What is in a morpheme? Theoretical, experimental and computational approaches to the relation of meaning and form in morphology

Workshop at the 50th Annual Meeting of the *Societas Linguistica Europaea*, Zurich, 2017

Organizers: Stela Manova[1], Harald Hammarström[2] and Itamar Kastner[3]
[1]University of Vienna, http://homepage.univie.ac.at/stela.manova/
[2]Max Planck Institute for the Science of Human History (Jena), http://www.shh.mpg.de/employees/48214/55811
[3]Humboldt University of Berlin, http://itamarkast.net/

There are enough examples in science that obvious things are the most difficult to explain: issues such as how inorganic matter turns into organic or how a child learns to understand language. There is a similar problem in morphology: morphemes consist of phonemes but only the former can be associated with meaning (systematically) and it is a non-trivial question how this association happens.

There are three possible ways to approach the relation of meaning and form:

A. Form and meaning emerge simultaneously
B. The association is from meaning to form
C. The association is from form to meaning.

The most important difference between these scenarios consists in the fact that in scenarios B and C meaning may be assigned at the level of word, i.e. one may claim that morphemes do not have meaning of their own or even that there are no morphemes at all (in scenario B). (Information (syntactic/morphological/morphosyntactic) that does not refer to (phonological) form is called 'meaning' herein.)

Theoretical, experimental and computational linguistics approach word structure from different perspectives and seem to diverge with respect to which is the "right" scenario. Theoretical linguistics is interested in generalizations over meaning (features) (scenarios A and B), both within languages and typologically: e.g., only a language with PLURAL can have DUAL (Greenberg 1963). Experimental linguistics researches perception, parsing, processing and production of word structure; computational linguistics is focused on parsing and distribution of word structure. Consequently, both experimental and computational linguistics follow scenario C and their findings seem to contradict theoretical linguistics. Nevertheless, theoretical linguists (seem to) agree that speakers have somewhat reliable intuitions about n-gram frequency over sub-word units. Thus, the goals of this workshop are threefold: to encourage interdisciplinary discussion; to clarify and unify assumptions; and to pave the way for collaboration.

Let us illustrate the different scenarios. *Minimalist Morphology (MM)* (Wunderlich 1996) is an example of **scenario A**. In MM, a morpheme has form and meaning; (inflectional) morphemes are heads; a morpheme minimally includes a representation of its phonological form, a specification of the base's category and an output specification:

(1) German: /st/; [+min]; [+2]/+V (Stiebels 2011)

[+min] indicates that the form is a morpheme; [+2] = 2 person; "+V" indicates that *-st* attaches to verbs; the slash / stands for "output/input".

In *Realizational Morphology (RM)*, theories such as *Paradigm Function Morphology (PFM)* (Stump 2001) and *Distributed Morphology (DM)* (Halle & Marantz 1993), meaning and form are modeled separately and semantic derivation precedes formal derivation, the so-called late insertion (**scenario B**). Roughly, one can predict form based on meaning, while the opposite does not hold and therefore the form-to-meaning direction is not activated in RM.

PFM manipulates morphosyntactic property sets:

(2) $PF(\langle L,\sigma\rangle) = \langle R,\sigma\rangle$ (Stewart & Stump 2007)

The value of the paradigm function (PF) of a paradigm cell ⟨L,σ⟩ (L=LEXEME) is the pairing of this cell's realization R with the morphosyntactic property set σ. Such a theory does not necessarily need morphemes.

DM relies on syntactic structures and 'morpheme', [PAST] in (3), is an abstract unit that refers to a syntactic terminal node (INFL in this case) and its content, not to the phonological expression of that terminal:

(3) Vocabulary of English (fragment) (Bobaljik 2015)
a. [PAST] ↔ -t /]$_V$ _ ; where V ∈ {*dream, dwell* etc.}
b. [PAST] ↔ Ø /]$_V$ _ ; where V ∈ {*run, hit, fly* etc.}
c. [PAST] ↔ -d /]$_V$ _

To explain the fact that in DM syntactic structure derives morphological structure, Müller (2016) refers to the meaning-form dichotomy as two different dimensions of a linguistic unit: a representational and an algorithmic one respectively.

The assumption that meaning precedes exponence is claimed to make RM superior in comparison to incremental theories of morphology that follow scenario A because in RM derivation takes place at an abstract level and is always compositional, while exponence often entails idiosyncrasies.

On the other hand, affixes are directly accessible through their form (**scenario C**) and can be identified and processed even without having a contentful stem to attach to, as evidenced by recent *psycholinguistic studies*. Crepaldi et al. (2016) demonstrates that prime non-words facilitate lexical decisions to target words ending with the same suffix and that the priming effect depends on the affix position in the non-word. Lázaro et al. (2016) uses suffixes as primes and shows that the prime suffix facilitates the recognition of words ending with that suffix. Both studies conclude that the priming effect of suffixes is not orthographic but morphological. Similar findings are reported in Beyersmann et al. (2016). Manova & Brzoza (2015) shows that if provided with a list of existing and non-existing suffix combinations without stems, native speakers can judge which combinations exist and which do not.

A subfield of *Computational Linguistics* called *Unsupervised Learning of Morphology (ULM)* is concerned with learning natural language's morphology from unannotated text corpora. Since form is given (but meaning is not), to segment words into morphemes, ULM relies on comparison, grouping and weighting of substrings (of letters) and their frequencies (see Hammarström & Borin 2011 for an overview of ULM research). Semantic representations of extracted form-based morphemes may also be inferred using the principle that semantically related morphemes tend to occur in similar contexts (e.g., Baroni et al. 2002).

The workshop will provide a platform for exchange of ideas and for an interdisciplinary discussion of the meaning-form issue in morphology. The questions to be addressed include, but are not limited to, the following:

1. What information is encoded in a morpheme?
2. Does an analysis with emphasis on either meaning (scenario B) or form (scenario C) provide evidence for a (complete) separation of form and meaning in the morpheme?
3. Could it be that a morpheme relates meaning and form and semantic stimuli activate derivation through meaning, while formal stimuli activate access through form?
4. How does morphology "emerge" in fieldwork, i.e. how does a fieldworker decide that something is a morpheme, is it according to A, B or C?
5. How does morphology "emerge" in child language?
6. What exactly does a language borrow when it borrows morphological structure such as, e.g., a PLURAL nominal marker, if that language already has PLURAL and its speakers are not expected to be able to perform a morpheme analysis of the donor language's words?

7. If important generalizations are (necessarily) stated over either meaning or form, how are the two types of generalizations related to one another; and are they both needed for an adequate characterization of speakers' knowledge of their language?
8. Can a computational analysis based on n-gram frequency distributions and distributional semantics account for the kinds of generalizations that interest theoretical linguists and motivate the B (or A) perspective?

As an alternative, non-linguistic source of inspiration, we would like to turn your attention to the following video on how computers learn to understand pictures: https://www.youtube.com/watch?v=40riCqvRoMs (the speaker, Fei-Fei Li, is an Associate Professor of Computer Science at Stanford University). Computer vision is one of the most important areas of research in machine learning and many striking analogies with linguistic analyses can be made.

Important Dates

15 January 2017: deadline for submission of 500-word abstracts (excluding references) to SLE for review (submission open to everybody interested in the workshop, submission guidelines at: http://sle2017.eu/submission-guidelines)
31 March 2017: notification of paper acceptance
10-13 September 2017: SLE conference in Zurich

References

Baroni, M., J. Matiasek & H. Trost (2002). Unsupervised discovery of morphologically related words based on orthographic and semantic similarity. In Mike Maxwell (ed.), *Proceedings of the Workshop on Morphological and Phonological Learning of ACL/SIGPHON-2002*, 48-57. East Stroudsburg PA: ACL.

Beyersmann, E., J.C. Ziegler, A. Castles, M. Coltheart, Y. Kezilas & J. Grainger (2016). Morpho-orthographic segmentation without semantics. *Psychon Bull Rev* 23(2): 533-9. DOI: 10.3758/s13423-015-0927-z.

Bobaljik, J. (2015). Distributed Morphology. Ms. University of Connecticut. Available at: http://bobaljik.uconn.edu/papers/DM_ORE.pdf

Crepaldi D., L. Hemsworth, C. J. Davis & K. Rastle (2016). Masked suffix priming and morpheme positional constraints. *Q J Exp Psychol (Hove)* 69(1): 113-28. DOI: 10.1080/17470218.2015.1027713.

Greenberg, J. H. (ed.) (1963). *Universals of Human Language*. Cambridge, Mass: MIT Press.

Halle, M. & A. Marantz (1993). Distributed morphology and the pieces of inflection. In Hale K. and S. J. Keyser (eds.), *The view from building 20*, 111-176. Cambridge, MA: MIT Press.

Hammarström, H. & L. Borin (2011). Unsupervised Learning of Morphology. *Computational Linguistics* 37(2): 309-350.

Lázaro, M, V. Illera & J. Sainz (2016). The suffix priming effect: Further evidence for an early morpho-orthographic segmentation process independent of its semantic content. *Q J Exp Psychol (Hove)* 69(1): 197-208. DOI: 10.1080/17470218.2015.1031146.

Manova, S. & B. Brzoza (2015). Suffix combinability and the organization of the mental lexicon. Paper presented at the *10th Mediterranean Morphology Meeting (MMM)*, Haifa, Israel, 7-10 September 2015. Handout available at: http://homepage.univie.ac.at/stela.manova/uploads/1/2/2/4/12243901/manova_brzoza_mmm2015.pdf

Müller, G. (2016). Minimize Satisfaction in Harmonic Serialism. Paper presented at the workshop *The Word and the Morpheme*. Humboldt University of Berlin, 22.9-24.9.2016. Abstract available at: https://www.angl.hu-berlin.de/department/staff/artemis_alexiadou/abstracts/muller.pdf

Stewart, T. & G. Stump (2007). Paradigm Function Morphology and the Morphology–Syntax Interface. In G. Ramchand & C. Reiss (eds.), *The Oxford handbook of linguistic interfaces*, 383-421. Oxford: Oxford University Press.

Stiebels, B. (2011). Minimalist Morphology. Ms. University of Leipzig. Available at: http://home.uni-leipzig.de/stiebels/papers/handbook_morphology_stiebels_mm-2011.pdf

Stump, G. T. (2001). *Inflectional morphology*. Cambridge: Cambridge University Press.

Wunderlich, D. (1996). Minimalist morphology: the role of paradigms. In G. Booij & J. van Marle (eds.), *Yearbook of Morphology 1995*, 93–114. Dordrecht: Kluwer.