

# Navigable atom-rule interactions in PSL models enhanced by rule verbalizations, with an application to etymological inference

Verena Blaschke, Thora Daneyko, Jekaterina Kaparina,  
Zhuge Gao & Johannes Dellert

ILP 2022, Windsor (UK), September 29, 2022

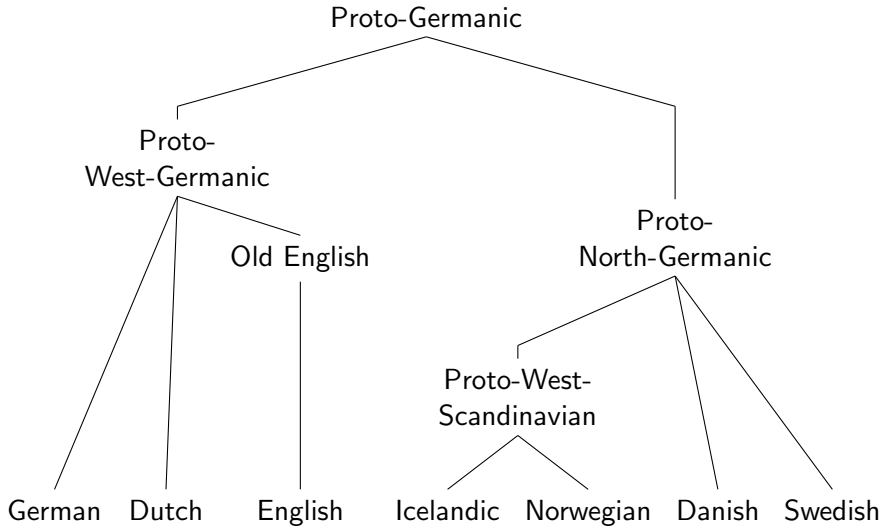
University of Tübingen

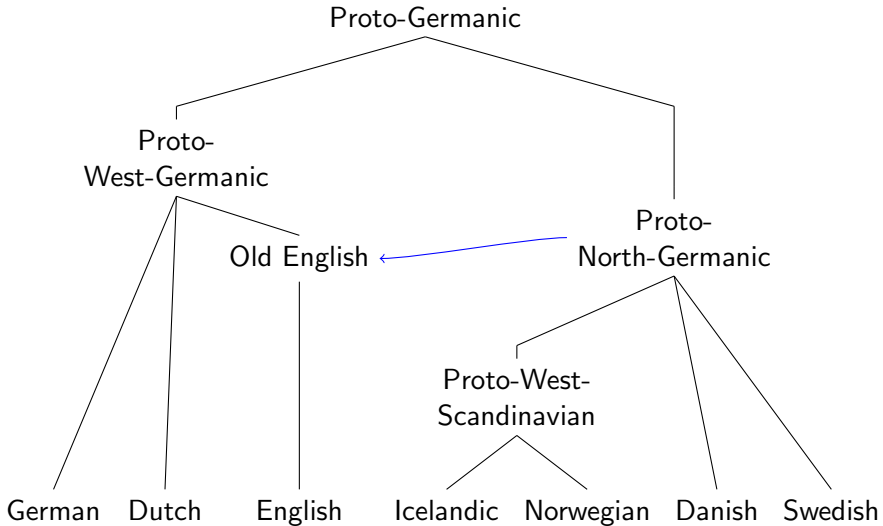
# Probabilistic soft logic<sup>1</sup>

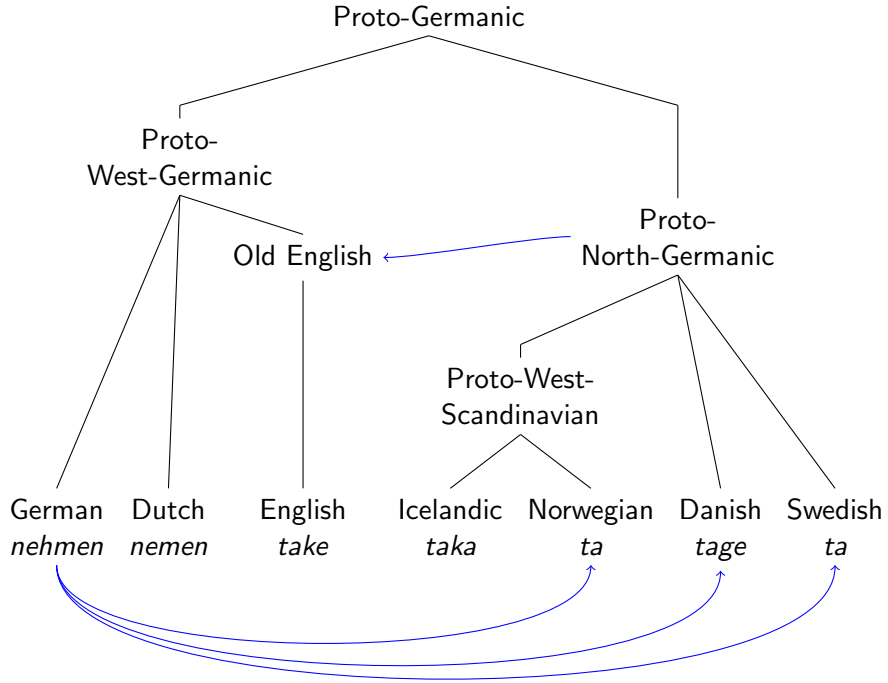
- ▶ Templating language for a probabilistic graphical model (hinge-loss Markov random fields)
- ▶ First-order logic + statistics
- ▶ Used for modelling relation graphs – in our case in the context of historical linguistics

---

<sup>1</sup>Bach, S. H., Broecheler, M., Huang, B., & Getoor, L. (2017). “Hinge-Loss Markov Random Fields and Probabilistic Soft Logic.” *JMLR* **18**(109)







## PSL atoms

- ▶ **Atom**

Inherited(X,Y), Borrowed(X,Y), UnknownOrigin(X),  
Similar(X,Y), ...

- ▶ **Ground atom**

Inherited("English 'take'", "Old English '...')")

Borrowed("Old English '...'",  
          "Proto-North-Germanic '...')")

UnknownOrigin("Proto-West-Germanic '...')")

Similar("English 'take'", "Norwegian 'ta'") = 0.7

...

→ We want to assign scores  $\in [0; 1]$  to the ones whose values  
are initially undefined

## PSL rules

- ▶ **Logical rules**

Disjuncts of literals, can be written as implications of the form

$$P1(A,B) \ \& \ P2(A,B) \ \rightarrow \ P3(A,B) \ \mid \ P4(A,B)$$
$$\text{Inherited}(X,Z) \ \& \ \text{Inherited}(Y,Z) \ \& \ (X \neq Y) \\ \rightarrow \text{Similar}(X,Y)$$

- ▶ **Arithmetic rules:** (in)equalities

$$\text{Similar}(X,Y) = \text{Similar}(Y,X) \ .$$
$$\text{Borrowed}(X,Y) + \text{Borrowed}(Y,X) \leq 1 \ .$$
$$\text{Inherited}(X,+Y) + \text{Borrowed}(X,+Z) \\ + \text{UnknownOrigin}(X) = 1 \ .$$

- ▶ **Ground rules:** where all atoms are ground atoms

$$\text{Similar}(\text{"German 'nehmen'"}, \text{"Dutch 'nemen'"}) = \\ \text{Similar}(\text{"Dutch 'nemen'"}, \text{"German 'nehmen'"}) \ .$$

## Distance to satisfaction

- ▶ Arithmetic rules are **satisfied** if their (in)equalities are fulfilled
- ▶ Logical rules are **satisfied** if the consequent's score is at least as high as the antecedent's score:
  - ▶ Łukasiewicz logic, e.g. conjunction:  
 $A \& B := \max\{A + B - 1, 0\}$

```
Inherited(X,Z) & Inherited(Y,Z) & (X != Y)
-> Similar(X,Y)
```

```
Inherited("German 'nehmen'", "P-W-Gmc '...'")    0.9
& Inherited("Dutch 'nemen'", "P-W-Gmc '...'")    0.8
& ("German 'nehmen'" != "Dutch 'nemen'")        1.0
-> Similar("German 'nehmen'", "Dutch 'nemen'")    0.9
```



## Distance to satisfaction

- ▶ Arithmetic rules are **satisfied** if their (in)equalities are fulfilled
- ▶ Logical rules are **satisfied** if the consequent's score is at least as high as the antecedent's score:
  - ▶ Łukasiewicz logic, e.g. conjunction:  
 $A \& B := \max\{A + B - 1, 0\}$

Inherited(X,Z) & Inherited(Y,Z) & (X != Y)  
-> Similar(X,Y)

Inherited("German 'nehmen'", "P-W-Gmc '...'" )	0.9	} 0.7
& Inherited("Dutch 'nemen'", "P-W-Gmc '...'" )	0.8	
& ("German 'nehmen'" != "Dutch 'nemen'" )	1.0	
-> Similar("German 'nehmen'", "Dutch 'nemen'" )	0.9	≤



## Distance to satisfaction

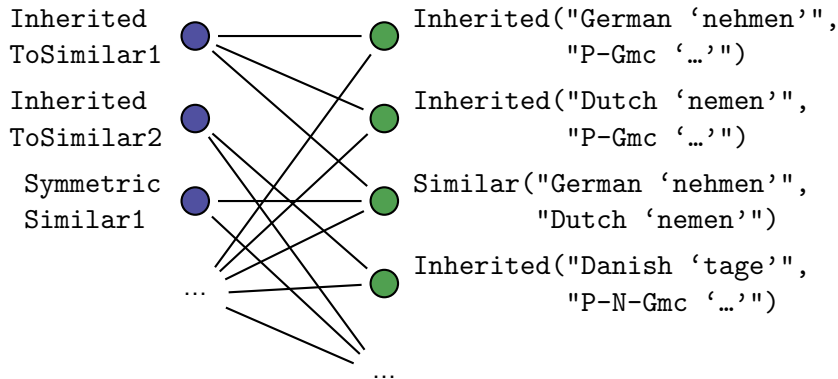
```
Inherited("Icelandic 'taka'", "P-W-Sca '...'" ) 0.9 }  
& Inherited("Norwegian 'ta'", "P-W-Sca '...'" ) 0.9 } 0.8  
& ("Icelandic 'taka'" != "Norwegian 'ta'" ) 1.0 } ≤  
-> Similar("Icelandic 'taka'", "Norwegian 'ta'" ) 0.7  
X
```

- ▶ **Distance to satisfaction:** *How* unsatisfied is a rule?  
 $\max\{\text{antecedent score} - \text{consequent score}, 0\}$   
(arithmetic constraints are a natural extension)

## Distance to satisfaction

- ▶ **Distance to satisfaction:** *How* unsatisfied is a rule?  
 $\max\{\text{antecedent score} - \text{consequent score}, 0\}$   
(arithmetic constraints are a natural extension)
- ▶ Rules can be constraints (that should always be satisfied) or weighted according to how much their dissatisfaction should be penalized
  - ▶ Rule weights can be learned!
- ▶ PSL inference goal: assign atom values in such a way that the weighted sum of distances to satisfaction is minimized (MAP estimate)

## Rule-atom graph



Very large → hard to interpret

# Existing debugging tool: VMI-PSL<sup>2</sup>

Rule Overview			
ID	Rule	Weighted Count	Total Dissatisfaction / Mean Dissatisfaction
1	SGD_RATING(U, I) → RATING(U, I)	true 19799	567.63 / 0.03
2	USER(U) & AVG_USER_RATING(U) & ITEM(I) & RATED(U, I) → RATING(U, I)	true 15120	721.17 / 0.05
3	RATING(U, I) → SGD_RATING(U, I)	true 18623	692.60 / 0.04
4	ITEM_PEARSON_RATING(U, I) → RATING(U, I)	true 19769	369.31 / 0.02
5	RATING(U1, I) & RATED(U1, I) & RATED(U2, I) & SIM_PEARSON_USERS(U1, U2) → RATING(U2, I)	true 14505	1137.08 / 0.08
6	RATING(U1, I) & RATED(U1, I) & USERS_ARE_FRIENDS(U1, U2) & RATED(U2, I) → RATING(U2, I)	true 39662	2629.30 / 0.07
7	BPMF_RATING(U, I) → RATING(U, I)	true 19809	312.74 / 0.02
8	USER(U) & ITEM(I1) & RATED(U, I1) & ITEM(I2) & RATED(U, I2) & SIM_MF_EUCLIDEAN_ITEMS(I1, I2) & RATING(U, I1) → RATING(U, I2)	true 11623	797.98 / 0.07
9	USER(U) & ITEM(I1) & RATED(U, I1) & ITEM(I2) & RATED(U, I2) & SIM_MF_COSINE_ITEMS(I1, I2) & RATING(U, I1) → RATING(U, I2)	true 11111	754.27 / 0.07
10	RATED(U, I1) & RATED(U, I2) & SIM_ADJCS_ITEMS(I1, I2) & RATING(U, I1) → RATING(U, I2)	true 0	0.00 / 0.00
11	USER(U) & RATING(U, I) & ITEM(I) & RATED(U, I) → AVG_USER_RATING(U)	true 18159	4348.77 / 0.24
12	USER(U1) & RATING(U1, I) & RATED(U1, I) & SIM_MF_EUCLIDEAN_USERS(U1, U2) & ITEM(I) & RATED(U2, I) & USER(U2) → RATING(U2, I)	true 5678	414.48 / 0.07
13	RATING(U, I) → ITEM_PEARSON_RATING(U, I)	true 15051	484.80 / 0.03
14	USER(U1) & RATING(U1, I) & RATED(U1, I) & ITEM(I) & RATED(U2, I) & SIM_MF_COSINE_USERS(U1, U2) & USER(U2) → RATING(U2, I)	true 5892	446.23 / 0.08
15	RATING(U1, I) & RATED(U1, I) & SIM_COSINE_USERS(U1, U2) & RATED(U2, I) → RATING(U2, I)	true 22874	1431.46 / 0.06
16	RATING(U, I) → BPMF_RATING(U, I)	true 19807	559.84 / 0.03
17	RATED(U, I1) & SIM_PEARSON_ITEMS(I1, I2) & RATED(U, I2) & RATING(U, I1) → RATING(U, I2)	true 20179	1394.86 / 0.07

- ▶ (Largely) focused on analyzing rules (aggregate (dis)satisfaction, number of groundings)
- ▶ Our approach: focus on ground atoms (examine atoms with surprising/undesired inferred values)

<sup>2</sup>Rodden, A., Salh, T., Augustine, E., & Getoor, L. (2020). "VMI-PSL: Visual Model Inspector for Probabilistic Soft Logic." *RecSys '20*, pp. 604–606. Graphic via supplementary material (demo video)

# Our RAG inspection tool

## Inherited("Old English '...'", "P-West-Germanic '...')"

The interface displays a list of nodes on the left and a detailed view of a selected node on the right.

**Left Panel (List of Nodes):**

- Einh(PWSca for TAKE, PNGer for TAKE)
- Einh(Icelandic (taka) /tʰaka/, PWSc)
- Einh(English (take) /teɪk/, PEng for TAKE)
- Einh(Dutch (nemen) /ne:ma/, PWG)
- Einh(German (nehmen) /ne:mən/, F)
- Einh(Danish (tage) /tɑː/, PNGer for TAKE)
- Einh(Swedish (ta) /taː/, PNGer for TAKE)
- Einh(PWGer for TAKE, PGer for TAKE)
- Einh(PNGer for TAKE, PGer for TAKE)
- Einh(Norwegian (ta) /taː/, PWSca for TAKE)
- Einh(PEng for TAKE, PWGer for TAKE)

**Right Panel (Detailed View):**

**Header:** Einh(PEng for TAKE, PWGer for TAKE) 8 %

**Text:** The Old English form for TAKE is probably not inherited from the Proto-West-Germanic form for TAKE.

**Why is the value not higher?**

- EinhOrEloaOrEunk:  $1.0 * \text{Eunk}(\text{PEng for TAKE}) [0.00] + 1.0 * \text{Eloa}(\text{PEng for TAKE, PNGer for TAKE}) [0.92] + 1.0 * \text{Eloa}(\text{PEng for TAKE, <ctrlArgForGrounding>}) [0.00] + 1.0 * \text{Einh}(\text{PEng for TAKE, PWGer for TAKE}) [0.08] = 1.0$
- EinhToFhom:  $\sim(\text{Fhom}(\text{PWGer for TAKE, German (nehmen) /ne:mən/}) [0.91]) | \sim(\text{Einh}(\text{PEng for TAKE, PWGer for TAKE}) [0.08]) | \text{Fhom}(\text{PEng for TAKE, German (nehmen) /ne:mən/}) [0.00]$
- EinhToFhom:  $\sim(\text{Fhom}(\text{PWGer for TAKE, English (take) /teɪk/}) [0.09]) | \sim(\text{Einh}(\text{PEng for TAKE, PWGer for TAKE}) [0.08]) | \text{Fhom}(\text{PEng for TAKE, English (take) /teɪk/}) [1.00]$
- EinhToSim ( $\times 2.0$ ):  $\sim(\text{Einh}(\text{PEng for TAKE, PWGer for TAKE}) [0.08]) | \sim(\text{Einh}(\text{Dutch (nemen) /ne:ma/, PWGer for TAKE}) [0.00])$

**Why is the value not lower?**

- EinhOrEloaOrEunk:  $1.0 * \text{Eunk}(\text{PEng for TAKE}) [0.00] + 1.0 * \text{Eloa}(\text{PEng for TAKE, PNGer for TAKE}) [0.92] + 1.0 * \text{Eloa}(\text{PEng for TAKE, <ctrlArgForGrounding>}) [0.00] + 1.0 * \text{Einh}(\text{PEng for TAKE, PWGer for TAKE}) [0.08] = 1.0$
- FhomToEinh ( $\times 0.4$ ):  $\sim(\text{Fhom}(\text{PWGer for TAKE, English (take) /teɪk/}) [0.09]) | \sim(\text{Xinh}(\text{the Old English form for TAKE, the Proto-West-Germanic form for TAKE}) [1.00]) | \sim(\text{Fhom}(\text{PEng for TAKE, English (take) /teɪk/}) [1.00]) | \text{Einh}(\text{PEng for TAKE, PWGer for TAKE}) [0.08]$
- FhomToEinh ( $\times 0.4$ ):  $\sim(\text{Fhom}(\text{PWGer for TAKE, German (nehmen) /ne:mən/}) [0.91]) | \sim(\text{Xinh}(\text{the Old English form for TAKE, the Proto-West-Germanic form for TAKE}) [1.00]) | \sim(\text{Fhom}(\text{PEng for TAKE, German (nehmen) /ne:mən/}) [0.00]) | \text{Einh}(\text{PEng for TAKE, PWGer for TAKE}) [0.08]$

# Atom sorting/filtering

**Einh(PEng for TAKE, PWGer for TAKE)** 8 %

The Old English form for TAKE is probably not inherited from the Proto-West-Germanic form for TAKE.

Why is the value not higher?

- $\text{EinhOrEloaOrEunk: } 1.0 * \text{Eunk(PEng for TAKE)} [0.00] + 1.0 * \text{Eloa(PEng for TAKE, PNGer for TAKE)} [0.92] + 1.0 * \text{Eloa(PEng for TAKE, <ctr(ArgForGrounding>)} [0.00] + 1.0 * \text{Einh(PEng for TAKE, PWGer for TAKE)} [0.08] = 1.0$ .
- $\text{EinhToFhom: } \neg(\text{Fhom(PWGer for TAKE, German (nehmen) /ne:man/}) [0.91]) | \neg(\text{Einh(PEng for TAKE, PWGer for TAKE)} [0.08]) | \text{Fhom(PEng for TAKE, German (nehmen) /ne:man/}) [0.00]}$ .
- $\text{EinhToFhom: } \neg(\text{Fhom(PWGer for TAKE, English (take) /te:k/}) [0.09]) | \neg(\text{Einh(PEng for TAKE, PWGer for TAKE)} [0.08]) | \text{Fhom(PEng for TAKE, English (take) /te:k/}) [1.00]}$ .
- $\text{EinhToFsims} (\times 2.0): \neg(\text{Einh(PEng for TAKE, PWGer for TAKE)} [0.08]) | \neg(\text{Einh(Dutch (nemen) /ne:ma/, PWGer for TAKE)} [0.08])$

Why is the value not lower?

- $\text{EinhOrEloaOrEunk: } 1.0 * \text{Eunk(PEng for TAKE)} [0.00] + 1.0 * \text{Eloa(PEng for TAKE, PNGer for TAKE)} [0.92] + 1.0 * \text{Eloa(PEng for TAKE, <ctr(ArgForGrounding>)} [0.00] + 1.0 * \text{Einh(PEng for TAKE, PWGer for TAKE)} [0.08] = 1.0$ .
- $\text{FhomToEinh} (\times 0.4): \neg(\text{Fhom(PWGer for TAKE, English (take) /te:k/}) [0.09]) | \neg(\text{Xinh(the Old English form for TAKE, the Proto-West-Germanic form for TAKE)} [1.00]) | \neg(\text{Fhom(PEng for TAKE, English (take) /te:k/}) [1.00]) | \text{Einh(PEng for TAKE, PWGer for TAKE)} [0.08]$
- $\text{FhomToEinh} (\times 0.4): \neg(\text{Fhom(PWGer for TAKE, German (nehmen) /ne:man/}) [0.91]) | \neg(\text{Xinh(the Old English form for TAKE, the Proto-West-Germanic form for TAKE)} [1.00]) | \neg(\text{Fhom(PEng for TAKE, German (nehmen) /ne:man/}) [0.00]) | \text{Einh(PEng for TAKE, PWGer for TAKE)} [0.08]$

- Einh(PWSca for TAKE, PNGer for TAKE)
- Einh(Icelandic (taka) /'tʰaka/, PWSca for TAKE)
- Einh(English (take) /te:k/, PEng for TAKE)
- Einh(Dutch (nemen) /ne:ma/, PWGer for TAKE)
- Einh(German (nehmen) /ne:man/, PWSca for TAKE)
- Einh(Danish (tage) /'ta:ʔ/, PNGer for TAKE)
- Einh(Swedish (ta) /ta:/, PNGer for TAKE)
- Einh(PWGer for TAKE, PGer for TAKE)
- Einh(PNGer for TAKE, PGer for TAKE)
- Einh(Norwegian (ta) /ta:/, PWSca for TAKE)
- Einh(PEng for TAKE, PWGer for TAKE)

Einh

# Rules and associated atoms

Ein(PWSc for TAKE, PNGer for TAKE)

Ein(Icelandic (taka) /ˈtaka/, PWSc for TAKE)

Ein(English (take) /tesk/, PEng for TAKE)

Ein(Dutch (nemen) /neːma/, PWGer for TAKE)

Ein(German (nehmen) /neːmən/, Fhom for TAKE)

Ein(Danish (tage) /ˈtaː/, PNGer for TAKE)

Ein(Swedish (ta) /taː/, PNGer for TAKE)

Ein(PWGer for TAKE, PGer for TAKE)

Ein(PNGer for TAKE, PGer for TAKE)

Ein(Norwegian (ta) /taː/, PWSc for TAKE)

Ein(PEng for TAKE, PWGer for TAKE)

Ein(PEng for TAKE, PWGer for TAKE) 8 %

The Old English form for TAKE is probably not inherited from the Proto-West-Germanic form for TAKE.

Why is the value not higher?

- EinOrEloaOrEunk:  $1.0 * \text{Eunk}(\text{PEng for TAKE}) [0.00] + 1.0 * \text{Eloa}(\text{PEng for TAKE, PNGer for TAKE}) [0.92] + 1.0 * \text{Eloa}(\text{PEng for TAKE, <ctrlArgForGrounding>}) [0.00] + 1.0 * \text{Ein}(\text{PEng for TAKE, PWGer for TAKE}) [0.08] = 1.0$ .
- EinToFhom:  $\sim(\text{Fhom}(\text{PWGer for TAKE, German (nehmen) /neːmən/}) [0.91]) | \sim(\text{Ein}(\text{PEng for TAKE, PWGer for TAKE}) [0.08]) | \text{Fhom}(\text{PEng for TAKE, German (nehmen) /neːmən/}) [0.00]$ .
- EinToFhom:  $\sim(\text{Fhom}(\text{PWGer for TAKE, English (take) /tesk/}) [0.09]) | \sim(\text{Ein}(\text{PEng for TAKE, PWGer for TAKE}) [0.08]) | \text{Fhom}(\text{PEng for TAKE, English (take) /tesk/}) [1.00]$ .
- EinToFsim (x2.0):  $\sim(\text{Ein}(\text{PEng for TAKE, PWGer for TAKE}) [0.08]) | \sim(\text{Ein}(\text{Dutch (nemen) /neːma/}, \text{PWGer for TAKE}) [0.08])$ .

Why is the value not lower?

- EinOrEloaOrEunk:  $1.0 * \text{Eunk}(\text{PEng for TAKE}) [0.00] + 1.0 * \text{Eloa}(\text{PEng for TAKE, PNGer for TAKE}) [0.92] + 1.0 * \text{Eloa}(\text{PEng for TAKE, <ctrlArgForGrounding>}) [0.00] + 1.0 * \text{Ein}(\text{PEng for TAKE, PWGer for TAKE}) [0.08] = 1.0$ .
- FhomToEin (x0.4):  $\sim(\text{Fhom}(\text{PWGer for TAKE, English (take) /tesk/}) [0.09]) | \sim(\text{Xinh}(\text{the Old English form for TAKE, the Proto-West-Germanic form for TAKE}) [1.00]) | \sim(\text{Fhom}(\text{PEng for TAKE, English (take) /tesk/}) [1.00]) | \text{Ein}(\text{PEng for TAKE, PWGer for TAKE}) [0.08]$ .
- FhomToEin (x0.4):  $\sim(\text{Fhom}(\text{PWGer for TAKE, German (nehmen) /neːmən/}) [0.91]) | \sim(\text{Xinh}(\text{the Old English form for TAKE, the Proto-West-Germanic form for TAKE}) [1.00]) | \sim(\text{Fhom}(\text{PEng for TAKE, German (nehmen) /neːmən/}) [0.00]) | \text{Ein}(\text{PEng for TAKE, PWGer for TAKE}) [0.08]$ .



# Upward/downward pressure

**Einh(PEng for TAKE, PWGer for TAKE)** 8 %

The Old English form for TAKE is probably not inherited from the Proto-West-Germanic form for TAKE.

**Why is the value not higher?**

- $\text{EinhOrEloaOrEunk: } 1.0 * \text{Eunk(PEng for TAKE)} [0.00] + 1.0 * \text{Eloa(PEng for TAKE, PNGer for TAKE)} [0.92] + 1.0 * \text{Eloa(PEng for TAKE, <ctr(ArgForGrounding>)} [0.00] + 1.0 * \text{Einh(PEng for TAKE, PWGer for TAKE)} [0.08] = 1.0 .$
- $\text{EinhToFhom: } \sim(\text{Fhom(PWGer for TAKE, German (nehmen) /ne:man/) } [0.91]) | \sim(\text{Einh(PEng for TAKE, PWGer for TAKE)} [0.08]) | \text{Fhom(PEng for TAKE, German (nehmen) /ne:man/) } [0.00]) .$
- $\text{EinhToFhom: } \sim(\text{Fhom(PWGer for TAKE, English (take) /tesk/} [0.09]) | \sim(\text{Einh(PEng for TAKE, PWGer for TAKE)} [0.08]) | \text{Fhom(PEng for TAKE, English (take) /tesk/} [1.00]) .$
- $\text{EinhToFsimsim} (*2.0): \sim(\text{Einh(PEng for TAKE, PWGer for TAKE)} [0.08]) | \sim(\text{Einh(Dutch (nemen) /ne:ma/, PWGer for TAKE)} [0.08]) .$

**Why is the value not lower?**

- $\text{EinhOrEloaOrEunk: } 1.0 * \text{Eunk(PEng for TAKE)} [0.00] + 1.0 * \text{Eloa(PEng for TAKE, PNGer for TAKE)} [0.92] + 1.0 * \text{Eloa(PEng for TAKE, <ctr(ArgForGrounding>)} [0.00] + 1.0 * \text{Einh(PEng for TAKE, PWGer for TAKE)} [0.08] = 1.0 .$
- $\text{FhomToEinh} (*0.4): \sim(\text{Fhom(PWGer for TAKE, English (take) /tesk/} [0.09]) | \sim(\text{Xinh(the Old English form for TAKE, the Proto-West-Germanic form for TAKE)} [1.00]) | \sim(\text{Fhom(PEng for TAKE, English (take) /tesk/} [1.00]) | \text{Einh(PEng for TAKE, PWGer for TAKE)} [0.08])$
- $\text{FhomToEinh} (*0.4): \sim(\text{Fhom(PWGer for TAKE, German (nehmen) /ne:man/) } [0.91]) | \sim(\text{Xinh(the Old English form for TAKE, the Proto-West-Germanic form for TAKE)} [1.00]) | \sim(\text{Fhom(PEng for TAKE, German (nehmen) /ne:man/) } [0.00]) | \text{Einh(PEng for TAKE, PWGer for TAKE)} [0.08]$

## Upward/downward pressure

- ▶ Each rule-atom link exerts upward and/or downward pressure on the value of the atom

Inherited(X,Z) & Inherited(Y,Z) & (X != Y)  
-> Similar(X,Y)

- ▶ If unsatisfied:
  - ▶ upward pressure on the value of Similar(X,Y)
  - ▶ downward pressure on Inherited(X,Z) and Inherited(Y,Z)

## Upward/downward pressure

Inherited(X,Z) & Inherited(Y,Z) & (X != Y)  
-> Similar(X,Y)

- ▶ If satisfied: potentially relevant for explaining
  - ▶ why Similar(X,Y) receives a higher value than expected
  - ▶ why Inherited(X,Z) or Inherited(Y,Z) receives a lower value than expected

## Upward/downward pressure

$\text{Borrowed}(X,Y) + \text{Borrowed}(Y,X) \leq 1$  .

- ▶ If unsatisfied:
  - ▶ downward pressure on  $\text{Borrowed}(X,Y)$  and  $\text{Borrowed}(Y,X)$

# Rule activity

← → **Einh(PEng for TAKE, PWGer for TAKE)** 8 %

The Old English form for TAKE is probably not inherited from the Proto-West-Germanic form for TAKE.

Why is the value not higher?

- **EinhOrEloaOrEunk:**  $1.0 * \text{Eunk}(\text{PEng for TAKE}) [0.00] + 1.0 * \text{Eloa}(\text{PEng for TAKE, PNGer for TAKE}) [0.92] + 1.0 * \text{Eloa}(\text{PEng for TAKE, <ctrlArgForGrounding>}) [0.00] + 1.0 * \text{Einh}(\text{PEng for TAKE, PWGer for TAKE}) [0.08] = 1.0$ .
- **EinhToFhom:**  $\sim(\text{Fhom}(\text{PWGer for TAKE, German (nehmen) /ne:man/}) [0.91]) | \sim(\text{Einh}(\text{PEng for TAKE, PWGer for TAKE}) [0.08]) | \text{Fhom}(\text{PEng for TAKE, German (nehmen) /ne:man/}) [0.00])$ .
- **EinhToFhom:**  $\sim(\text{Fhom}(\text{PWGer for TAKE, English (take) /tesk/}) [0.09]) | \sim(\text{Einh}(\text{PEng for TAKE, PWGer for TAKE}) [0.08]) | \text{Fhom}(\text{PEng for TAKE, English (take) /tesk/}) [1.00])$ .
- **EinhToFsim (x2.0):**  $\sim(\text{Einh}(\text{PEng for TAKE, PWGer for TAKE}) [0.08]) | \sim(\text{Einh}(\text{Dutch (nemen) /ne:ma/}, \text{PWGer for TAKE}) [0.08])$

Why is the value not lower?

- **EinhOrEloaOrEunk:**  $1.0 * \text{Eunk}(\text{PEng for TAKE}) [0.00] + 1.0 * \text{Eloa}(\text{PEng for TAKE, PNGer for TAKE}) [0.92] + 1.0 * \text{Eloa}(\text{PEng for TAKE, <ctrlArgForGrounding>}) [0.00] + 1.0 * \text{Einh}(\text{PEng for TAKE, PWGer for TAKE}) [0.08] = 1.0$ .
- **FhomToEinh (x0.4):**  $\sim(\text{Fhom}(\text{PWGer for TAKE, English (take) /tesk/}) [0.09]) | \sim(\text{Xinh}(\text{the Old English form for TAKE, the Proto-West-Germanic form for TAKE}) [1.00]) | \sim(\text{Fhom}(\text{PEng for TAKE, English (take) /tesk/}) [1.00]) | \text{Einh}(\text{PEng for TAKE, PWGer for TAKE}) [0.08]$
- **FhomToEinh (x0.4):**  $\sim(\text{Fhom}(\text{PWGer for TAKE, German (nehmen) /ne:man/}) [0.91]) | \sim(\text{Xinh}(\text{the Old English form for TAKE, the Proto-West-Germanic form for TAKE}) [1.00]) | \sim(\text{Fhom}(\text{PEng for TAKE, German (nehmen) /ne:man/}) [0.00]) | \text{Einh}(\text{PEng for TAKE, PWGer for TAKE}) [0.08]$

← →

Einh

## Rule activity

- ▶ A ground rule is **active** with respect to a ground atom if
  - ▶ it is dissatisfied (changing the atom's value in the direction of the rule pressure would alleviate the dissatisfaction)
  - ▶ it *would* be dissatisfied if the atom's value were changed slightly (against the rule pressure) – counterfactual test
- ▶ Typically, most ground rules will be inactive for any given atom
- ▶ Only active rules actually contribute to the MAP estimate and therefore its explanation
- ▶ If many groundings of some rule turn out to be inactive,
  - ▶ its reasoning pattern is already covered by (a combination of) other rules, which might mean the rule is not needed, or
  - ▶ the rule needs to be adjusted in order to increase its influence on the results

# Atom & rule verbalization

Einh(PEng for TAKE, PWGer for TAKE) 8 %

The Old English form for TAKE is probably not inherited from the Proto-West-Germanic form for TAKE.

Why is the value not higher?

- The last step in a word's history will be an inheritance or a borrowing, unless its origin is out of scope. An alternative explanation is that the Old English form for TAKE is **borrowed** from the Proto-North-Germanic form for TAKE, which is likely.
- When we reconstruct an inheritance and assign some belief to the homologue status of the parent, we must assign at least as much belief to the child's inclusion in the same homologue set. Applying this logic to the homologue set for German (nehmen) /ne:mən/, the Proto-West-Germanic form for TAKE **probably belongs to this set** but the Old English form for TAKE **certainly does not**, so reconstructing an inheritance becomes problematic.
- When we reconstruct an inheritance and assign some belief to the homologue status of the parent, we must assign at least as much belief to the child's inclusion in the same homologue set. Applying this logic to the homologue set for English (take) /teɪk/, we are already certain that the Old English form for TAKE **belongs to this set**. (Therefore, the inheritance relationship

Why is the value not lower?

- The last step in a word's history will be an inheritance or a borrowing, unless its origin is out of scope. An alternative explanation is that the Old English form for TAKE is **borrowed** from the Proto-North-Germanic form for TAKE, which is likely.
- If the forms in a language and its parent are assigned to the same homologue set, this suggests that the form in the child language was inherited. Applying this logic to the homologue set for English (take) /teɪk/, since neither homology judgment is entirely unlikely (the Old English form for TAKE is **certainly** a homologue of English (take) /teɪk/ although the Proto-West-Germanic form for TAKE is **probably not** a homologue of English (take) /teɪk/), we cannot disregard the possibility that the Proto-West-Germanic form for TAKE is inherited from the Old English form for TAKE.
- If the forms in a language and its parent are assigned to the same homologue set, this suggests that the form in the child language was inherited. Applying this logic to the homologue set for German (nehmen) /ne:mən/, since the \*Old English

Einh

# Atom verbalization

Einh(PWSc for TAKE, PNGer for T) 8 %

The Old English form for TAKE is probably not inherited from the Proto-West-Germanic form for TAKE.

Why is the value not higher?

- The last step in a word's history will be an inheritance or a borrowing, unless its origin is out of scope. An alternative explanation is that the Old English form for TAKE is **borrowed** from the Proto-North-Germanic form for TAKE, which is likely.
- When we reconstruct an inheritance and assign some belief to the homologue status of the parent, we must assign at least as much belief to the child's inclusion in the same homologue set. Applying this logic to the homologue set for German (nehmen) /ne:man/, the Proto-West-Germanic form for TAKE **probably belongs to this set** but the Old English form for TAKE **certainly does not**, so reconstructing an inheritance becomes problematic.
- When we reconstruct an inheritance and assign some belief to the homologue status of the parent, we must assign at least as much belief to the child's inclusion in the same homologue set. Applying this logic to the homologue set for English (take) /te:k/, we are already certain that the Old English form for TAKE **belongs to this set**. (Therefore, the inheritance relationship

Why is the value not lower?

- The last step in a word's history will be an inheritance or a borrowing, unless its origin is out of scope. An alternative explanation is that the Old English form for TAKE is **borrowed** from the Proto-North-Germanic form for TAKE, which is likely.
- If the forms in a language and its parent are assigned to the same homologue set, this suggests that the form in the child language was **inherited**. Applying this logic to the homologue set for English (take) /te:k/, since neither homology judgment is entirely unlikely (the Old English form for TAKE is **certainly** a homologue of English (take) /te:k/ although the Proto-West-Germanic form for TAKE is **probably not** a homologue of English (take) /te:k/), we cannot disregard the possibility that the Proto-West-Germanic form for TAKE is inherited from the Old English form for TAKE.
- If the forms in a language and its parent are assigned to the same homologue set, this suggests that the form in the child language was **inherited**. Applying this logic to the homologue set for German (nehmen) /ne:man/, since the \*Old English

Einh



# Rule verbalization

- Einh(PWSc for TAKE, PNGer for TA
- Einh(Icelandic (taka) /ˈtɑkɑ/, PWSc
- Einh(English (take) /teɪk/, PEng for
- Einh(Dutch (nemen) /neːmə/, PWG
- Einh(German (nehmen) /neːmən/, P
- Einh(Danish (tage) /ˈtɑː/, PNGer fo
- Einh(Swedish (ta) /tɑː/, PNGer for
- Einh(PWGer for TAKE, PGer for TA
- Einh(PNGer for TAKE, PGer for TAK
- Einh(Norwegian (ta) /tɑː/, PWSc fa
- Einh(PEng for TAKE, PWGer for TA

Ein

## Einh(PEng for TAKE, PWGer for TAKE) 8 %

The Old English form for TAKE is probably not inherited from the Proto-West-Germanic form for TAKE.

Why is the value not higher?

- The last step in a word's history will be an inheritance or a borrowing, unless its origin is out of scope. An alternative explanation is that the Old English form for TAKE is **borrowed** from the Proto-North-Germanic form for TAKE, which is **likely**.
- When we reconstruct an inheritance and assign some belief to the homologue status of the parent, we must assign at least as much belief to the child's inclusion in the same homologue set. Applying this logic to the homologue set for German (nehmen) /neːmən/, the Proto-West-Germanic form for TAKE **probably belongs to this set** but the Old English form for TAKE **certainly does not**, so reconstructing an inheritance becomes problematic.
- When we reconstruct an inheritance and assign some belief to the homologue status of the parent, we must assign at least as much belief to the child's inclusion in the same homologue set. Applying this logic to the homologue set for English (take) /teɪk/, we are already certain that the Old English form for TAKE **belongs to this set**. (Therefore, the inheritance relationship

Why is the value not lower?

- The last step in a word's history will be an inheritance or a borrowing, unless its origin is out of scope. An alternative explanation is that the Old English form for TAKE is **borrowed** from the Proto-North-Germanic form for TAKE, which is **likely**.
- If the forms in a language and its parent are assigned to the same homologue set, this suggests that the form in the child language was inherited. Applying this logic to the homologue set for English (take) /teɪk/, since neither homology judgment is entirely unlikely (the Old English form for TAKE is **certainly** a homologue of English (take) /teɪk/ although the Proto-West-Germanic form for TAKE is **probably not** a homologue of English (take) /teɪk/), we cannot disregard the possibility that the Proto-West-Germanic form for TAKE is inherited from the Old English form for TAKE.
- If the forms in a language and its parent are assigned to the same homologue set, this suggests that the form in the child language was inherited. Applying this logic to the homologue set for German (nehmen) /neːmən/, since the \*Old English

# Rule verbalization

- ▶ Expressing the mechanics of each ground rule in terms of domain-specific natural language
- ▶ Templates for various scenarios
  - ▶ Which position within the rule does the inspected atom have?
  - ▶ Upwards/downwards pressure

## Rule verbalization

```
Inherited(X,Z) & Inherited(Y,Z) & (X != Y)
  -> Similar(X,Y)
```

```
Inherited("German 'nehmen'", "P-West-Gmc '...') 0.9 }
& Inherited("Old Eng '...'", "P-West-Gmc '...') 0.1 } 0.0
& ("German 'nehmen'" != "Old Eng '...') 1.0 }
-> Similar("German 'nehmen'", "Old Eng '...') 0.4
```

## Rule verbalization

Expressing the mechanics of each ground rule in terms of domain-specific natural language

```
Inherited(X,Z) & Inherited(Y,Z) & (X != Y)  
  -> Similar(X,Y)
```

- ▶ Introductory sentence summarizing the reasoning pattern expressed by the rule

*"If two words are inherited from the same source word, they should be phonetically similar."*

## Rule verbalization

```
Inherited("German 'nehmen'", "P-West-Gmc '...'") 0.9  
& Inherited("Old Eng '...'", "P-West-Gmc '...'") 0.1  
& ("German 'nehmen'" != "Old Eng '...'") 1.0  
-> Similar("German 'nehmen'", "Old Eng '...'") 0.4
```

- ▶ Introductory sentence

*"If two words are inherited from the same source word, they should be phonetically similar."*

- ▶ Then fill in specifics of grounding & inference

*"However, since it is very likely that German 'nehmen' was inherited from Proto-West-Germanic, but it is very unlikely that the Old English word was as well, the similarity score is actually not constrained."*

## Rule verbalization

```
Inherited("German 'nehmen'", "P-West-Gmc '...') 0.9  
& Inherited("Old Eng '...'", "P-West-Gmc '...') 0.1  
& ("German 'nehmen'" != "Old Eng '...'") 1.0  
-> Similar("German 'nehmen'", "Old Eng '...'") 0.4
```

► Introductory sentence

*"If two words are inherited from the same source word, they should be phonetically similar."*

► Then fill in specifics of grounding & inference

*"Since it is very likely that German 'nehmen' was inherited from Proto-West-Germanic, but the German and Old English words are moderately dissimilar, it should be at least somewhat unlikely that the Old English word comes from the same source."*

Code can be used for any PSL model:

- ▶ [github.com/jdellert/psl-infrastructure](https://github.com/jdellert/psl-infrastructure)
  - ▶ Java API for defining/running PSL inferences
  - ▶ Rule-atom graph analysis tools (rule pressure, rule activity, verbalizations)
  - ▶ (More PSL tools released and to come: inference parallelization, model evaluation, ...)
- ▶ [github.com/verenablaschke/psl-ragviewer](https://github.com/verenablaschke/psl-ragviewer)
  - ▶ GUI

Thank you!  
Questions?

[github.com/jdellert/psl-infrastructure](https://github.com/jdellert/psl-infrastructure)  
[github.com/verenablaschke/psl-ragviewer](https://github.com/verenablaschke/psl-ragviewer)

[verena.blaschke@cis.lmu.de](mailto:verena.blaschke@cis.lmu.de)  
[johannes.dellert@uni-tuebingen.de](mailto:johannes.dellert@uni-tuebingen.de)



**European Research Council**  
Established by the European Commission

This work is part of a project that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (CrossLingference, grant agreement no. 834050).



## Appendix

## Etymological inference: predicates

Closed predicates:

InheritancePoss(X,Y) lang of Y is parent of lang of X

BorrowingPoss(X,Y) lang of Y influenced lang of X

Open predicates used for reasoning:

Homologue(X,H) belief that word X belongs to homologue set H

Similar(X,Y) phonetic similarity of the words X and Y  
(already known for pairs of modern words)

Open predicates for encoding etymologies (= the inference result):

Inherited(X,Y) belief that X was inherited Y

Borrowed(X,Y) belief that X was borrowed from Y

UnknownOrigin(X) belief that etymology of X is outside scope

## Etymological Inference: constraints and priors

Homologue set assignment for each form is a distribution:

$$\text{Homologue}(X, +H) = 1.$$

Form similarity is symmetric and (roughly) transitive:

$$\text{Similar}(X, Y) = \text{Similar}(Y, X) .$$

$$(X \neq Y) \ \& \ (X \neq Z) \ \& \ (Y \neq Z) \ \& \\ \text{Similar}(X, Y) \ \& \ \text{Similar}(Y, Z) \ \rightarrow \ \text{Similar}(X, Z) .$$

Weak negative prior on borrowing (= prefer inheritance by default):

$$0.5: \ !\text{Borrowed}(X, Y)$$

Strong negative prior on unknown (= try hard to find etymology):

$$2.5: \ !\text{UnknownOrigin}(X)$$

## Etymological Inference: rules

Belief assigned to etymologies for each form is a distribution:

$$\text{Inherited}(X,+Y) + \text{Borrowed}(X,+Z) + \text{UnknownOrigin}(X) = 1 .$$

Borrowing is monodirectional:

$$\text{Borrowed}(X,Y) + \text{Borrowed}(Y,X) \leq 1 .$$

Etymological links are only possible between homologues:

$$\text{Inherited}(X,Y) \ \& \ \text{Homologue}(Y,H) \ \rightarrow \ \text{Homologue}(X,H) .$$

$$\text{Borrowed}(X,Y) \ \& \ \text{Homologue}(Y,H) \ \rightarrow \ \text{Homologue}(X,H) .$$

## Etymological inference: rules

Strong preference for two words which are inherited from the same word to be phonetically similar:

2.0:  $\text{Inherited}(X,Z) \ \& \ \text{Inherited}(Y,Z) \ \& \ (X \neq Y)$   
 $\rightarrow \text{Similar}(X,Y)$

All pairs of inherited words are at least as similar as their parents (preventing forms from becoming more similar with time):

1.0:  $\text{Similar}(X,Y) \ \& \ \text{Inherited}(X,W) \ \& \ \text{Inherited}(Y,Z)$   
 $\ \& \ (W \neq Z) \rightarrow \text{Similar}(W,Z)$

A loan should be more similar to its source than to any other word:

1.0:  $\text{Borrowed}(X,Y) \ \& \ \text{Similar}(X,Z) \ \& \ (Y \neq Z) \ \& \ (X \neq Z)$   
 $\rightarrow \text{Similar}(X,Y)$

## Etymological inference: rules

Evidence of homologue set presence is propagated along parent-child links, with the child-to-parent direction being dominant:

0.6:  $\text{Homologue}(X,H) \ \& \ \text{InheritancePoss}(X,Z)$   
     $\rightarrow \text{Homologue}(Z,H)$

0.2:  $\text{Homologue}(Z,H) \ \& \ \text{InheritancePoss}(X,Z)$   
     $\rightarrow \text{Homologue}(X,H)$

## Etymological inference: rules

If parent and child are homologues, that suggests inheritance:

0.4: Homologue(X,H) & Homologue(Y,H)  
& InheritancePoss(X,Y) -> Inherited(X,Y)

By contrast, if there is any reason to doubt the presence of a homologue set in the parent, an available loanword etymology becomes much more likely:

1.0: Homologue(X,H) & !Homologue(Y,H) &  
InheritancePoss(X,Y) & BorrowingPoss(X,Z)  
-> Borrowed(X,Z)