GASC: Genre-Aware Semantic Change for Ancient Greek

Valerio Perrone[1], Marco Palma[2], **Simon Hengchen**[3], Alessandro Vatri[4,6], Jim Q. Smith[2,6], and **Barbara McGillivray**[5,6]

Amazon[1], University of Warwick[2], University of Helsinki[3], University of Oxford[4], University of Cambridge[5], The Alan Turing Institute[6]

vperrone@amazon.com, simon.hengchen@helsinki.fi, bmcgillivray@turing.ac.uk

amazon

The Alan Turing Institute

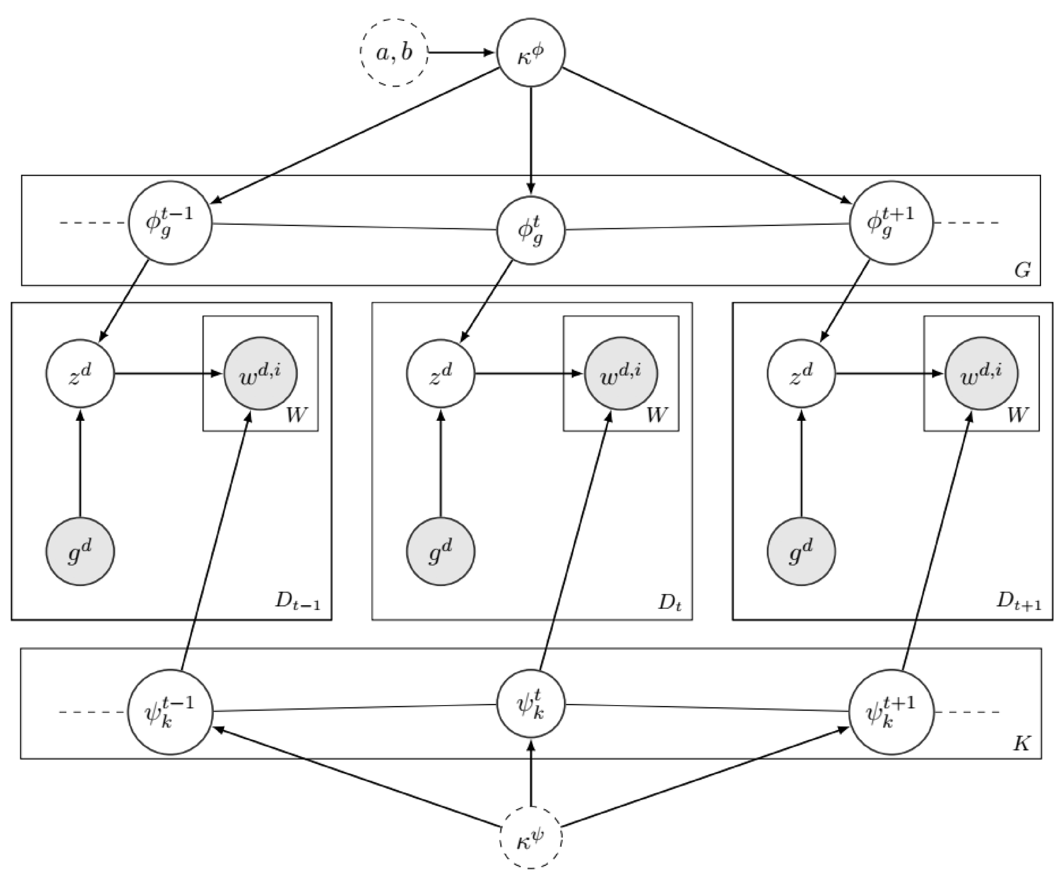
Semantic change for Ancient Greek

Bayesian models have emerged as a powerful tool to provide explicit and interpretable representations of semantic change phenomena. However, for ancient languages, a lack of data and a long diachronic span make it harder to draw a clear distinction between polysemy and semantic change, and current systems perform poorly in these settings.

We introduce GASC (**G**enre-**A**ware **S**emantic **C**hange), a novel dynamic Bayesian mixture model for semantic change. In this model, the evolution of word senses over time is based not only on distributional information of lexical nature, but also on additional features, specifically genre. This allows GASC to decouple sense probabilities and genre prevalence, which is critical with genre-unbalanced data such as ancient languages corpora.

GASC

We start with a lemmatized corpus pre-processed into text snippets, each containing an instance of the word under study ("target word"). Each snippet is a fixed-sized window W of 5 words to the left and right of the target word. The goal is to detect the sense associated to the target word in the given context, and describe the evolution of sense proportions. Suppose the target word is used with K different senses. A sense at time t is a distribution ψ_k^t over words from the dictionary. These distributions are used to generate text snippets by drawing each of their words from the dictionary based on a Multinomial distribution.



Algorithm 1: GASC generative model

Draw $K^\phi \sim Gamma(a, b)$;
for *time* $t = 1, \ldots, T$ **do**
  for *genre* $g = 1, \ldots, G$ **do**
    Draw sense distribution
    $\phi_g^t \mid \phi_g^{-t}, K^\phi \sim N(\frac{1}{2}(\phi_g^{t-1} + \phi_g^{t+1}), K^\phi)$
  end
  for *sense* $k = 1, \ldots, K$ **do**
    Draw word distribution
    $\psi_k^t \mid \psi^{-t}, K^\psi \sim N(\frac{1}{2}(\psi_k^{t-1} + \psi_k^{t+1}), K^\psi)$
  end
  for *document* $d = 1, \ldots, D_t$ **do**
    Let g^d be the observed genre;
    Draw sense $z^d \mid g^d \sim \text{Mult}(\text{softmax}(\phi_{g^d}^t))$;
    for *context position* $i = 1, \ldots, W$ **do**
      Draw word $w^{d,i} \sim \text{Mult}(\text{softmax}(\psi^{t,z^d}))$;
    end
  end
end

Two sources of variation are captured:

- *Time-variation*: word and sense distributions evolve with Gaussian changes. The coupling between sense probabilities over time is controlled by K^ϕ. We place a Gamma prior over K^ϕ with hyperparameters a and b. We fix K^ψ, the word probability precision parameter.
- *Genre-variation*: each of G genres determines a genre-specific distribution over senses $\phi_{g^d}^t$ at time t where g^d is the observed genre for document d. This accounts for genre-specific word usage patterns.

Posterior inference

For posterior inference, we extend the blocked Gibbs sampler proposed in SCAN [1]:

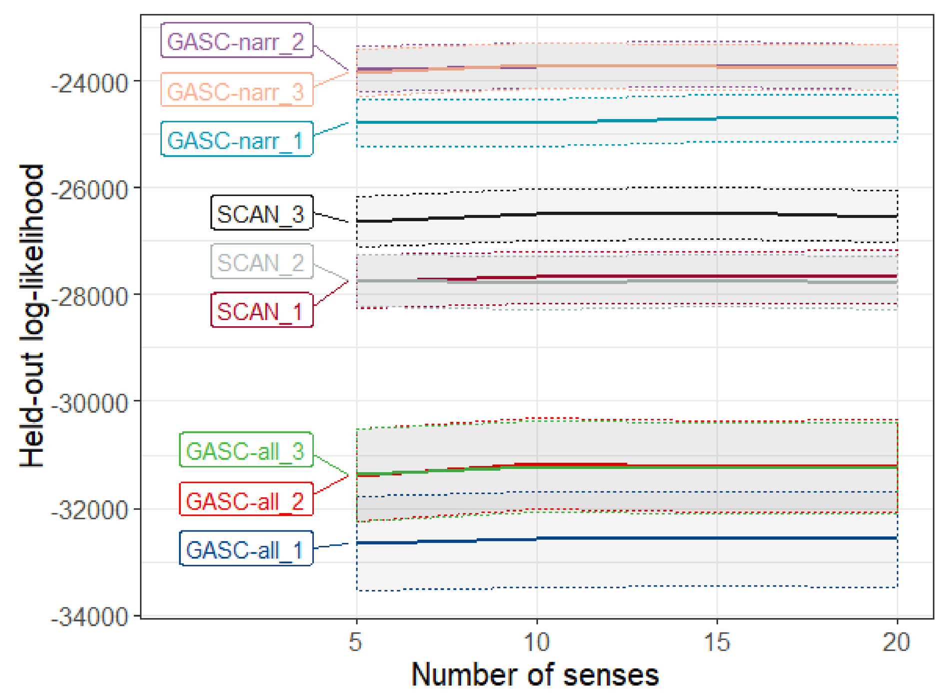
- Sample the snippet-sense assignments from their full conditional, the sense precision parameters from their conjugate Gamma priors, and the sense and word distributions with the auxiliary-variable approach from [2].
- For the distribution over genres we proceed as follows. First, sample the distribution over senses ϕ_g^t for each genre $g = 1, \ldots, G$ following [2]. Then, sample the sense assignment conditioned on the observed genre from its full conditional:
$p(z^d \mid g^d, \mathbf{w}, t, \phi, \psi) \propto p(z^d \mid g^d, t) p(\mathbf{w} \mid t, z^d) = \phi_g^t \prod_{w \in \mathbf{w}} \psi_w^{t,z^d}$.

Ancient Greek dataset

Diorisis Annotated Ancient Greek Corpus [3]: 10,206,421 lemmatized and PoS-tagged words, 8th cent. BC - 5th cent. AD. Various literary and technical genres.

Predictive performance

We evaluated predictive log-likelihood of held-out data on a 50-word dataset for SCAN (not using any genre information), GASC-all (GASC with all the $G = 10$ available genres) and GASC-narr (GASC with 2 genres, Narrative vs. non Narrative).



Predictive log-likelihood over 50 leave-one-out folds for different K and 3 main settings: 1) $a = 7, b = 3, K^\psi = 10$ as [1], 2) $a = 7, b = 3, K^\psi = 100$, to reduce time-variation within senses, 3) $a = b = 1, K^\psi = 100$, to allow probabilities to vary widely across centuries.

- GASC-narr consistently outperforms SCAN.
- SCAN has higher held-out log-likelihood than GASC-all.

Exploiting some information on the genre yields better predictions, while using all genres attested in the corpus is not effective.

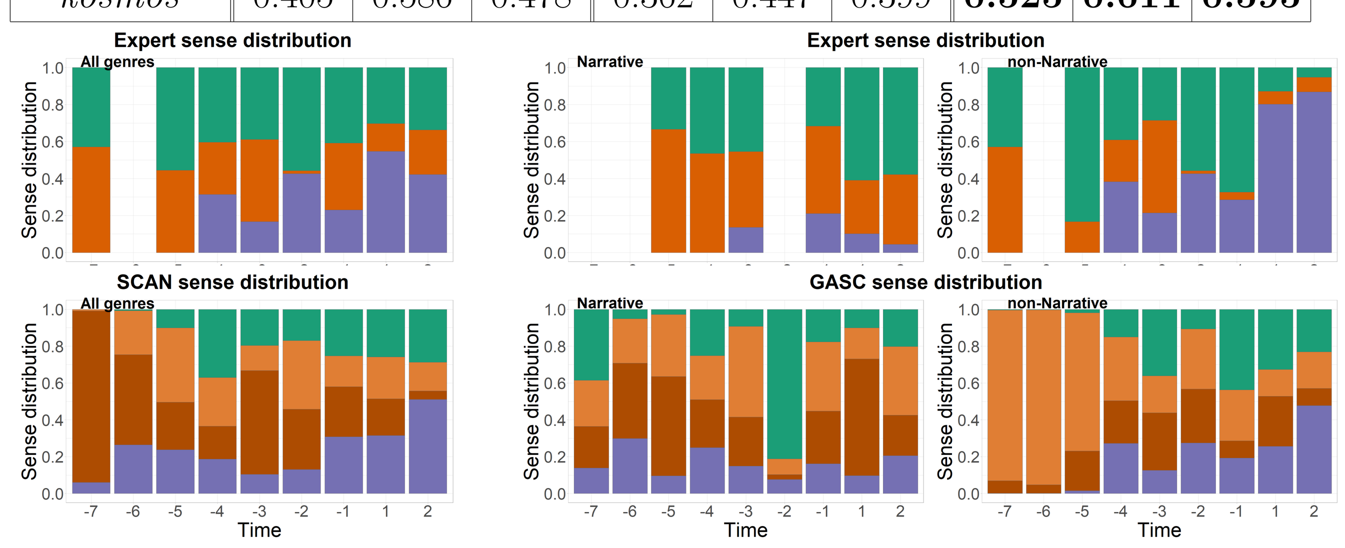
Ground truth evaluation

Semantically annotated 3 words: *mus* 'mouse'/'muscle'/'mussel', *harmonia*, 'fastening'/'agreement'/'musical scale, melody', *kosmos* 'order'/'world'/'decoration').
Confidence score:

$$conf(k, s) = \sum_i P(w_i|k) * m(w_i, s),$$

k model sense, s annotated sense, $P(w_i|k)$ normalised probabilities with which words w_i are associated to k, $m(w_i, s)$ expert scores of words w_i for s. For every s we select k for which conf(k, s) is higher than the random baseline *and* the sum of the 2nd and 3rd confidence scores. Otherwise, s is labelled NA. For every (s, k), a word is considered correctly assigned to sense k if it also appears within a 5-word window of the target word in the expert annotation for s.
Precision (P) = # words correctly assigned to k weighted by their respective normalised model-estimated probabilities / number of words assigned to k by the model.
Recall (R) = # words correctly assigned to k weighted by their probabilities / # words assigned to sense s by the experts weighted by their expert scores

Word/Model	SCAN			GASC-independent			GASC		
	P	R	F1	P	R	F1	P	R	F1
mus	**0.430**	**0.477**	**0.452**	0.420	0.442	0.431	0.224	0.298	0.253
harmonia	0.527	0.708	0.603	**0.582**	**0.729**	**0.646**	0.497	0.481	0.484
kosmos	0.405	0.586	0.478	0.362	0.447	0.399	**0.525**	**0.611**	**0.595**



References

[1] L. Frermann and M. Lapata. A Bayesian model of diachronic meaning change. *Transactions of the Association for Computational Linguistics*, 4:31–45, 2016.

[2] D. Mimno, H. Wallach, and A. McCallum. Gibbs sampling for logistic normal topic models with graph-based priors. In *NIPS workshop*, 2008.

[3] A. Vatri and B. McGillivray. The Diorisis Ancient Greek Corpus. *Research Data Journal for the Humanities and Social Sciences*, 2018.