*Li8: Morphology/Lent 2018*

# Modularity and the Paradigm Cell Filling Problem

Jim Blevins (jpb39)
M11-12/LB9/29.01

# Challenges to morphological modularity

- Attempts to establish a fixed demarcation between morphology and phonology faces a range of theoretical challenges:
  - The status of phenomena as 'morphological' or 'phonological' depends on the abstractness of phonological analyses.
    - Interleaving stress assignment with affixation in English permits a phonological analysis to recapitulate historical developments that produced the word stock of the language.
  - The appeal to 'morphological conditioning' allows formerly productive phonological processes to continue to operate, over domains that cannot be characterized in phonological terms.
    - Phonological truncation in Nominative Singular in Lardil represents the phonologization of a morphological pattern.

# Sub-phonemic subversion

- The 'sub-phonemic' contrasts summarized in the previous lecture are often interpreted conservatively, as evidence that there is no true homonymy, and/or that phonemic transcriptions are deficient.

- Yet these contrasts have more far-reaching implications:

  - Phonemicization provides a discrete description of continuous variation, but discrete descriptions are task-dependent and can't encapsulate the discriminable contrasts of a sound system.

  - As Kemps et al. (2005) show, sub-phonemic cues can be misaligned with morphotactic structure, which subverts naive notions about (i) form identity, and (ii) the synchronic relevance of morphotactic divisions and the functions of their parts.

  - Thus phonemic transcriptions may be glorified orthographies and morphotactic structure a record of morphologization.

# The logic of sub-phonemic diachrony

- ❖ Although the implications of sub-phonemic contrasts for synchronic systems have only recently been recognized, discriminable sub-phonemic contrasts have long played a fundamental role in diachronic analyses of morphologization.

- ❖ As in the synchronic cases discussed earlier, a contrast may arise in a context where it is initially not phonologically distinctive.

- ❖ However, if a contrast comes to be perceived and produced systematically by speakers, it may facilitate sound changes that create contexts in which it becomes synchronically phonemic.

- ❖ It is usually assumed that the contrast arises first, and preserves its functional load in the new structure that it facilitates, rather than that a phonemic contrast arises spontaneously in the structure.

# The sub-phonemic origin of Finnic gradation

Recall the stem alternations in noun declensions exhibiting 'weakening' gradation:

|  | Sg | Plu |
|---|---|---|
| Nom | vakk | vakad |
| Gen | vaka | vakkade |
| Part | vakka | vakkasid |
| Illa | vakasse | vakkadesse |
| ⋮ | ⋮ | ⋮ |
| Com | vakaga | vakkadega |

'bushel'

# The discriminability of stem variation as a prerequisite for the grammaticalization of a sub-phonemic contrast

* Viitso (2003: 163) traces gradation to contrasts between "the nominative and genitive forms of the stem for 'bushel', cf. *vakka* : *vakan*" in which "the geminate stop was shortened before a closed second syllable".

* As in Modern Dutch, prosodic differences conditioned by syllable structure in the Finnic languages must have been discriminable, functional load-bearing contrasts at the point where the conditioning environments were still present.

* At this point, *vakka* and *vakan* would not have defined a minimal pair with respect to length, since they are distinguished by the presence/absence of *-n*.

* Hence, at this stage, an analysis of *vakan* that associated genitive case features exclusively (or even primarily) with the ending *-n* would have been as misleading as an analysis of *geiten* that associates plurality exclusively with *-en*.

# The sub-phonemic origin of Finnic grade

*there are no other known phonetic preconditions for the rise of quantity correlation of long syllables than the phonetic alternation of first syllables depending on the openness vs. closedness of the following unstressed syllable in Finnish …* ***In order to develop into a morphophonological gradation, this automatic alternation needed only to be perceived****. Only a perceived alternation could be accepted as a linguistic norm and subjected to further polarization and reinterpretation of sound patterns caused by openness vs. closedness of the second syllable of a foot as a function of syllabic accentuation in initial syllables … This polarization as a norm caused the quantity alternation of stems in their morphological paradigms so that the stressed syllable was longer before an open syllable and shorter before a closed syllable.* ***This quantity alternation was retained, in principle, also after the former conditioning environments changed.*** *(Viitso 2003: 164f.)*

# Implications for hierarchies of units and levels?

- ❖ The kinds of challenges that arise in distinguishing morphology from phonology also arise in distinguishing syntax.

- ❖ There is a long and mostly inconclusive literature on the status of 'clitics', periphrastic constructions, multiword expressions, etc.

- ❖ These challenges call into question conceptions of a hierarchy of levels, where a level is defined by simple units that

  - ❖ encapsulate the contrastive features at that level, and

  - ❖ are combined to form complex units that function as the simple units of the next level up in the hierarchy.

- ❖ Yet it is possible that similar operations and relations apply to sequences of different size, with different effects at different scales.
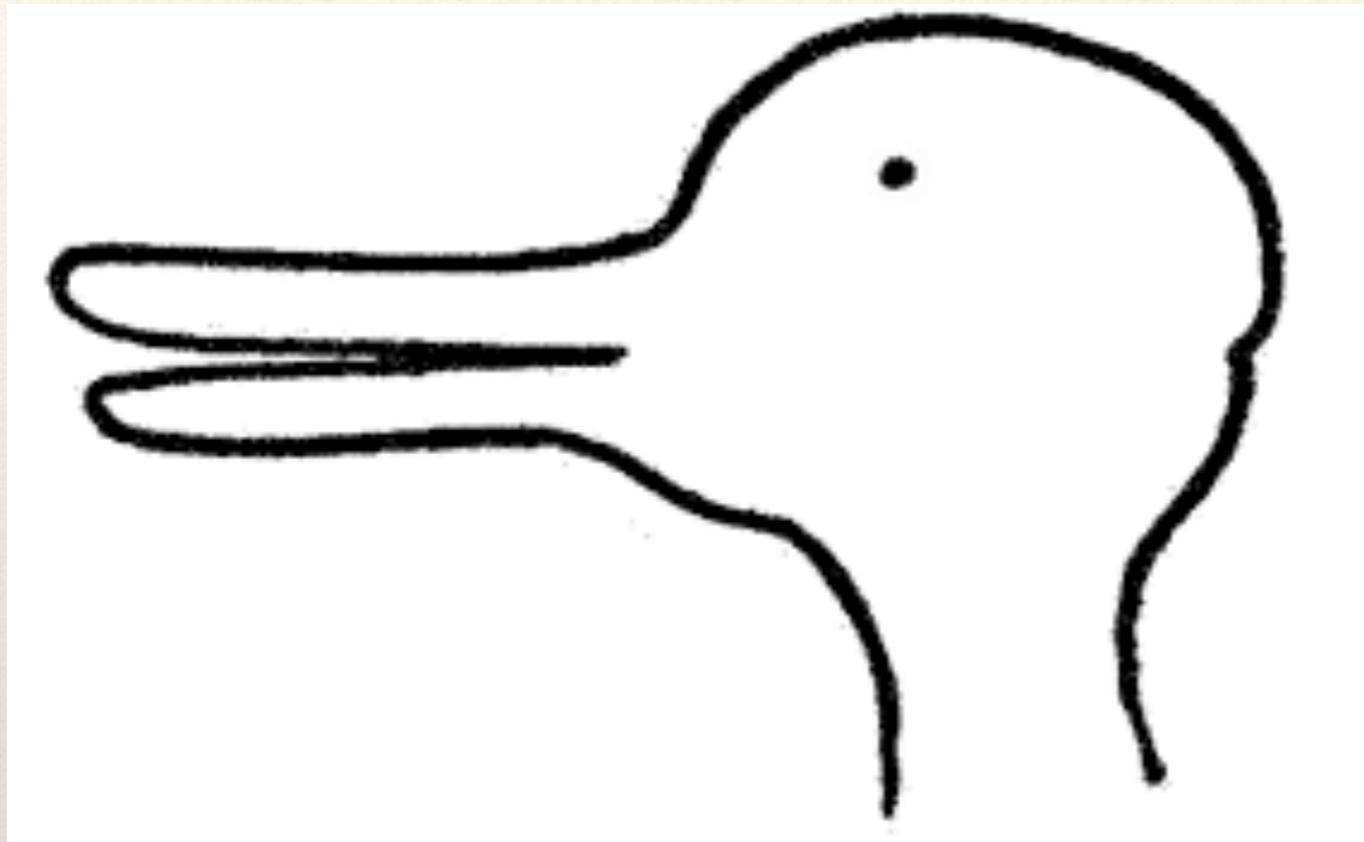
# The Paradigm Cell Filling Problem (PCFP)

❖ How does a speaker come to know the inflected (and derived) forms of the lexical items of a language?[†]

❖ [†]Recently formulated in these terms in Ackerman et al. (2009) but core motivation for the proportional analogies in Paul (1880), and implicit as far back as Aristotle's position on the 'analogy/anomaly' debate.

# A Duck-Rabbit Problem



❖ At first glance, the PCFP seems to be a learning problem. On further reflection, it appears to raise questions about the structure of a system. On yet further reflection, it seems to suggest that questions about structure are ultimately learning problems.

❖ So we start with a duck, then we see a rabbit, and we end up with a duck-rabbit.

Ludwig Wittgenstein
(*Philosophical Investigations* 1953)

# A brute force solution?

❖ Could speakers memorize the forms that they encounter until they have filled in all (or most) of the cells?

# The challenge of inflectional profusion

❖ Conservative estimates of speakers' lexical knowledge place adult vocabulary size in the neighbourhood of 20-35K words.

❖ In Estonian, nouns have at least 28 inflected forms, and verbs have upwards of 30 basic forms (which occur in a range of periphrastic perfect and negative formations that triple the size of paradigms).

❖ On the most conservative assumptions about vocabulary size and paradigm size, a speaker would need to encounter and learn (possibly on a single exposure) at least 500K inflectional forms.

# The challenge of inflectional exuberance

❖ Estonian is by no means an extreme example:

❖ In Georgian, transitive verbs have 250 inflected forms, distributed across 10 paradigms in three inflectional series (Tschenkéli 1958).

❖ Kibrik (1998) estimates that a single Archi verb may (at least in principle) have up to 1.5M (inflected and derived) forms.

❖ Even if we regard this estimate as exaggerated, it is clear that languages may have large inflectional paradigms and that learning these form inventories poses a significant acquisition challenge.

# The 'Too Big to Store' argument

❖ One influential argument against a brute-force memorization learning strategy is that there would be too many forms to store:

all versions of the FLH [Full Listing Hypothesis] are untenable as hypotheses about speakers of human languages in general. The argument is based on an examination of morphological complexity in agglutinative languages (with Turkish as the central example) and on a demonstration that the number of forms corresponding to a single noun or verb root is considerably larger than would be consistent with the assumption that all forms are listed in a mental lexicon. (Hankamer:1989: 393)

# The contents of the 'mental lexicon'

❖ Linguistic arguments against the FLH are directed against a conception of the mental lexicon in which all of the **potential** word forms of a language(many looping through derivation) are listed.

❖ This conception might have practical advantages for NLP systems in which the lexicon was populated at 'compile time' and efficient lexical lookup could be executed as part of the 'runtime' system.

❖ For pedagogical purposes, it might be useful to multiply out the possible forms for learners of morphologically intricate languages.

❖ However, no cognitively plausible model of a speaker's lexical knowledge would assume anything like a lexicon that lists all the potential forms of a language (as opposed to those in circulation).

# The format of the 'mental lexicon'

❖ It is also unclear what conclusions can be drawn about the 'mental lexicon' based on assessments of storage demands that abstract away from:

> ❖ a specification of the format in which the contents would be 'stored',
>
> ❖ informed estimates of the 'storage capacity' of the brain,
>
> ❖ information about 'compression techniques' that might be exploited by the brain, and other considerations of this kind.

❖ The omission of these factors fits with the abstractness of the estimated form set, highlighting the complete dissociation between the "mental lexicon" in Hankamer (1989: 393) and any cognitively-relevant notion of a speaker's lexical knowledge.

❖ Thus the argument against the FLH operates with a fully idealized measure of the size of the mental lexicon and an equally idealized notion of its composition.

❖ The qualification 'mental' in the 'mental lexicon' is a misnomer; the linguist's 'lexicon' is a purely theoretical construct, modelled on the written dictionaries of the pedagogical tradition and the dedicated storage of a digital computer