Stop the morphological cycle, I want to get off:
Modeling the development of fusion

Micha Elsner, Martha Booker Johnson, Stephanie Antetomaso and Andrea D. Sims

The Ohio State University
An old idea: change in morphological typology

Over time, morphologically agglutinative structures can become fusional…

But this doesn’t always happen!

Which language features accelerate these changes?

Which ones slow or stop them?
# Agglutination vs. fusion

| Indo-European (reconstructed) | Ancient Greek ("give") |
|------------------------------|-------------------------|
|                              | PRS | AOR       | PRS     | AOR       |
| 1SG                          | -m-i | -m        | 1SG     | dídō-mi   | edō-n |
| 2SG                          | -s-i | -s        | 2SG     | dídō-s    | edō-s |
| 3SG                          | -t-i | -t        | 3SG     | dídō-si   | edō   |

**PRS**: Present tense, **AOR**: Aorist tense.
### Agglutination vs. Fusion

| Indo-European (reconstructed) | Ancient Greek (“give”) |
|-------------------------------|------------------------|
| **PRS** | **AOR** | **PRS** | **AOR** |
| 1SG  | -m-i  | -m   | dídō-mi | edō-n  |
| 2SG  | -s-i  | -s   | dídō-s  | edō-s  |
| 3SG  | -t-i  | -t   | dídō-si | edō    |
But what is fusion, really?

Plank (1999) lists many, many properties

**Typical of agglutination**
- identifiable exponents
- no inflection classes
- no syncretism
- zero and multiple exponence
- large paradigms
- weak phonological cohesion
- many optional elements

**Typical of fusion**
- fused exponents
- inflection classes
- plentiful syncretism
- little zero/multiple exponence
- small paradigms
- strong phonological cohesion
- few optional elements
But what *is* fusion, really?

Plank (1999) lists many, many properties

Typical of agglutination

- identifiable exponents
- no inflection classes
- no syncretism
- zero and multiple exponence
- large paradigms
- weak phonological cohesion
- many optional elements

Typical of fusion

- fused exponents
- inflection classes
- plentiful syncretism
- little zero/multiple exponence
- small paradigms
- strong phonological cohesion
- few optional elements
A huge list isn’t very satisfying

These properties don’t always go together…

Or even usually go together (Haspelmath 2009, for instance)

It’s not clear which of them are causes and which are effects

So it’s hard to frame hypotheses about why they may, or may not, coincide for a particular inflection in a particular language
History is the missing piece

By understanding how these systems arise over time, we can see whether some properties contribute to or prevent the development of others

(see Nichols 1992, Harris 2008, Murawaki 2018 and others)
One process of change

Indo-European (reconstructed)

|       | PRS | AOR |
|-------|-----|-----|
| 1SG   | -m-i | -m  |
| 2SG   | -s-i | -s  |
| 3SG   | -t-i | -t  |
### One process of change

|        | Indo-European (reconstructed) | Proto-Greek (reconstructed) |
|--------|-------------------------------|-----------------------------|
|        | PRS  | AOR | PRS  | AOR |
| 1SG    | -m-i | -m  | -m-i | -n  |
| 2SG    | -s-i | -s  | -s   | -s  |
| 3SG    | -t-i | -t  | -t-i | -t  |

(change)
One process of change

Proto-Greek (reconstructed)

|   | PRS  | AOR |
|---|------|-----|
| 1SG | -m-i | -n  |
| 2SG | -s-∅ | -s  |
| 3SG | -t-i | -t  |

Proto-Greek (reconstructed)

|   | PRS  | AOR |
|---|------|-----|
| 1SG | -mi | -n  |
| 2SG | -s  | -s  |
| 3SG | -ti | -t  |
One process of change

Proto-Greek (reconstructed)

|   | PRS | AOR |
|---|-----|-----|
| 1SG | -m-i | -n |
| 2SG | -s-∅ | -s |
| 3SG | -t-i | -t |

reanalysis

Proto-Greek (reconstructed)

|   | PRS | AOR |
|---|-----|-----|
| 1SG | -mi | -n |
| 2SG | -s | -s |
| 3SG | -ti | -t |

...followed by dialectal generalization of -mi, -n to new verbs
The “morphological cycle”

Not the only way morphological fusion can arise…

But one important way:

   Proposed by Schleicher (1850); newer survey in Igartua (2015)

Morphological change is driven by phonology and linear adjacency

Thus, the phonological and linear factors are causal!
Language learners face a choice

### Learning outcome 1

|   | PRS | AOR |
|---|-----|-----|
| 1SG | -m-i | -n |
| 2SG | -s-∅ | -s |
| 3SG | -t-i | -t |

plus **phonological rule**: m# → n

and more...

### Learning outcome 2

|   | PRS | AOR |
|---|-----|-----|
| 1SG | -mi | -n |
| 2SG | -s | -s |
| 3SG | -ti | -t |
What determines the outcome?

Environments in which the morphemes appear

Frequency of the different combinations

Evidence for the rule elsewhere in the language

Bybee (2002): “Items that are used together, fuse together”
This study

A proof-of-concept simulation using artificial data
Bayesian model of learner predicts when reanalysis might occur

What features of change process / input system predict the outcome?
Two specific claims

Case study 1

**Variable realization** (morphological slots which may contain $\emptyset$) preserves **agglutinative** structures: Plank (1999), Comrie (1989)

Case study 2

**Stress-based vowel reduction** encourages **fusional** structures: Zingler (2018)
Model framework

High priority for **interpretability**: we want to examine the output and see if the learned system is fusional or not

Makes neural models (Kann et al. 2016) less appealing; use older-fashioned transducer cascade (Cotterell et al. 2015; Dreyer and Eisner 2008) instead

Everything implemented using Carmel toolkit (Graehl et al. 1997, Chiang et al. 2010)
# The model

| Input | lex="give", tense=PRS, prs/num=1SG |
|-------|-----------------------------------|
| Transducer 1: fusion |
| Abstract morphs | lex="give", tense=PRS | prs/num=1SG |
| Transducer 2: lexicon |
| Underlying | díðō - mi |
| Transducer 3: phonology |
| Surface | díðōmi |
Model biases

Balance between two opposing pressures:

Learn a small inventory of morphemes (pressure for more agglutinative analyses)

Do not learn unnecessary phonological rules (pressure for more fusional analyses)
Case study 1: variable realization

Data loosely based on Kihehe (Bantu, Tanzania; Johnson 2015, Odden and Odden 1999)

\[ \text{twikomala} \]

\[
\begin{array}{ccc}
\text{tu-} & \text{i-} & \text{komala} \\
\text{SM} & \text{PROG} & \text{sit} \\
\end{array}
\]

“we are sitting”

Kihehe verbs are marked for subject agreement (SM) and tense/aspect; phonological rules prevent VV on the surface

Ambiguous: may or may not be fused in speakers’ minds
Why would variable elements matter?

An agglutinative analysis has \((\text{num SMs} + \text{num TMA})\) morphemes;

A fused analysis has \((\text{num SMs} \times \text{num TMA})\) morphemes

\[ \text{twikomala} \]

\[
\begin{array}{ccc}
\text{tu-} & \text{i-} & \text{komala} \\
\text{SM} & \text{PROG} & \text{sit} \\
\end{array}
\]

“we are sitting”
What if the template had more slots?

tu- ??- i- komala

SM ?? PROG sit

More combinations to memorize; fused analysis less appealing
Polysynthesis: many variably-filled slots

Choguita Raramuri: Uto-Aztecan, Mexico (Caballero 2008)
Simulations with artificial data

Each language has:

200 random CV stems

1000 word forms
Model validity check

### Table A: Agglutinative

| $M_1$ | $M_2$ | $M_3$ | \(\text{lex}\) | \(\text{under}\) | \(\text{surface}\) |
|-------|-------|-------|-------------|----------------|------------------|
| ta    | i     | mela  | ko-i-mela   | koimela        |                  |
| ko    | a     | tano  | mu-i-mela   | muimela        |                  |
| he    | de    |       |              |                |                  |
| mu    | no    |       |              |                |                  |
| gu    |       |       |              |                |                  |
| si    |       |       |              |                |                  |

### Table B: Fusional

| $M_1$ | $M_3$ | \(\text{lex}\) | \(\text{under}\) | \(\text{surface}\) |
|-------|-------|-------------|----------------|------------------|
| ya, se, dunu, lanu | mela | ya-mela | yamela |                  |
| ha, hi, si, yu...   | tano | dunu-mela | dunumela |                  |
|                   |      |            |              |                  |
As expected...

Langue A
- agglutinative

Langue B
- fusional

Fusion type
- M1-M3
- M1|M3
Add some phonology

Prior weight sets bias against phonological rule

| $M_1$ | $M_2$ | $M_3$ | lex   | under          | surface       |
|-------|-------|-------|-------|----------------|---------------|
| ta    | i     | mela  | ko-i-mela | ko-i-mela      | mwi-mela      |
| ko    | a     | tano  | mu-i-mela |               |               |
| he    | de    | ...   |       |                |               |
| mu    | no    |       |       |                |               |
| gu    |       |       |       |                |               |
| si    |       |       |       |                |               |

ambiguous

high V + V $\rightarrow$ glide + V
low V + V $\rightarrow$ _ + V
Bias determines outcome

![Graph showing the relationship between posterior and prior weight for different fusion types.](image)

- **Fusion type**
  - M1-M3
  - M1|M3

**Rule of thumb:**
- *Rule is ok* if the posterior is high for the desired condition.
- *Don’t learn rule* if the posterior is low for the desired condition.
Fix prior, then vary morphological template

rule is ok

Prior weight

don’t learn rule

Fusion type

- M1-M3
- M1|M3
Add a variable element

M₁  M₂  M₃  lex

D

ambiguous

ta  sa  i  mela  mu-i-mela
ko  ø   a  tano  mu-sa-i-mela
he  de  no  ...
mu  gu  si

M₂ filled with variable probability

under

surface

high V + V → glide + V
low V + V → _ + V

mwimela
musimela
Categorical systems: full fusion

Graph showing the posterior probabilities against the probability of M2, with different fusion types indicated:
- M1-M2-M3
- M1-M2|M3
- M1|M2-M3
- M1|M2|M3

Key areas highlighted in red.
Variable $M_2$: partial or no fusion

Fusion type:
- $M1-M2-M3$
- $M1-M2|M3$
- $M1|M2-M3$
- $M1|M2|M3$
Interim conclusion

Adding variable template slots preserves agglutination

Even when outcome without them would be fusional

When fusion does result, tends to be more local (as in Caballero and Kapatsinski 2019)

Effect critically depends on variability, not just extra slots
Case study 2: stress-based reduction

Zingler (2018) argument for maintenance of Turkish agglutination: **vowel harmony** prevents **stress-based reduction**, which would in turn lead to more fusional system

The link between harmony and reduction is unclear... but we’ll focus on the second claim, that reduction leads to fusion
Why reduction?

Reducing vowels forces consonants into contact…

Result: phonological interactions across morpheme boundaries

Obscures the true underlying forms of morphemes

Especially when reduction effect targets the same syllables each time
**Language E has final stress**

| lex  | $M_1$ | $M_2$ | underlying |
|------|-------|-------|------------|
| dite | ta    | pi    | dite-ko-de |
| ...  | ko    | ka    |            |
| he   | de    | no    |            |

*E* final stress
Language E has final stress

Assign **SS** stress from right

Because there are always two monosyllabic suffixes, this stress system ensures $M_1$ will be unstressed and $M_2$ will be stressed.
Language E has final stress

underlying  
dite-ko-de  

Assign **SS** stress from right

Probabilistically reduce weak vowels

surface  
dtekde  
ddekte

Voicing assimilation

Test variants with different probability of reduction
Reduction encourages fusion

![Graph showing the relationship between posterior probability and probability of reduction for features M1-M2 and M1|M2 with reduction.](image)
Probability of reduction is underestimated
Reduction is “baked into” the lexicon

| % reduction | $M_1=1$ | $M_2=3$ | $M_1=1 \mid M_2=3$ |
|--------------|----------|----------|-------------------|
|              | (ta)     | (de)     | (ta-de)           |
| 0            | ta       | de       | -                 |
| 25           | ta (t, te)| de (te) | tade              |
| 50           | ta (t)   | te (de)  | tade              |
| 75           | ta (ti, t)| te (de) | tte               |
| 100          | -        | -        | tte               |
What if stress placement is less predictable?

Occurs in real languages with lexical stress or some kinds of predictable stress systems…

Language with initial stress:

Even and odd-length stems place different stress on suffixes
Language F has initial stress

underlying: ditko-de

Assign SS stress from left

Probabilistically reduce weak vowels

Voicing assimilation

ditkod

ditkod

test variants with different probability of reduction
Unpredictable stress: less fusion

![Graph showing the relationship between posterior and probability of reduction. The graph includes three lines: M1-M2 in black, M1|M2 in dotted black, and Red. in orange. The x-axis represents the probability of reduction, ranging from 0 to 100, and the y-axis represents the posterior, ranging from 0.00 to 1.00.]
| Plank’s list, revisited |
|-------------------------|
| **Typical of agglutination** |
| identifiable exponents |
| no inflection classes |
| no syncretism |
| zero and multiple exponentence |
| large paradigms |
| weak phonological cohesion |
| many optional elements |
| **Typical of fusion** |
| fused exponents |
| inflection classes |
| plentiful syncretism |
| little zero/multiple exponentence |
| small paradigms |
| strong phonological cohesion |
| few optional elements |
Conclusion

Typical of agglutination

- identifiable exponents
- no inflection classes
- no syncretism
- zero and multiple exponentce
- large paradigms
- weak phonological cohesion
- many optional elements
Conclusion

But we can also understand why (per Haspelmath and others) “agglutinative” and “fusional” features don’t always cluster…

Many ways for fusion to arise historically

Our model addresses only one mechanism
Future work

Test the model on data from real historical corpora!

Look at other language features (like the rest of Plank’s list)

More generally: historical explanations for typological correlations (Harris 2008 and others), combined with models of the learner
Thank you!

And thanks to:

Fall 2018 Seminar on Models of Morphological Learning and Change at Ohio State University

Brian Joseph

Three anonymous reviewers