (‘not’) (Lightfoot, 1991).

Unreliable Verb-Markers

Sometimes the Verb-Marker would not remain adjacent to the Object.

---

Verb-Markers

Sub-piece of the verbal complex that is semantically associated with a Verb, used to determine original position of Verb.

Examples: particle (‘up’, ‘out’), a non-tensed complement to tensed Verbs, a closed-class adverbial (‘never’), or a negative (‘not’) (Lightfoot, 1991).

Unreliable Verb-Markers

Sometimes the Verb-Marker would not remain adjacent to the Object.

---

Verb-Markers

Sub-piece of the verbal complex that is semantically associated with a Verb, used to determine original position of Verb.

Examples: particle (‘up’, ‘out’), a non-tensed complement to tensed Verbs, a closed-class adverbial (‘never’), or a negative (‘not’) (Lightfoot, 1991).

Unreliable Verb-Markers

Sometimes the Verb-Marker would not remain adjacent to the Object.

---

Verb-Markers

Sub-piece of the verbal complex that is semantically associated with a Verb, used to determine original position of Verb.

Examples: particle (‘up’, ‘out’), a non-tensed complement to tensed Verbs, a closed-class adverbial (‘never’), or a negative (‘not’) (Lightfoot, 1991).

Unreliable Verb-Markers

Sometimes the Verb-Marker would not remain adjacent to the Object.