

A Computational Approach to Quantifying Grammaticalization of English Deverbal Prepositions



Ryo Nagata (Konan University)

Yoshifumi Kawasaki (University of Tokyo)

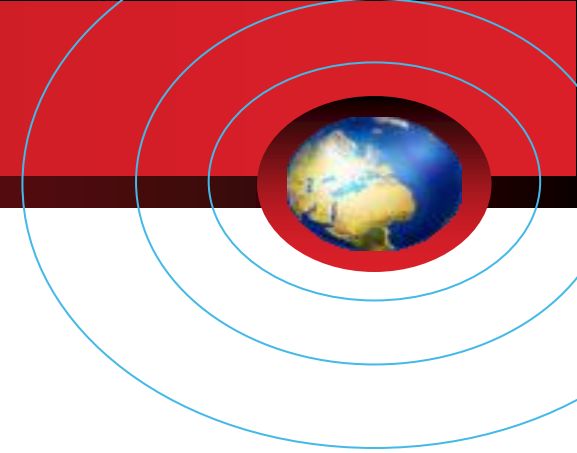
Naoki Otani (Tokyo University of Foreign Studies)

Hiroya Takamura (AIST)

Background

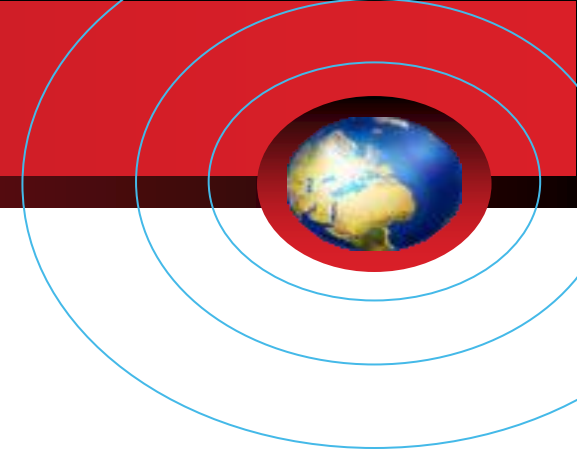


Grammaticalization?



- **Diachronic change: content → function words**
 - E.g., *can*
 - ~ 14th century: to know (verb) → now: to be able to (aux.)

What are deverbal prepositions?



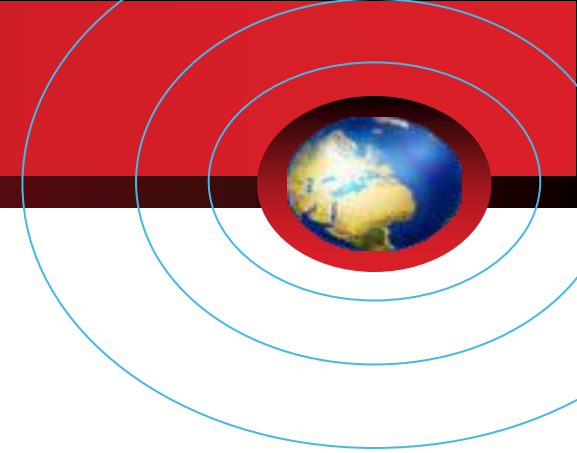
■ words

- derived from a verb
- behaving as a preposition

e.g., *during*, *considering*, *according to*

Hereafter, referred to as D-prep

Grammaticalization of D-preps



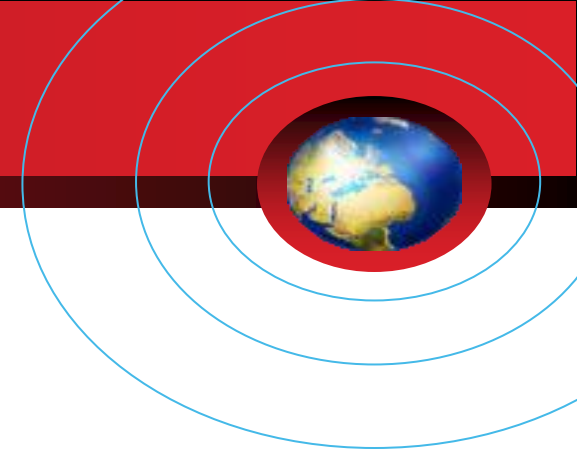
- **Basically**



- verb → preposition

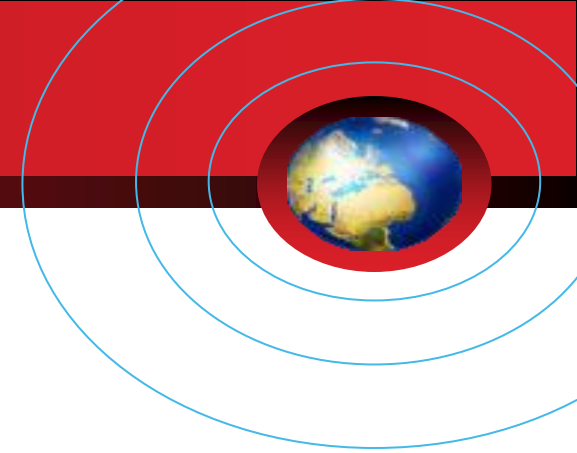
- **Findings in linguistics**

1. Not yet fully grammaticized; between prepositions and verbs
2. Degree of grammaticization is gradual; some are more like prepositions
e.g, high: *past, during, except*
low: *succeeding, considering, allowing for*
3. can be measured based on prepositionality and verbality

Related research questions



- Which d-preps are more/less like preposition?
- Is the change unidirectional?
 - Unidirectional hypothesis
 - content → function word 
 - function → content word 
- When did verbs become d-preps?



■ Mostly, manual analyses

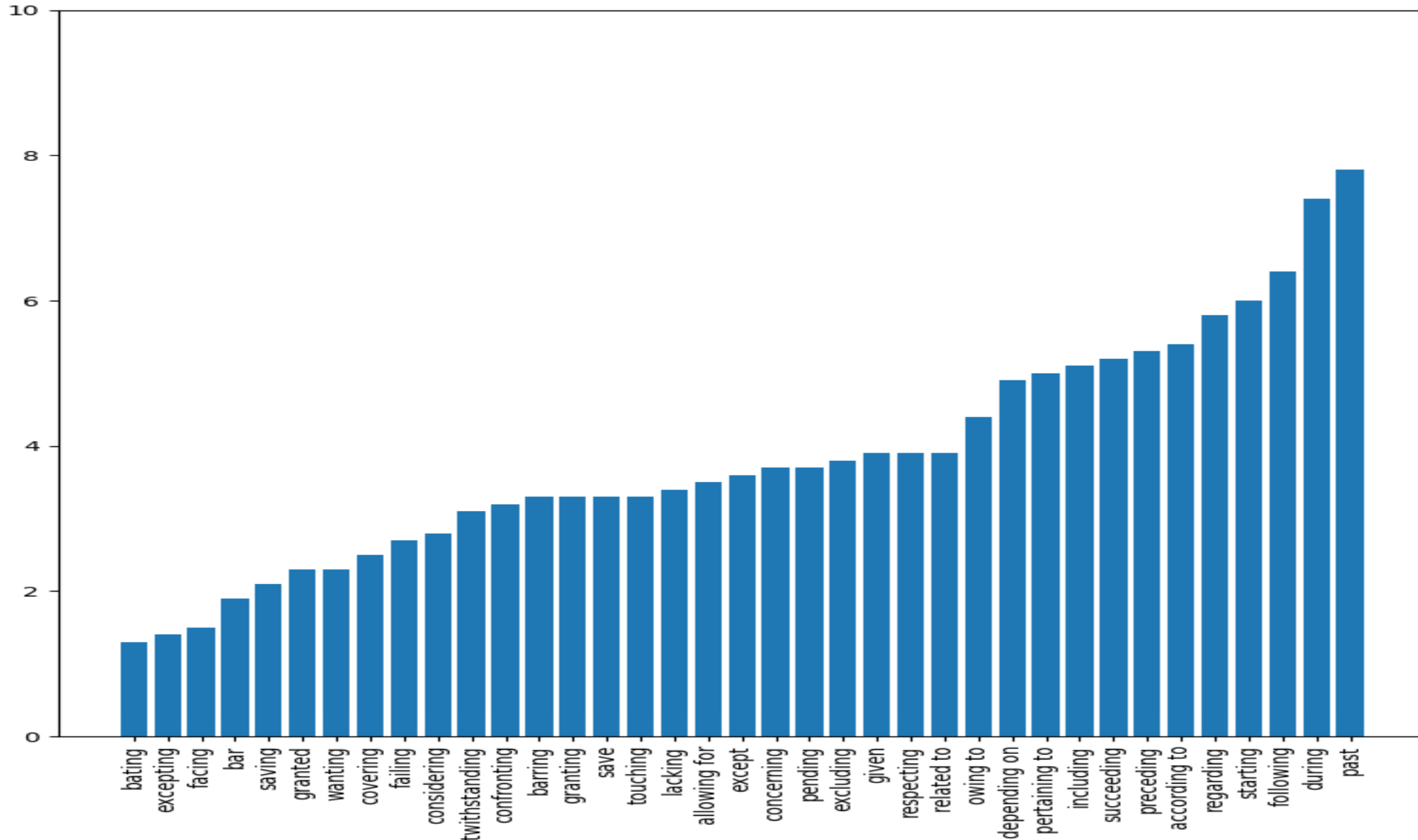
- corpus analysis

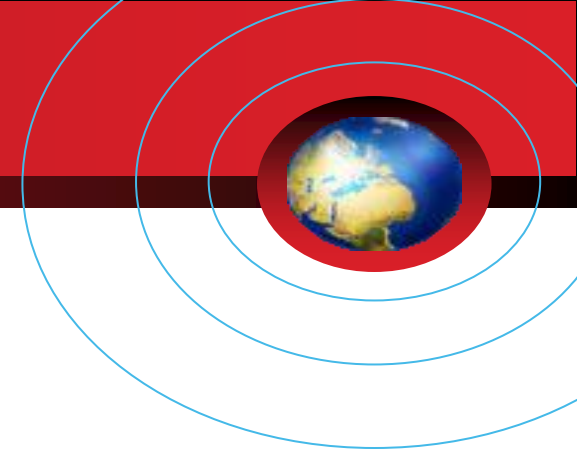
 - e.g., co-reference (same subject in main and participle clauses)

 - Considering its narrow beam, the boat is remarkably sea-worthy.*

- test based (ask native-speakers)

Prepositoinarity of d-preps (Hayashi 2015)





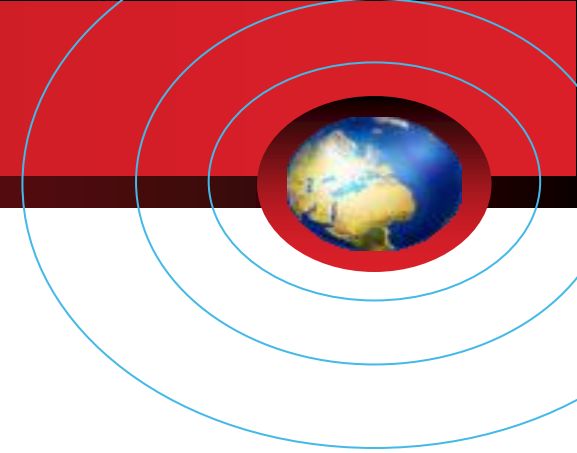
■ Labor-intensive

- Limited target d-preps
- Limited number of instances examined
- Limited number of examiners (few persons)

This work
word vector-based



Word vector-based approaches



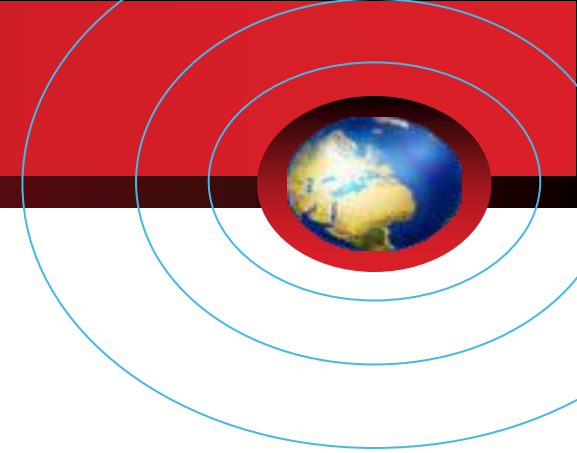
- **Uncontextualized word vectors (e.g., word2vec)**
 - direct comparison to prepositions ○
 - direct comparison to prepositions&verbs ◎

- **Contextualized word vectors (e.g., BERT)**
 - norm-based △

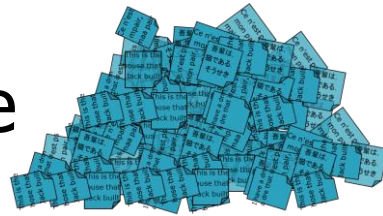
Uncontextualized word vector-based approaches



Uncontextualized word vector-based method



d-prep
vectorize



d

score:

$$\cos(d, \bar{p})$$

sim. to preps

preps
vectorize,
average

\bar{p}

-

verbs
vectorize,
average

\bar{v}

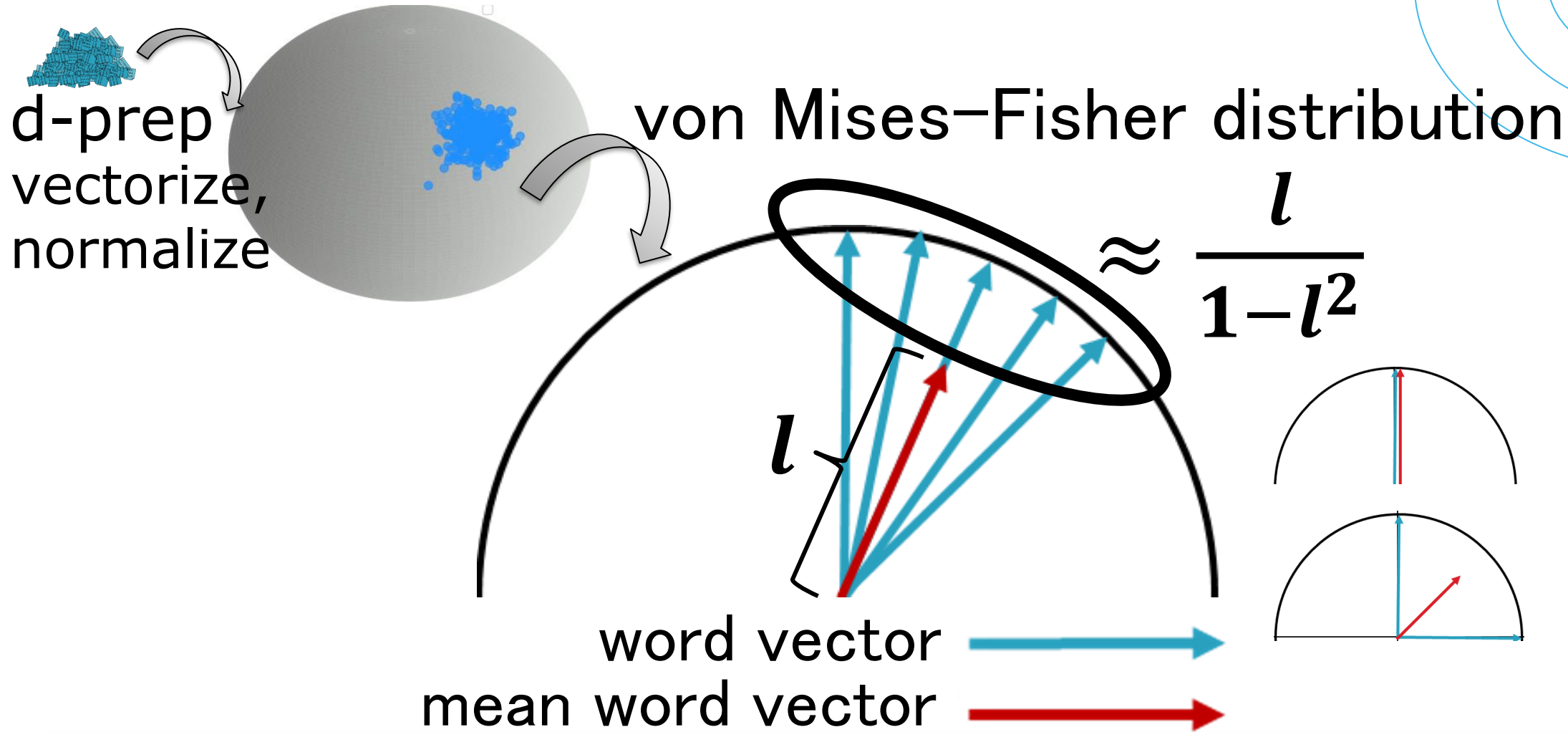
$$\cos(d, \bar{v})$$

disim to verbs

Contextualized word vector-based approaches



Based on concentration of contextualized word vectors

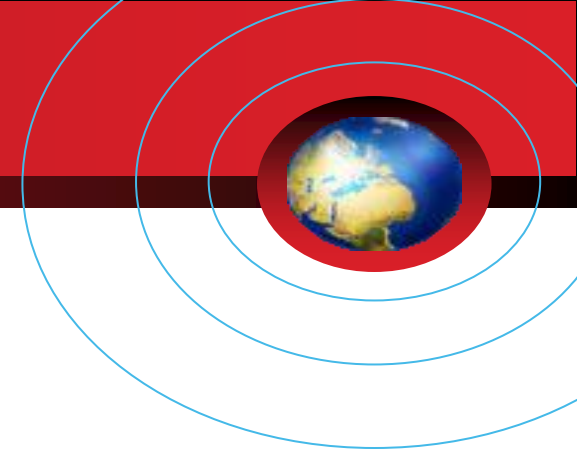


larger variety of contexts (as in function words), larger $\frac{1-l^2}{l}$

Evaluation



Evaluation conditions



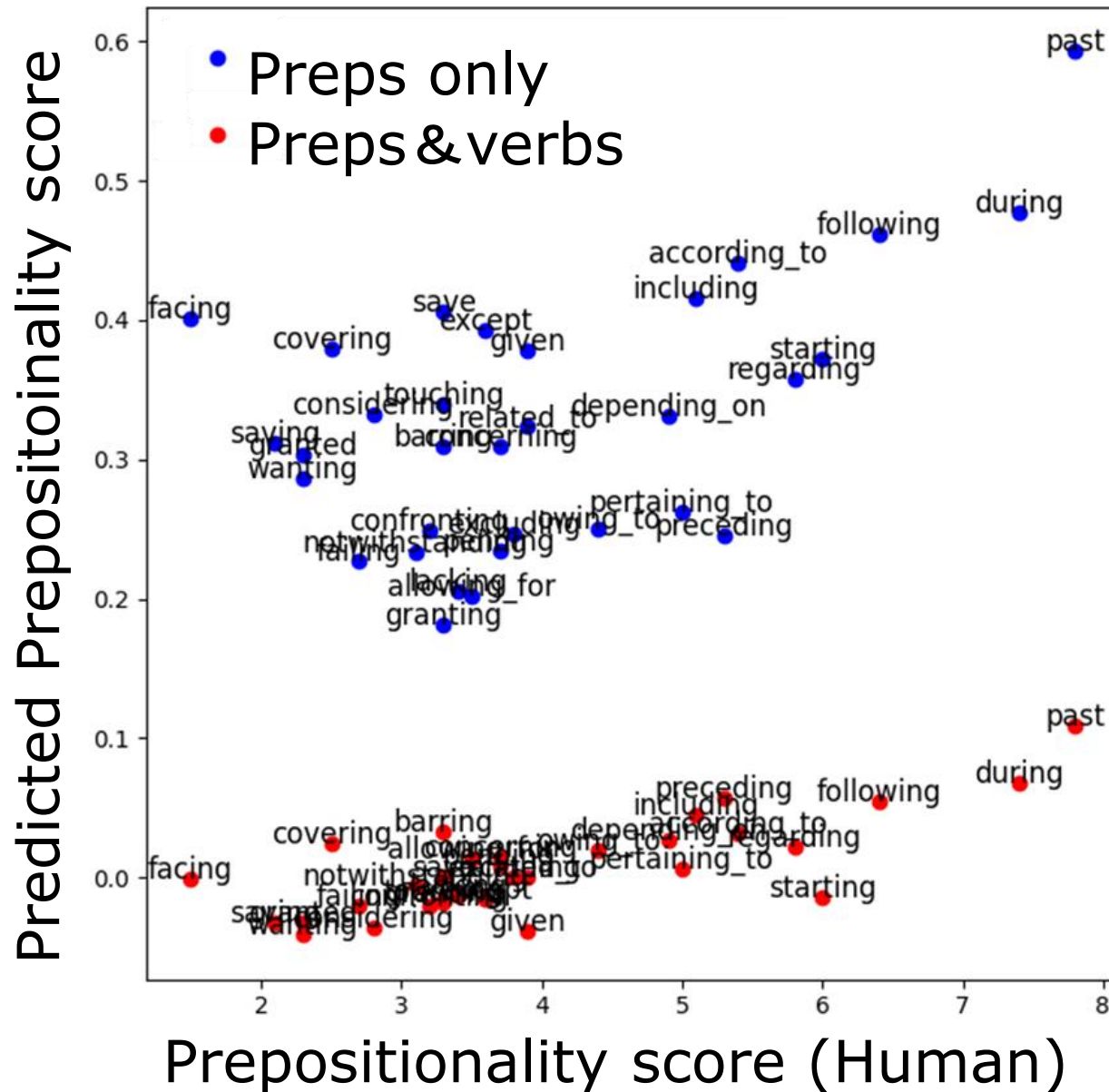
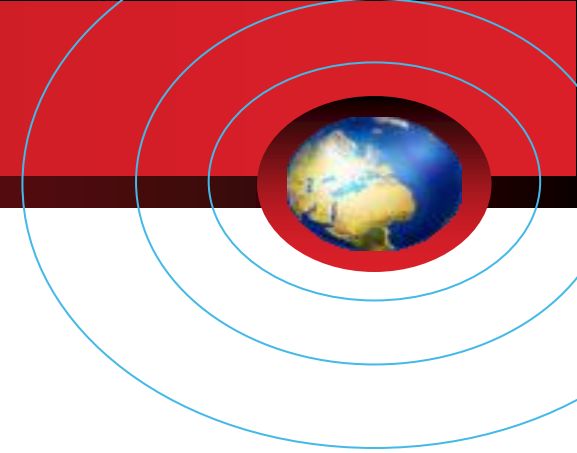
- **Corpus : COHA (2000s)**
- **Target D-preps : 32 (from Hayashi 2015)**
- **word2vec : dim200, w size: 5, skip-gram**
- **Prepositions : major 52**
- **Verbs : 200 most frequent ones**
- **Metrics : Spearman's rank co-relation coefficient
(w.r.t Hayashi 2015)**

Spearman's rank correlation to human judgement



Method	Coefficient (p-value)
Mean prep & mean verb	0.69 (0.1×10^{-6})
Mean prep only	0.37 (0.03)
Mean verb only	0.09 (0.64)
Contextualized vector-based	0.14 (0.43)

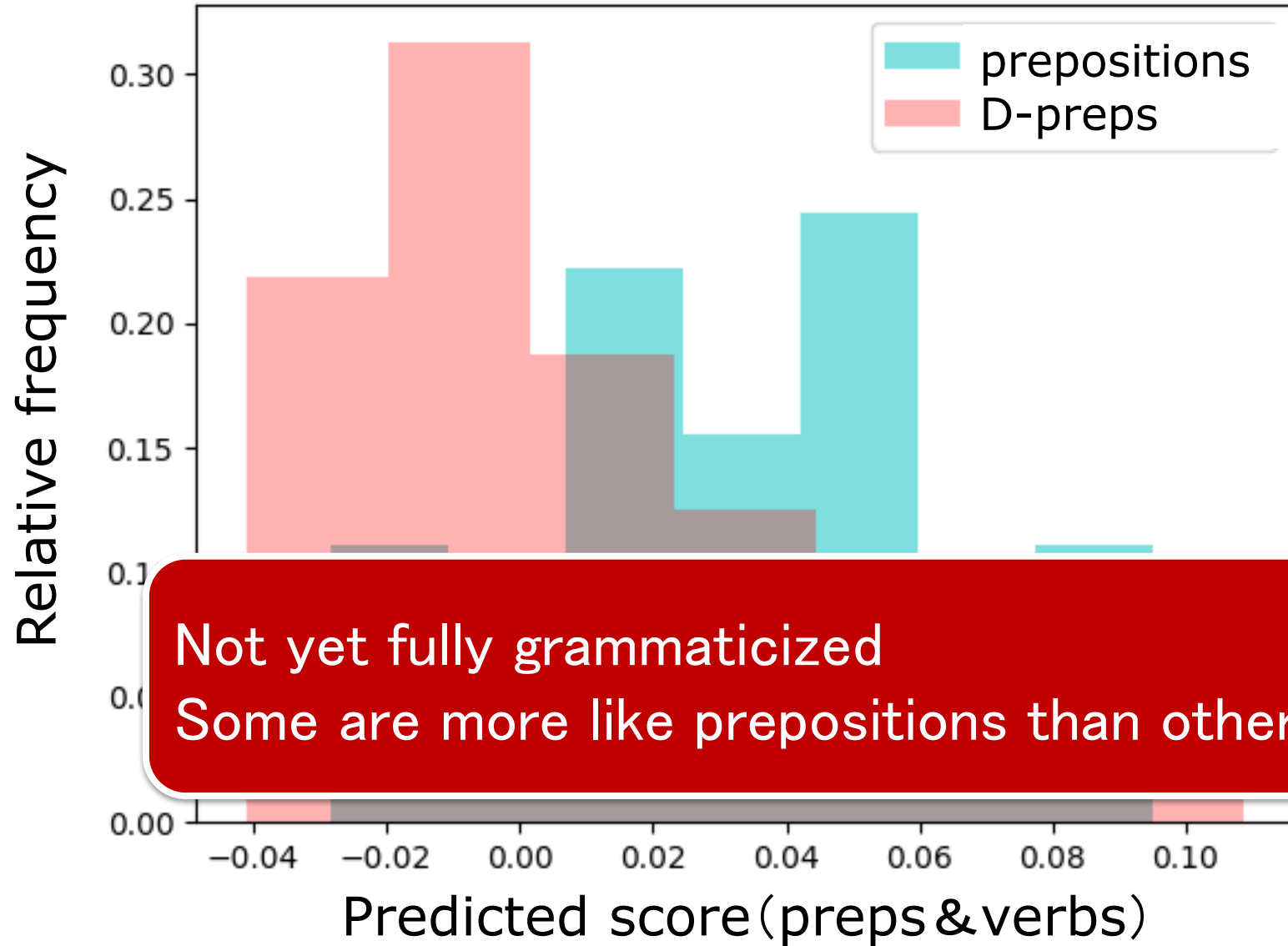
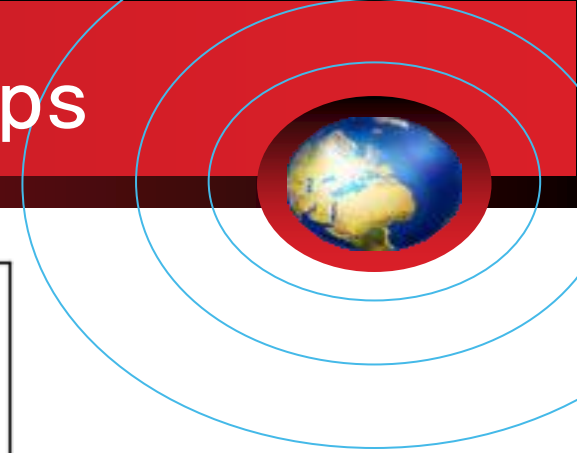
With/without information about verbs



Discussion

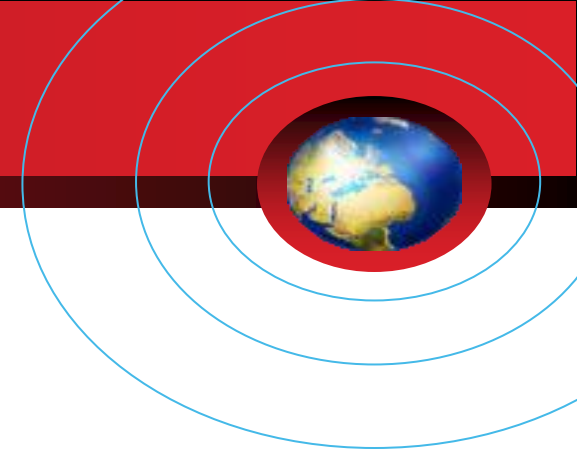


Degree of grammaticization of preps and D-preps

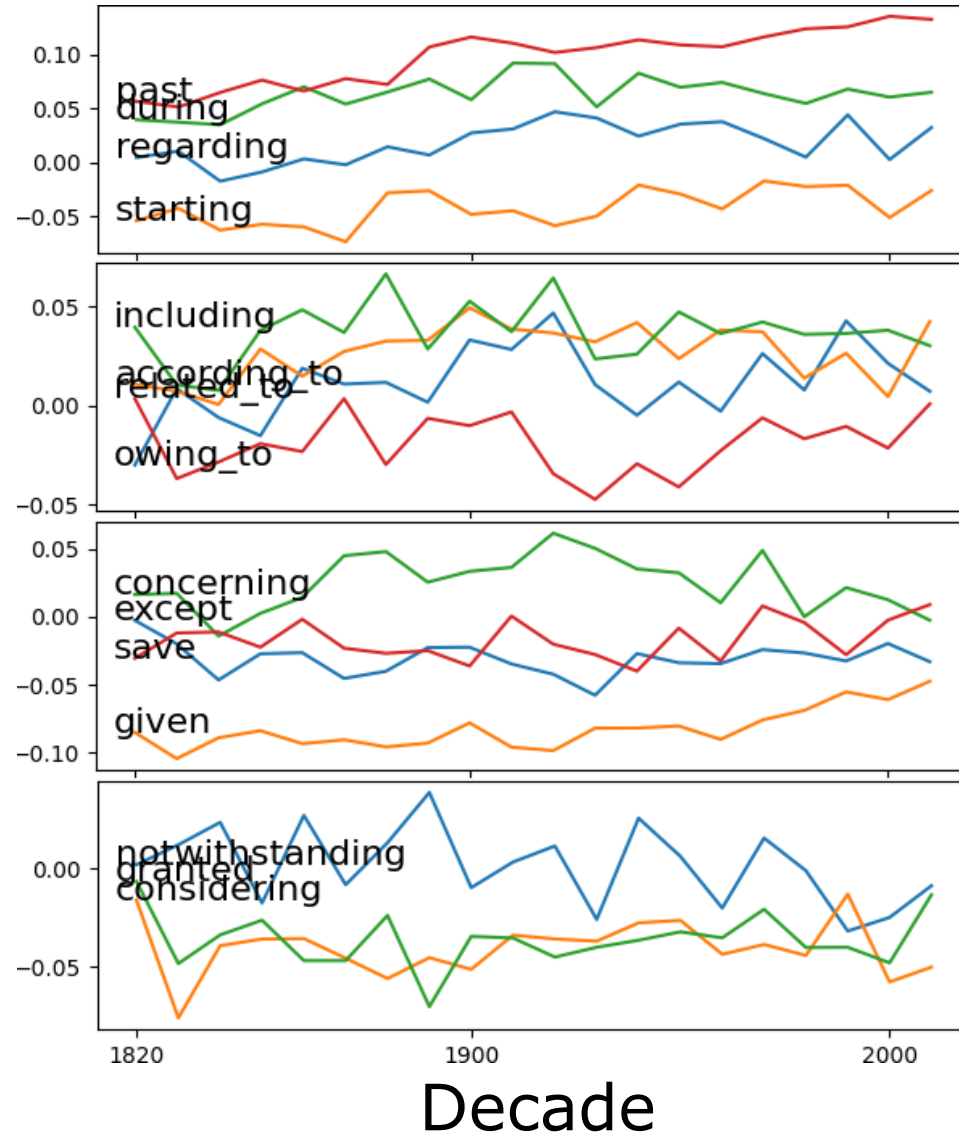


Not yet fully grammaticized
Some are more like prepositions than others

Diachronic change



Predicted score (preps&verbs)





- **Quantifying Grammaticalization of English Deverbal Prepositions**
 - Uncontextualized word vector-based ○
 - Contextualized word vector-based △ (but more general)
- **Support three previous, linguistic findings**
 1. Not yet fully grammaticized
 2. Degree of grammaticization is gradual
 3. its degree is measured based on prepositionality and verballity
- **Unidirectional hypothesis**
 - Requires more investigations