Grammatical Factors in Morphological Processing

Evidence from Allomorphy

Daniil Bondarenko, Onur Özsoy, Itamar Kastner

Humboldt-Universität zu Berlin
30. November 2019

Outline

Background: Allomorphy

Morphemes can have variants depending on distinct environments:

Phonologically conditioned
(rule-based)

cats~*dogs*~*buses*
(/ts/~/gz/~/səz/)

Background: Allomorphy

Morphemes can have variants depending on distinct environments:

Phonologically conditioned (rule-based)

cats~*dogs*~*buses*
(/ts/~/gz/~/səz/)

Syntactically conditioned (suppletive)

go~*went*~*gone*
present, past, past.part

Background: Allomorphy

Morphemes can have variants depending on distinct environments:

Phonologically conditioned (rule-based)

cats~*dogs*~*buses*
(/ts/~/gz/~/səz/)

Syntactically conditioned (suppletive)

go~*went*~*gone*
present, past, past.part

Q_1 : **Are the different kinds processed differently?**

Background: Allomorphy

Morphemes can have variants depending on distinct environments:

Phonologically conditioned (rule-based)

cats~*dogs*~*buses*
(/ts/~/gz/~/səz/)

Syntactically conditioned (suppletive)

go~*went*~*gone*
present, past, past.part

Q_1 : **Are the different kinds processed differently?**

Q_2 : **Does allomorphy impact processing in the first place?**

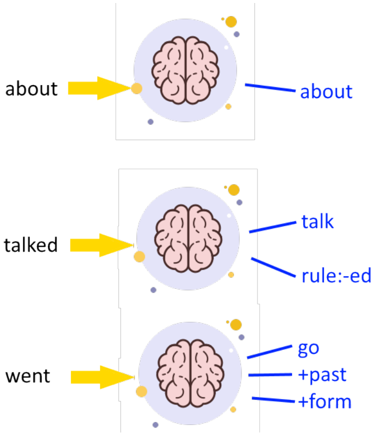
Hypothesis

- Allomorphy is unrelated to meaning and lexical categories (Nevins 2011) - it is a purely **grammatical factor**.
- Grammatical factors have been shown to impact lexical processing (Fruchter et al. 2013, but see Hay and Baayen 2005).

Hypothesis

Sensitivity to levels of **allomorphic complexity impacts processing time** along the cline *none<rule<suppletion*.

Hypothesis



Methods: Allomorphy

We introduced three predictors reflecting allomorphic complexity:

	beat	***beaten***	***beating***	***beats***	***dog***	***dogs***	***about***
ParHas	suppl	suppl	suppl	suppl	rule	rule	none
HasAllos	suppl	none	none	none	rule	none	none
IsAllo	suppl	suppl	none	rule	none	rule	none

Table: Examples of Allomorphy Coding

- Items were coded across lexical categories.
- Each item was assigned its highest level of complexity.
- Predictors were then contrast-coded for use in the models.

Methods: Corpus Data

We simulated an experiment via stepwise linear mixed-effects regression modelling with the following design:

- British Lexicon Project dataset (Keuleers et al. 2012);
- 804,633 observations across 37 participants;
- RTs and Accuracy as dependent variables;
- standard predictors (see below);
- allomorphy predictors across different models.

Findings: Paradigm-by-Paradigm

Predictor	Estimate	SE	*p*
(Intercept)	0.2482	0.0069	<0.001
Lemma Frequency	−0.0945	0.0005	<0.001
Orthographic Length	0.0731	0.0003	<0.001
Orthographic Similarity	−0.0004	0.0001	<0.001
Inflectional Entropy	−0.0079	0.0001	<0.001
ParHas - *allo*VS*none*	0.0654	0.0030	<0.001
ParHas - *suppl*VS*rule*	0.0929	0.0070	<0.001
ParHas - (*allo*VS*none*)*LemmaFreq	−0.0085	0.0004	<0.001
ParHas - (*suppl*VS*rule*)*LemmaFreq	−0.0135	0.0007	<0.001

Table: Regression Results for Paradigms (normalized RT)

- Directionalities of standard predictors looking normal.

Findings: Paradigm-by-Paradigm

Predictor	Estimate	SE	*p*
(Intercept)	0.2482	0.0069	<0.001
Lemma Frequency	−0.0945	0.0005	<0.001
Orthographic Length	0.0731	0.0003	<0.001
Orthographic Similarity	−0.0004	0.0001	<0.001
Inflectional Entropy	−0.0079	0.0001	<0.001
ParHas - *allo*VS*none*	0.0654	0.0030	<0.001
ParHas - *suppl*VS*rule*	0.0929	0.0070	<0.001
ParHas - (*allo*VS*none*)*LemmaFreq	−0.0085	0.0004	<0.001
ParHas - (*suppl*VS*rule*)*LemmaFreq	−0.0135	0.0007	<0.001

Table: Regression Results for Paradigms (normalized RT)

- Allomorphy generally inhibitory, suppletive more so than regular.

Findings: Paradigm-by-Paradigm

Predictor	**Estimate**	**SE**	***p***
(Intercept)	0.2482	0.0069	<0.001
Lemma Frequency	−0.0945	0.0005	<0.001
Orthographic Length	0.0731	0.0003	<0.001
Orthographic Similarity	−0.0004	0.0001	<0.001
Inflectional Entropy	−0.0079	0.0001	<0.001
ParHas - *allo*VS*none*	0.0654	0.0030	<0.001
ParHas - *suppl*VS*rule*	0.0929	0.0070	<0.001
ParHas - (*allo*VS*none*)*LemmaFreq	−0.0085	0.0004	<0.001
ParHas - (*suppl*VS*rule*)*LemmaFreq	−0.0135	0.0007	<0.001

Table: Regression Results for Paradigms (normalized RT)

- The more frequent a complex form, the quicker it is processed.

Findings: Paradigm-by-Paradigm

Predictor	**Estimate**	**SE**	***p***
(Intercept)	0.2482	0.0069	<0.001
Lemma Frequency	−0.0945	0.0005	<0.001
Orthographic Length	0.0731	0.0003	<0.001
Orthographic Similarity	−0.0004	0.0001	<0.001
Inflectional Entropy	−0.0079	0.0001	<0.001
ParHas - *allo*VS*none*	0.0654	0.0030	<0.001
ParHas - *suppl*VS*rule*	0.0929	0.0070	<0.001
ParHas - (*allo*VS*none*)*LemmaFreq	−0.0085	0.0004	<0.001
ParHas - (*suppl*VS*rule*)*LemmaFreq	−0.0135	0.0007	<0.001

Table: Regression Results for Paradigms (normalized RT)

- Same general pattern replicated within-paradigm.
- Same pattern for Accuracy, but the models failed to converge.

Findings: Discussion

- Allomorphy generally inhibitory (*none*<*rule*<*suppl*): words and paradigms with allomorphic complexity are slower to parse.
  - ⇒ **Obligatory decomposition (cf. Fruchter et al. 2013)?**

Findings: Discussion

- Allomorphy generally inhibitory (*none*<*rule*<*suppl*): words and paradigms with allomorphic complexity are slower to parse.
  - ⇒ **Obligatory decomposition (cf. Fruchter et al. 2013)?**
- Frequency interactions generally facilitatory (*suppl*<*rule*<*none*): if a complex form is really frequent, it gets processed more quickly despite its complexity.
  - ⇒ **Subregularity effects (cf. Albright and Hayes 2003)?**

Conclusions

- Sensitivity to allomorphy is a consistently significant factor.
  - ⇒ **We need to account for grammatical information in (visual) word processing.**
- Findings favour approaches that incorporate grammatical information, e.g. Marantz 2013, Fruchter et al. 2013.
- Further steps include, but are not limited to:
  1. Cross-linguistic replication on database corpora (e.g. in Dutch);
  2. Checking for sub-regularity effects (Albright and Hayes 2003).

Danke!

This study was supported by AL 554/8-1 (DFG Gottfried Wilhelm Leibniz Preis 2014 to Artemis Alexiadou) and the Research Unit on Experimental Syntax and Heritage Languages (**RUESHeL**).

References

Albright, A. & Hayes, B. (2003). Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition* 90(2):119-161.
Baayen, R. H, Milín, P., Durdević, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review* 118:438-481.
Fruchter, J., Stockall, L., & Marantz, A. (2013). MEG masked priming evidence for form based decomposition irregular verbs. *Frontiers in Human Neuroscience* 7(798):1-16.
Hay, J. & Baayen, R. H. (2005). Shifting paradigms: gradient structure in morphology. *Trends in Cognitive Sciences* 9:342-348.
Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods* 44:287-304.
Marantz, A. (2013). No escape from morphemes in morphological processing. *Language and Cognitive Processes* 28:905-916.
Marelli, M., Amenta, S., & Crepaldi, D. (2015). Semantic transparency in free stems: The effect of Orthography Semantics Consistency in word recognition. *Quarterly Journal of Experimental Psychology* 68:1571-1583.
Milín, P., Durdević, D. F., & Moscoso del Prado Martín, F. (2009). The simultaneous effects of inflectional paradigms and classes on lexical recognition: *Evidence from Serbian. Journal of Memory and Language* 60:50-64.
Moscoso del Prado Martín, F., Kostić , A., & Baayen , R. H. (2004). Putting the bits together: an information theoretical perspective on morphological processing. *Cognition* 94:1-18.
Nevins, Andrew. 2011. Phonologically conditioned allomorph selection. In: *The Blackwell companion to phonology*, 2357-2382.

Within-Paradigm Model: RT

Predictor	Estimate	SE	*p*
(Intercept)	0.1588	0.0080	<0.001
Frequency	−0.0837	0.0008	<0.001
Orthographic Length	0.0584	0.0003	<0.001
Orthographic Similarity	−0.0004	0.0001	<0.001
Inflectional Entropy	−0.0030	0.0001	<0.001
HasAllos - *alloVSnone*	0.0334	0.0028	<0.001
HasAllos - *supplVSrule*	0.1508	0.0076	<0.001
HasAllos - (*alloVSnone*)*Freq	−0.0078	0.0004	<0.001
HasAllos - (*supplVSrule*)*Freq	−0.0232	0.0010	<0.001
IsAllo - *alloVSnone*	−0.0205	0.0025	<0.001
IsAllo - *supplVSrule*	0.0528	0.0004	<0.001
IsAllo - (*alloVSnone*)*Freq	0.0022	0.0003	<0.001
IsAllo - (*supplVSrule*)*Freq	−0.0054	0.0009	<0.001

Omnibus Model: RT

Predictor	**Estimate**	**SE**	***p***
(Intercept)	0.2308	0.0087	<0.001
Frequency	−0.0889	0.0010	<0.001
Orthographic Length	0.0595	0.0004	<0.001
Orthographic Similarity	−0.0061	0.0001	<0.001
Inflectional Entropy	−0.0030	0.0001	<0.001
ParHas - *allo*VS*none*	−0.0603	0.0029	<0.001
ParHas - *suppl*VS*rule*	−0.1013	0.0070	<0.001
ParHas - (*allo*VS*none*)*Freq	0.0037	0.0004	<0.001
ParHas - (*suppl*VS*rule*)*Freq	0.0016	0.0010	<0.001
HasAllos - *allo*VS*none*	0.0796	0.0035	<0.001
HasAllos - *suppl*VS*rule*	0.2438	0.0097	<0.001
HasAllos - (*allo*VS*none*)*Freq	−0.0106	0.0005	<0.001
HasAllos - (*suppl*VS*rule*)*Freq	−0.0251	0.0013	<0.001
IsAllo - *allo*VS*none*	0.0164	0.0031	<0.001
IsAllo - *suppl*VS*rule*	0.1208	0.0083	<0.001
IsAllo - (*allo*VS*none*)*Freq	0.0002	0.0004	<0.001
IsAllo - (*suppl*VS*rule*)*Freq	−0.0057	0.0011	<0.001

Inflectional Entropy

Adapted from Information Theory, Inflectional Entropy reflects the uncertainty when choosing between members of a paradigm:

Entropy

$$H_f = -\sum \frac{F_i}{\sum_1^c F_m} log_2 \frac{F_i}{\sum_1^c F_m} \tag{1}$$

Inflectional Entropy

$$H_f = -\sum \frac{\frac{F_i}{R_i}}{\sum_1^c \frac{F_m}{R_m}} log_2 \frac{\frac{F_i}{R_i}}{\sum_1^c \frac{F_m}{R_m}} \tag{2}$$

- quantifies the energy spent when a paradigm is activated
- depends on the distributional probabilities of the paradigm's members

⇒ more uniform distributions = higher Entropy = faster RTs in processing