

Allomorphy: Old Concept, Big Data, New Model

New methods of analysis for rival polysemous affixes



Author: Dr. Anna Endresen

University of Tromsø: The Arctic University of Norway
Department of Language and Linguistics
CLEAR research group

ABSTRACT Many linguistic concepts were first introduced in the Structuralist Era, the time when linguists believed in clear-cut oppositions and did not have access to large corpora. I find that allomorphy is a scalar phenomenon that can be best captured in terms of a radial category.

The new model is based on quantitative methods and can handle semantic dissimilation of allomorphs as well as distributional overlap. I show how statistical models turn allomorphy into a measurable and verifiable correspondence of form and meaning.

THIS STUDY IN A NUTSHELL

Old Concept

I revisit an old concept of Allomorphy, which was first introduced to linguistics in the 1940s by American Structuralists. Despite fruitful discussions, the most rigid approach (Harris 1942) to Allomorphy persisted in the history of the field.

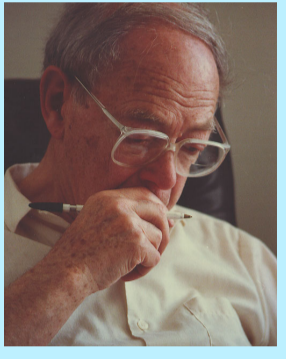
Big Data

I challenge this notion with data on 15 Russian prefixes (4,718 lexemes collected from the corpus and in 2 experiments with 60 and 120 subjects). I find that the conventional understanding of Allomorphy is a theoretical construct, an idealization. It fails to capture properties of data.

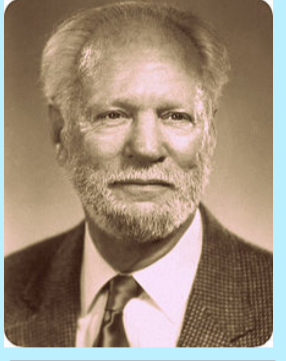
New Model

I propose an alternative model of Allomorphy. It is more accurate and realistic with regard to such properties of data as gradience, semantic dissimilation of allomorphs, and overlap in their distribution.

CONCEPT WITH STRUCTURALIST BAGGAGE (+/-)



Zellig Harris (1942): "We can arrange morpheme alternants into units in exactly the same manner as we arrange sound types into phonemes." →
• "A morpheme unit is a group of alternants which have the same meaning and complementary distribution."
MOST RIGID MODEL



Charles Hockett (1947): the analogy "(allo)phone : phoneme = morph : morpheme"
• Amendment: Non-contrastive distribution: i) **complementary distribution** or ii) **partial complementation**, i.e. free variation in the environments where both alternants can occur" (e.g. *you and me* vs. *you and I*).
LESS RIGID MODEL



Eugene Nida (1948): Morphemes are meaningful units, different from phonemes.
• Amendment: **No items that are different in form are absolutely identical in meaning.** → "From the difference in their distribution they acquire a certain difference in meaning."
FLEXIBLE MODEL

Coined the term **ALLOMORPH**

We can elaborate this flexible and non-absolute understanding of allomorphy and enrich it with advances of computational models, psycholinguistic experiments, and corpus data.

Prefixes	Number of analyzed verbs	Formal similarity	Etymological relationship	Semantics			Distribution		Status
				# of shared submeanings	Shared prototype	Distinct profiles	Size of overlap	Conditioning factors	
RAZ- RAS-	200	similar	related	share all 7 submeanings	share 'APART'	No	no overlap	phonology	Prototypical allomorphy
RAZ- RAZO-	210	similar	related	share all 7 submeanings	share 'APART'	No	no overlap	phonology & morphophonology	Standard allomorphy
S- SO-	1,156	similar	related	share all 6 submeanings	share both 'DOWNWARD' & 'CENTRIPETAL'	Yes: in 'CONCOMITANT ACTION'	15 minimal pairs	phonology, morphophonology, register, semantics	Non-Standard allomorphy
O- OB- OBO-	1,037	similar	related	share all 15 submeanings	share 'AROUND'	Yes: spatial vs. change-of-state	23 minimal pairs	phonology, semantics (type of base), prosody	Non-Standard allomorphy
PERE- PRE-	945	similar	related	share 8 out of 14 submeanings	share 'TRANSFER OVER/ACROSS'	Yes: spatial vs. intensity	22 minimal pairs	grammatical classes: verbs vs. non-verbs	Non-Standard allomorphy
VZ- VOZ-	384	similar	related	share all 9 submeanings	share 'UPWARD' but differ in height	Yes: spatial, metaphorical, aspectual	21 minimal pairs	semantics, register, aktionsart	Non-Standard allomorphy (borderline case)
VY- IZ-	998	not similar	different sources	share 10 out of 12 submeanings	share 'OUT of', but do not share 'ZIGZAG'	Yes: 'OUT of' vs. 'EXHAUST'	112 minimal pairs	semantics, register (prosody)	Non-Standard allomorphy (borderline case)
O- U-	155	not similar	unrelated	share the submeaning 'make X be Y'	different prototypes 'AROUND' and 'AWAY'	Not applicable	17 minimal pairs	qualitative vs. relational adjectival base	Non-Allomorphy; Closely associated rival morphemes
PRE- PRI- PRED-	10	similar	unrelated	no shared submeanings	different prototypes	Not applicable	some overlap	different semantics	Non-Allomorphy; Different morphemes with no association

NEW MODEL: ALLOMORPHY AS A RADIAL CATEGORY

Alternative to the all-or-nothing model:

Allomorphy is **broader** than its conventional understanding.

Allomorphy is a **scalar relationship** between morpheme variants – a relationship that can vary in terms of **closeness and regularity**.

Allomorphy is a **gradient** phenomenon – with a central prototype, standard exemplars and non-standard deviations.

The **core clear cases** of allomorphy can be viewed as prototypical rather than the only possible.

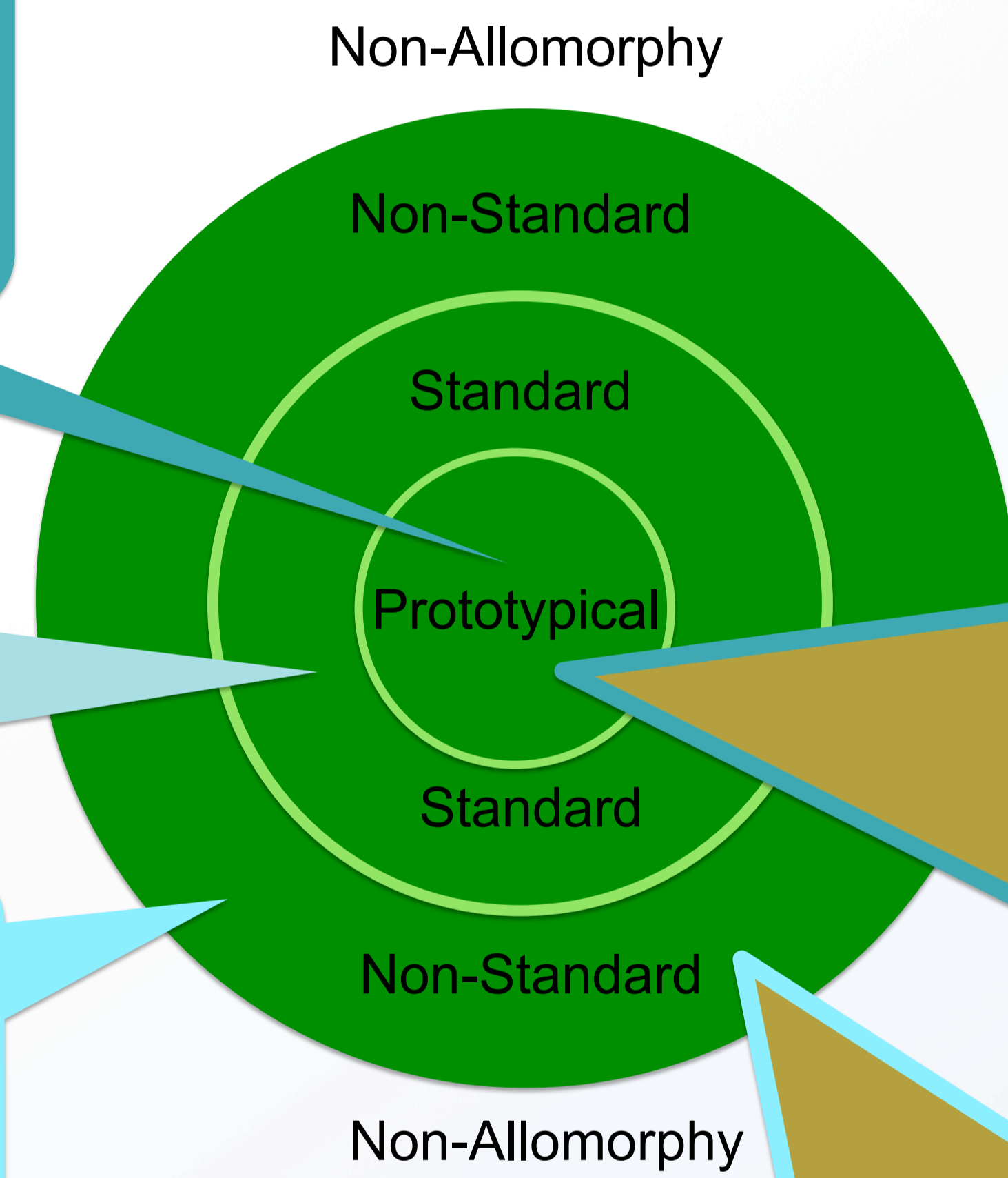
Deviations are recognized as Allomorphy or Non-Allomorphy on the basis of statistical measurements.

New distinctions:

Prototypical Allomorphy is characterized by the closest and most automatic association of formants. Typically phonologically conditioned by a regular, automatic, and productive phonological rule. E.g.: Russian prefixes RAZ-/RAS-

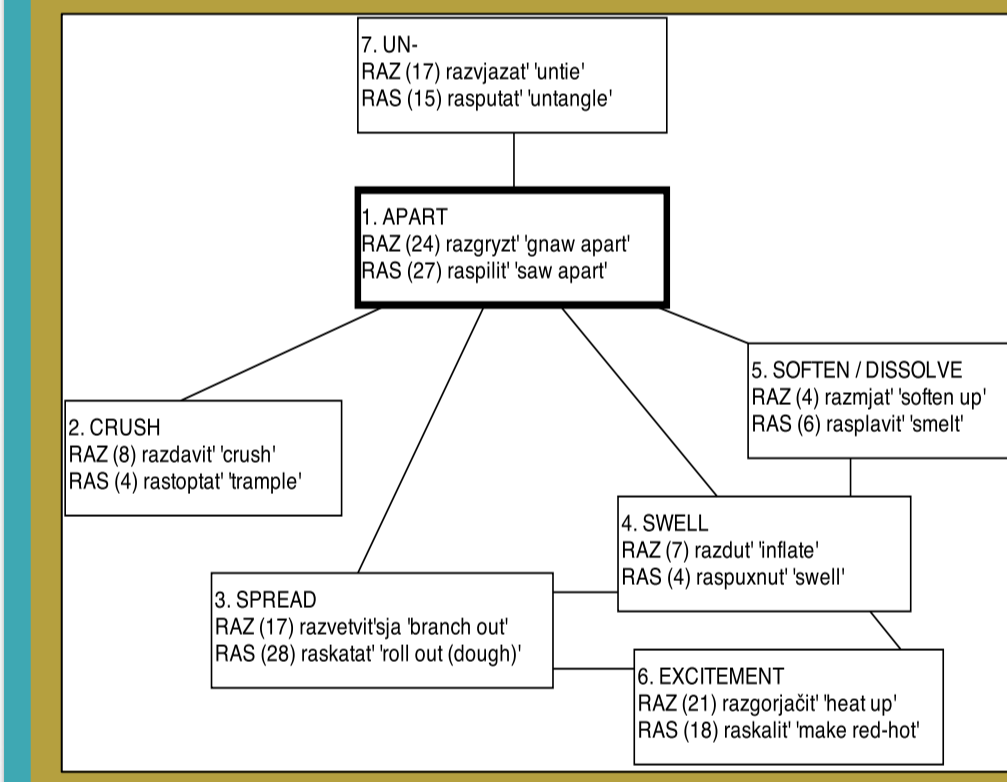
Standard Allomorphy – satisfies both criteria (identical meaning & complementary distribution), but is governed by factors other than (or in addition to) active phonology – morphophonology, register, semantics. E.g.: prefixes RAZ-/RAZO-

Non-Standard Allomorphy – violates one or both criteria BUT shows a strong semantic similarity or robust pattern of distribution. E.g.: Russian prefixes O-/OB-, S-/SO-, PERE-/PRE-, VZ-/VOZ-, VY-/IZ-

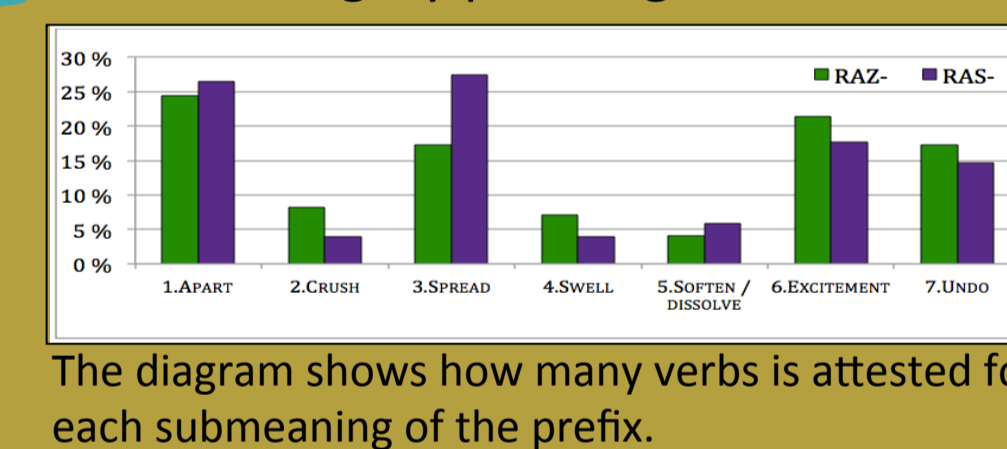


CASE STUDY OF THE PROTOTYPE: The Russian prefixes RAZ- / RAS- 'APART'

Modeling of prefix polysemy: 200 verbs



Radial category profiling:



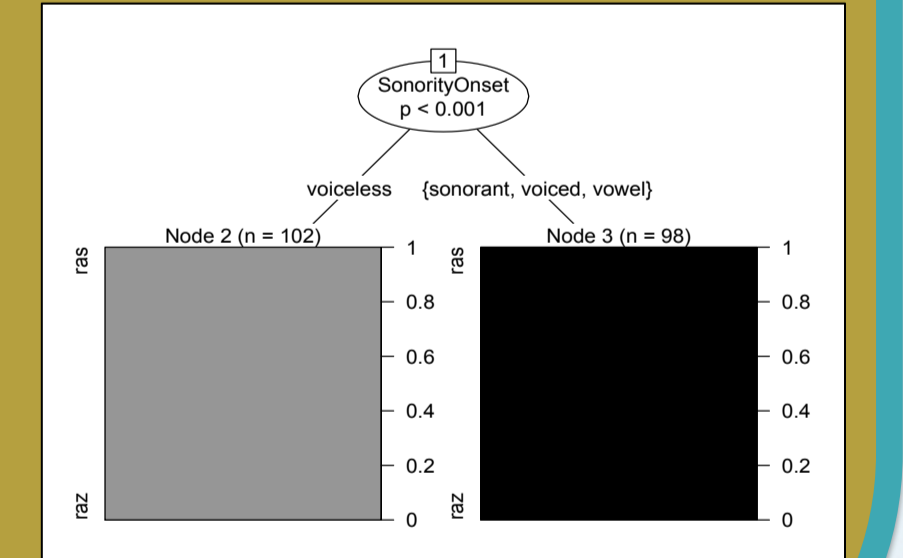
The diagram shows how many verbs is attested for each submeaning of the prefix. Distribution of RAZ- and RAS- across verbs and prefix submeanings is not significantly different: $p = 0.46$

→ Semantics plays no role in the distribution of RAZ- and RAS-.

The choice of RAZ- vs. RAS- is phonologically conditioned by a productive and exceptionless process of regressive voicing assimilation:



Sonority of the onset base (voiced vs. voiceless consonant) is the only predictor of the prefix:



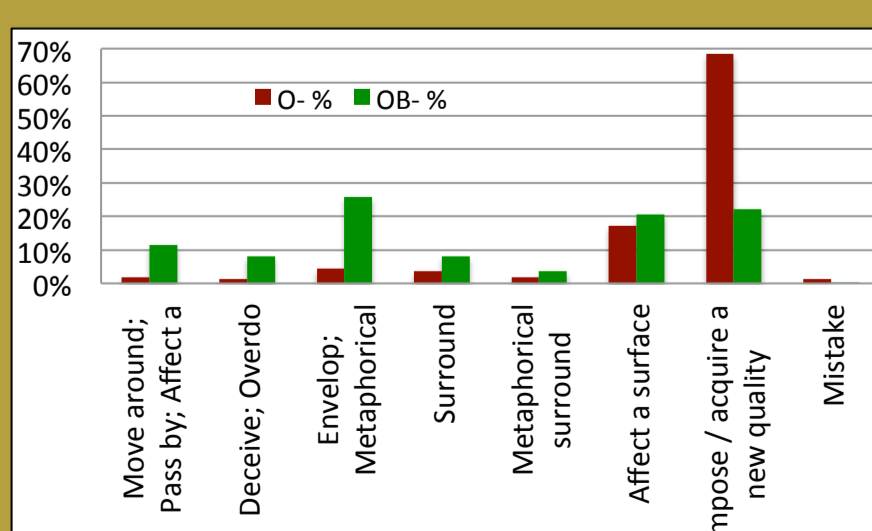
CASE STUDY OF NON-STANDARD ALLOMORPHY: The Russian prefixes O- / OB- 'AROUND'

SEMANTICS: Highly polysemous prefixes → How do we assess whether they are identical in meaning?

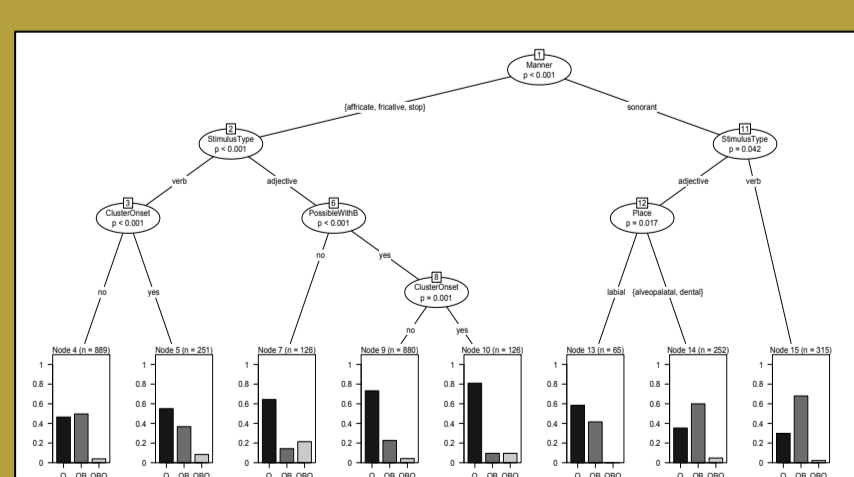
DISTRIBUTION: governed by several factors (phonological, semantic, prosodic) → How do we determine which factor is the most powerful?

Data: 1,037 verbs prefixed in O- and OB-
Single radial network of 15 submeanings

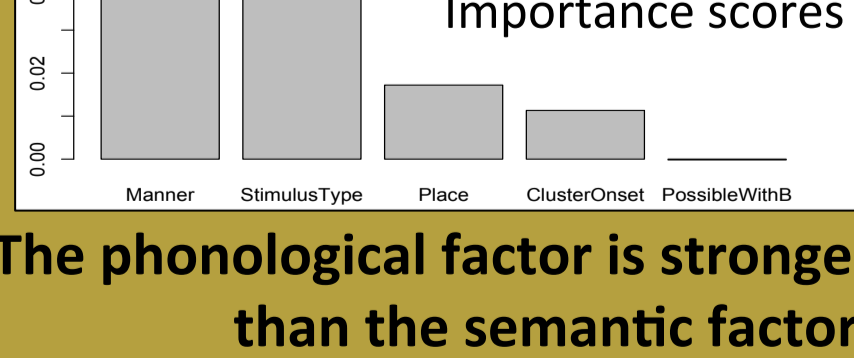
Radial category profiling: Different profiles of O- and OB- in terms of type frequency of verbs attested for each submeaning:



Classification tree model: Experimental data



Random Forests model: Importance scores

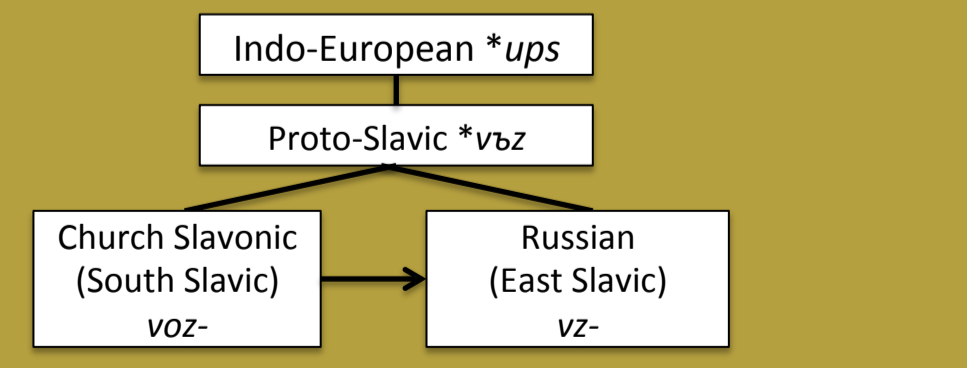


The phonological factor is stronger than the semantic factor.

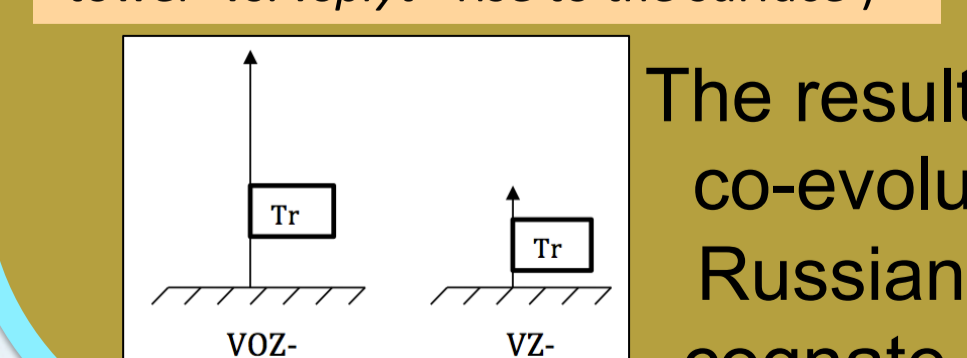
This results from the process of semantic dissimilation of former phonological variants.

CASE STUDY OF NON-STANDARD ALLOMORPHY: The Russian prefixes VZ- / VOZ- 'UPWARD'

Unique situation in Slavic: the native prefix VZ- and the loan prefix VOZ- have been coexisting in Russian since their formal differentiation emerged in the 14th c.



In Modern Russian VZ- and VOZ- encode different height ('vozvysit'sja' 'tower' vs. 'vsplyt' 'rise to the surface')

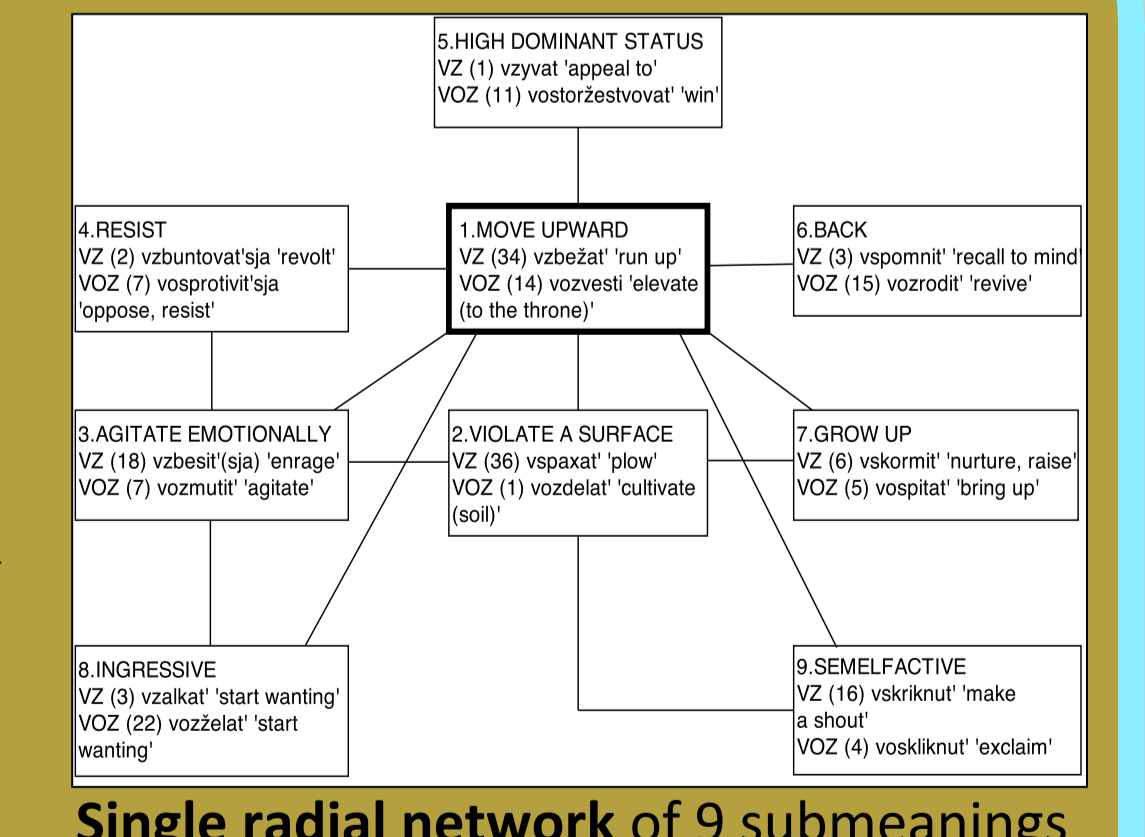


Data: 384 verbs prefixed in VZ- and VOZ-

The high altitude of VOZ- motivates gradual entry into a new state of affairs (Ingressive).

The short trajectory of VZ- justifies abruptness of a rapid momentary event (Semelfactive).

The result of interaction and co-evolution of the native Russian prefix VZ- and a cognate loan prefix VOZ-.



Single radial network of 9 submeanings

