

WORD SEMANTICS IN TERMS OF SUFFIX COMBINABILITY:

L2 ACQUISITION AND A SPECIALIZED ELECTRONIC CORPUS

12TH TEACHING AND LANGUAGE CORPORA CONFERENCE
GIESSEN, 21-23 JULY 2016



**universität
wien**



Stela Manova, University of Vienna
Bartosz Brzoza, Adam Mickiewicz University in Poznań

Structure of the talk

- Preliminaries
 - Approaches to affix order
 - Affixation in theory of grammar & computational linguistics
 - Our cognitive approach
- A psycholinguistic study
- The role of semantics in suffix combinability
- Facilitating vocabulary learning in L2
- Morphological parsing
- Conclusions

What is suffix ordering?



real →

real + **-ize** →

→ *real* + **-iz** + **-ation** →

→ *real* + **-iz** + **-ation** + **-al**

Goals



- To show that the findings of theoretical and psycholinguistic research on suffix ordering can:
 - 1) facilitate foreign language learning and
 - 2) be used for automatic morphological parsing of a lemmatized, POS-tagged corpora.

Approaches to affix ordering

- Overviews in Muysken (1986), Manova & Aronoff (2010), Rice (2011), and Manova (2014).
- According to the type of information used in affix ordering Manova & Aronoff (2010) define eight different approaches:
 - 1) phonological
 - 2) morphological
 - 3) syntactic
 - 4) semantic
 - 5) statistical
 - 6) psycholinguistic
 - 7) cognitive
 - 8) templatic

Affixation in theory of grammar & computational linguistics

- In theoretical linguistics, affixes have an unclear status:
 - affixes pair form and meaning (traditional grammar)
 - affixes do not have semantics, semantics is assigned at the level of word (Realizational morphology (Aronoff 1994, Stump 2001, a.o.), Construction morphology (Booij 2010))
 - affixes provide categorical information, roots are categoriless (Distributed morphology, Halle & Marantz 1993, a.o.)
- Affixation is the attachment of an affix to a morphological base (root, stem):
 - [...[[ROOT+SUFF1]+SUFF2]+...+SUFFIX]
- Computational linguistics: affixation is the attachment of an affix to a stem (ten Hacken & Lüdeling 2002, Trost 2005, a.o.)
- Research on affix ordering usually analyzes combinations of affixes without bases (such as roots or stems)

The combinability of the English suffix *-ist*

SUFF1	Lexical & semantic category of SUFF1	Followed by SUFF2
<i>-ist</i>	N person	<i>-dom, -ic, -y, -ize</i>

Data from Aronoff & Fuhrhop (2002), based on OED, CD 1994

English *-ist*: Our cognitive approach

SUFF1	Lexical & semantic category of SUFF1	SUFF2
<i>-ist</i>	N person	N: <i>-dom</i> (2) ADJ: <i>-ic</i> (631), <i>-y</i> (5) V: <i>-ize</i> (3)

Table from Manova (2011)
Data from Aronoff & Fuhrhop (2002), based on OED, CD 1994

Nouns, adjectives and verbs are seen as cognitive categories, cf. Langacker (1987).

-ist: Fixed combinations

SUFF1	Lexical & semantic category of SUFF1	SUFF2
<i>-ist</i>	N person	N: <i>-dom</i> (2) ADJ: <i>-ic</i> (631), <i>-y</i> (5) V: <i>-ize</i> (3)

Table from Manova (2011)
Data from Aronoff & Fuhrhop (2002)

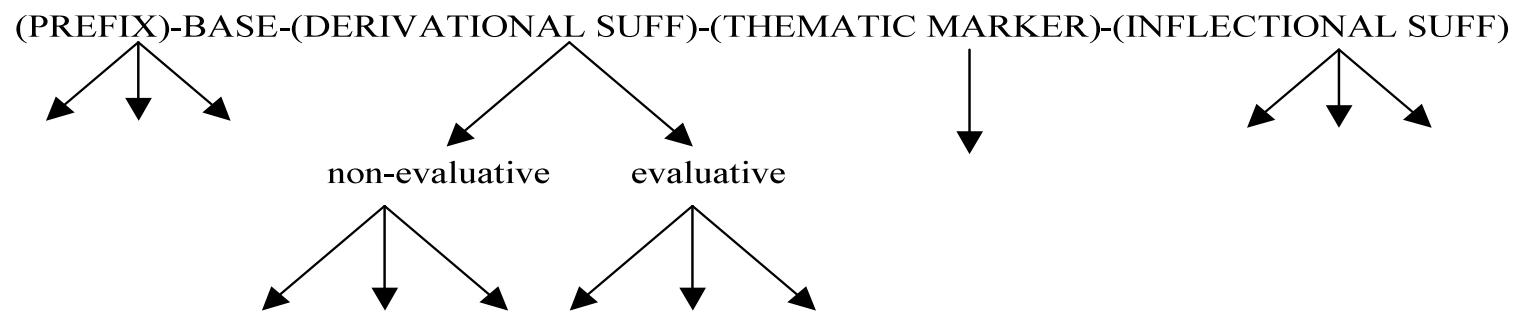
-ist: Predictable combinations

SUFF1	Lexical & semantic category of SUFF1	SUFF2
-ist	N person	N: <i>-dom</i> (2) ADJ: -ic (631), -y (5) V: -ize (3)

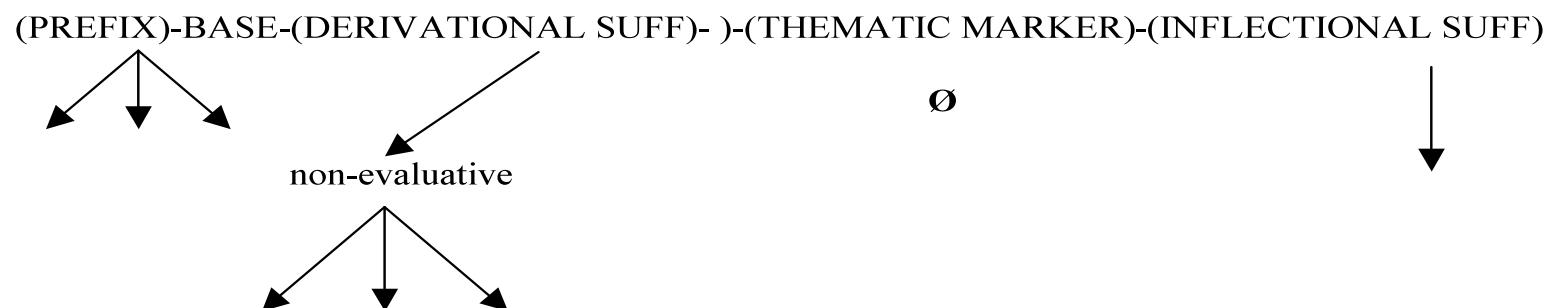
Table from Manova (2011)
Data from Aronoff & Fuhrhop (2002)

Slavic word versus English word

Slavic word



English word



Polish *-ak*: a traditional analysis

SUFF1	Lexical & semantic category of SUFF1	Followed by SUFF2
<i>-ak</i>	N person	<i>-two, -ówka, -ki, -owaty, -ny, -nieć</i>

Polish *-ak*: a cognitive analysis

SUFF1	Lexical & semantic category of SUFF1	SUFF2
<i>-ak</i>	N person	N: <i>-two, -ówka</i> (1) ADJ: <i>-ki, -owaty</i> (3), <i>-ny</i> (3) V: <i>-nieć</i> (1)

-ak: fixed combinations

SUFF1	Lexical & semantic category of SUFF1	SUFF2
-ak	N person	N: -two , -ówka (1) ADJ: -ki , -owaty (3), -ny (3) V: -nieć (1)

-ak: predictable combinations

SUFF1	Lexical & semantic category of SUFF1	SUFF2
-ak	N person	N: -two, -ówka (1) ADJ: -ki, -owaty (3), -ny (3) V: -nieć (1)

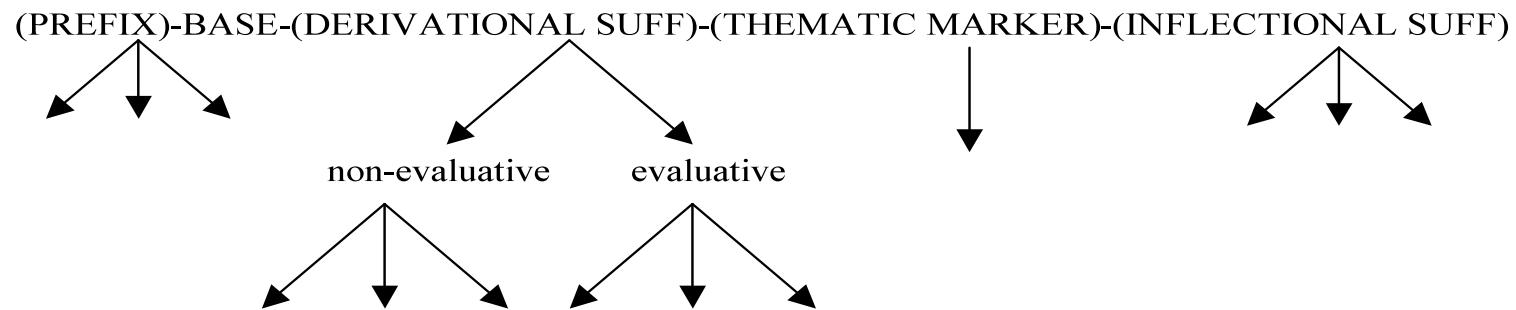
-ak: predictable combinations

SUFF1	Lexical & semantic category of SUFF1	SUFF2
-ak	N person	N: -two , -ówka (1) ADJ: -ki , -owaty (3), -ny (3) V: -nieć (1)

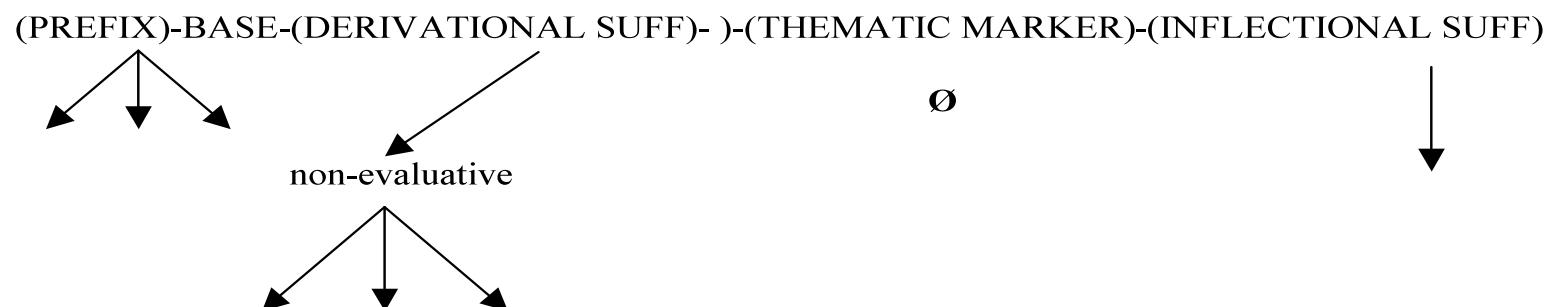


Slavic word versus English word

Slavic word



English word



Main sources of data

- A large dictionary (computer-searchable)
 - ▣ Saloni, Z., W. Gruszczyński, M. Woliński & R.Wołosz. 2007. *Słownik gramatyczny języka polskiego*. Warszawa: Wiedza Powszechna.
- An electronic corpus
 - ▣ Przepiórkowski, Adam, Mirosław Bańko, Rafał L. Górska and Barbara Lewandowska-Tomaszczyk. (2012) *Narodowy korpus języka polskiego*. Warszawa: PWN.

A psycholinguistic study: Hypotheses

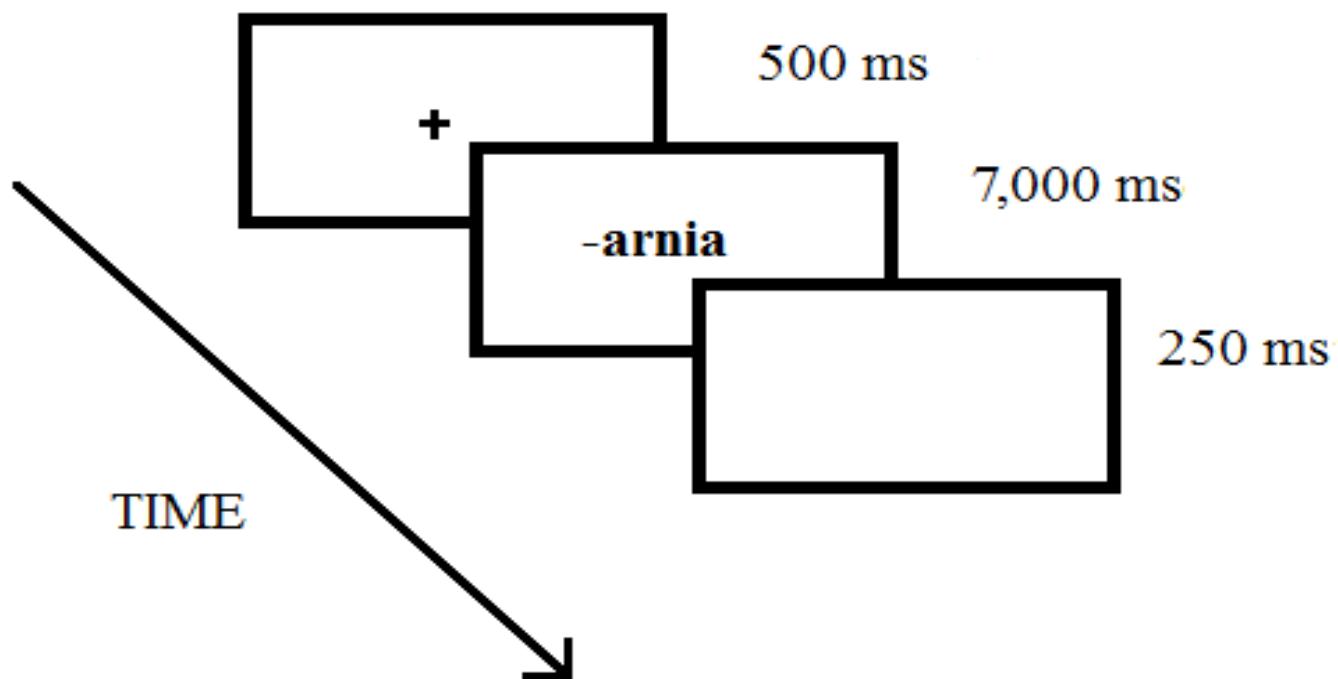


- **H1:** If SUFF1 tends to combine with only one SUFF2 of a major lexical category (N, ADJ, V), SUFF1-SUFF2 combinations are unique pieces of structure and speakers should know them by heart.
- **H2:** If speakers know suffix combinations by heart, existing combinations should be recognised with higher accuracy and faster than non-existing ones.

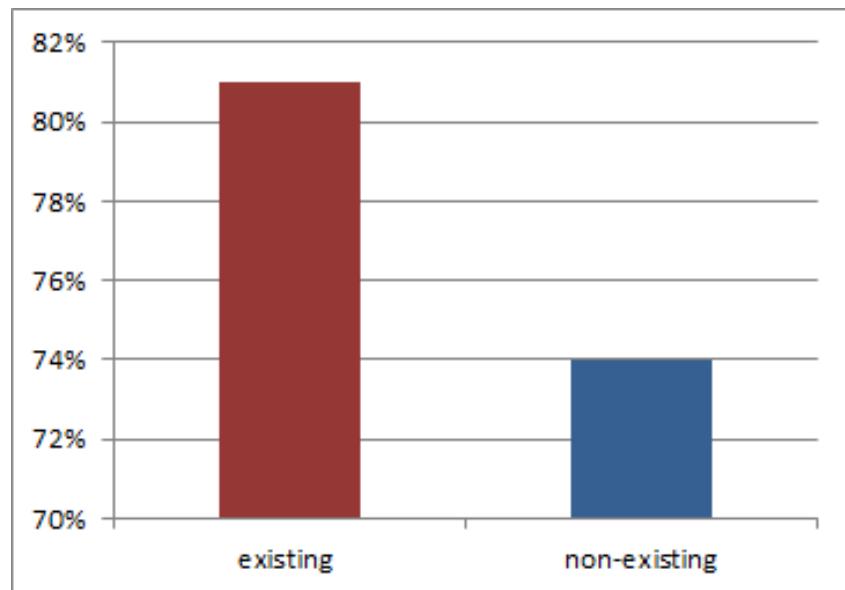
A psycholinguistic experiment

- **Participants:** 53 native speakers of Polish
 - age: M=21.43, SD=1.83
 - no history of developmental dyslexia or reading disabilities
 - non-linguists
- **Task:** Press the right arrow button if a string of letters is an existing combination or the left CTRL button if it is not. In case of a doubt, behave as if a stimulus does not exist.
- **Materials:** 88 items, randomized with the E-prime 2.0 software
 - 44 existing and 44 non-existing suffix combinations
 - The **non-existing combinations** were created by changing the order of suffixes from the existing combinations (e.g. **-c-aw** formed from **-aw-ca**, as in *łaskawca* ‘gracious person’) or by manipulating a single phoneme of the existing combinations (e.g. **-ost-ość** formed from **-ist-ość**, as in *osobistość* ‘personage’).
 - 2 lists
 - each with the suffixes of the other in reverse order
 - each participant saw all combinations

Procedure



Results: Accuracy



Existing combinations:

$$M_{Acc} = 81\%, SD=.09$$

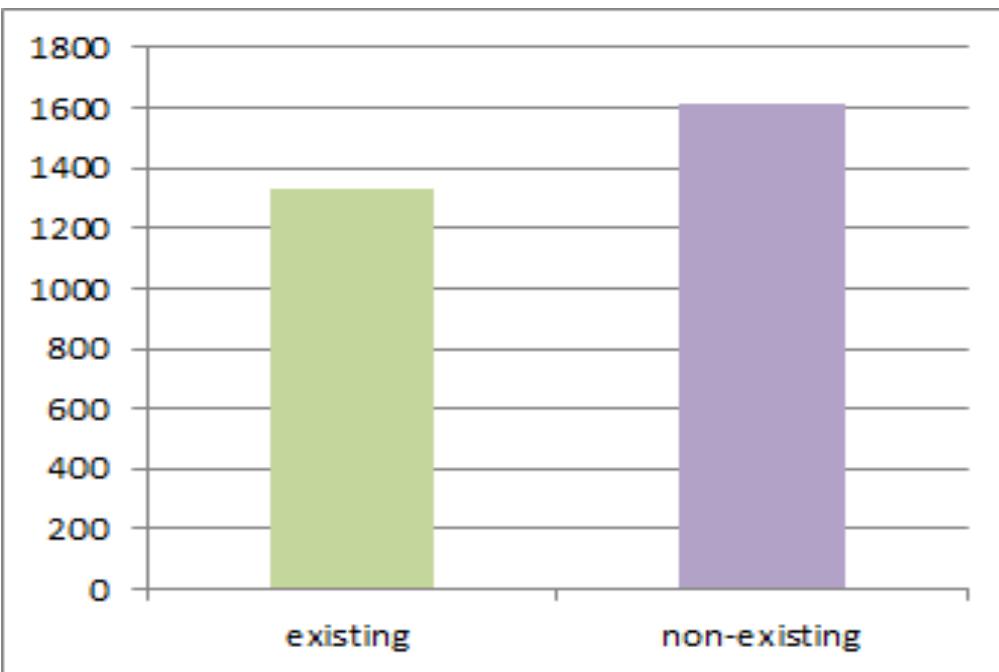
Non-existing combinations:

$$M_{Acc} = 74\%, SD=.12$$

The result is statistically significant:

$$t(52)=3.03, p=0.004$$

Results: RTs



Existing combinations:

1333 ms

$M_{RT}=1333.14, SD=420.57$

Non-existing combinations:

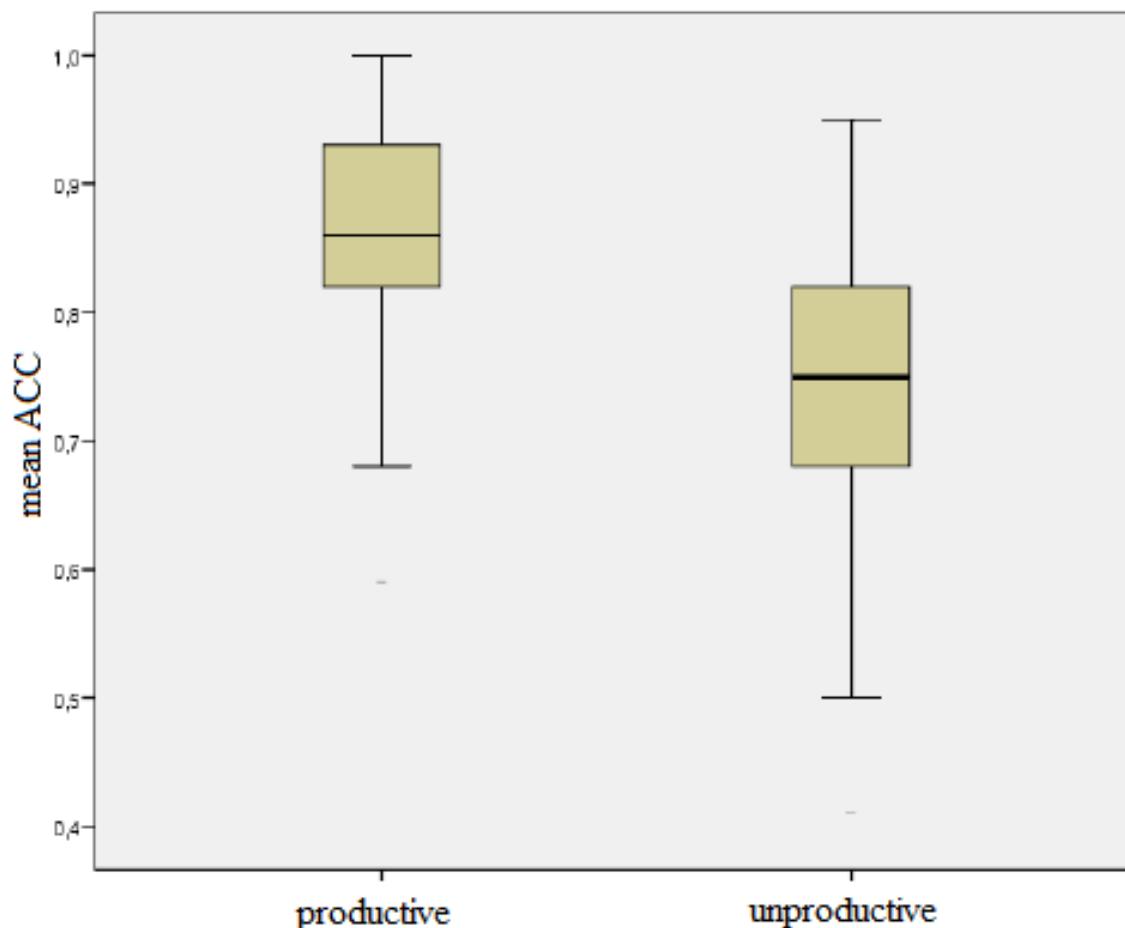
1610 ms

$M_{RT}=1610.38; SD=556.02$

The difference is statistically significant:

$t(51)=-7.53, p<0.001$

Mean accuracy of the productive combinations



Productive combinations:

$M_{Acc} = 86\%$, $SD=.09$

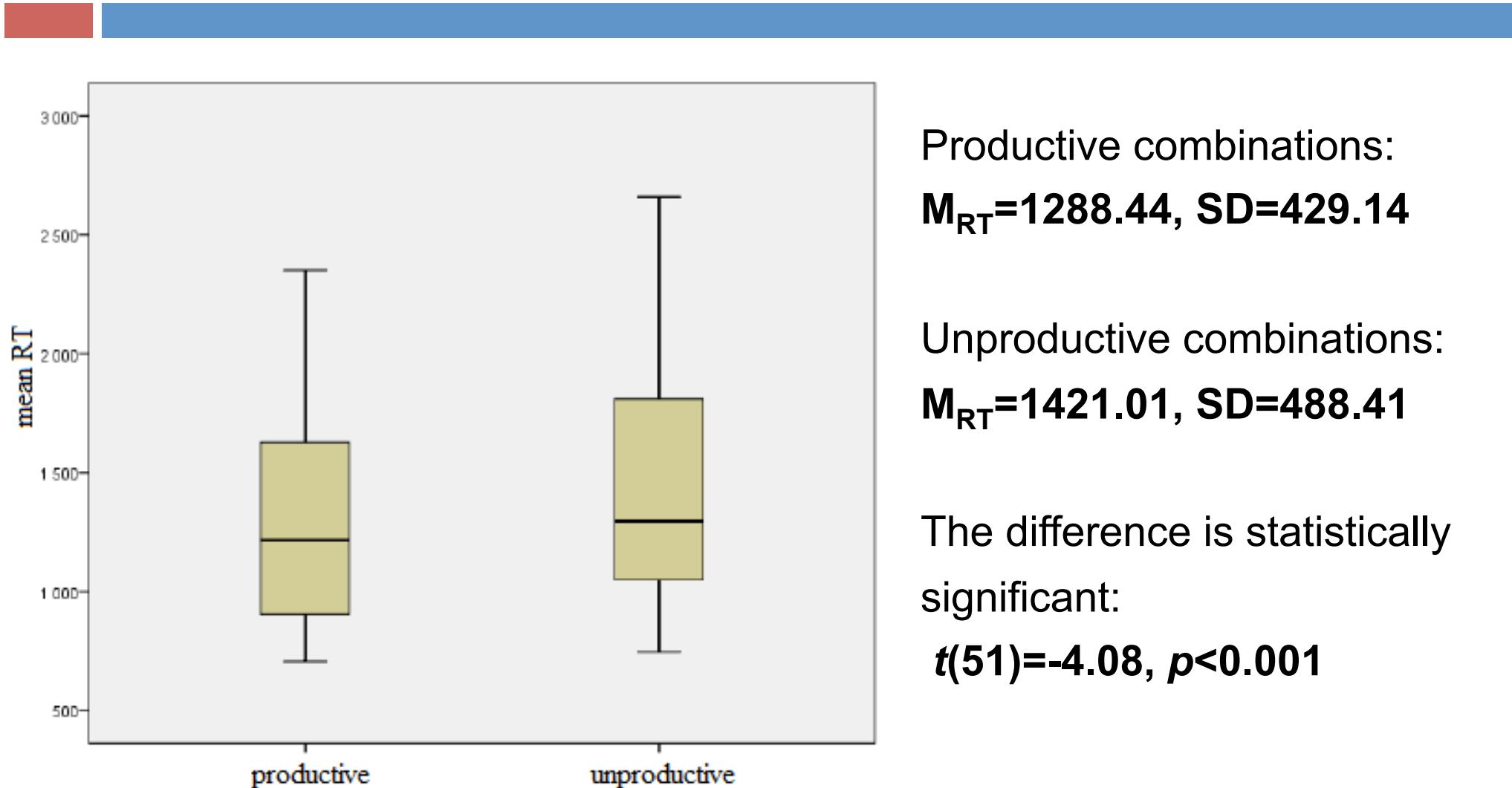
Unproductive combinations:

$M_{Acc} = 75\%$, $SD=.11$

The difference is statistically significant:

$t(51)=7.81$, $p<0.001$

Mean RTs of the productive combinations



Summing up

- The accuracy of recognition of the existing combinations is significantly higher than the accuracy of recognition of the non-existing combinations.
- The reaction times to the existing combinations are significantly shorter than to the non-existing ones.
- Thus, recognition of suffix combinations seems to resemble recognition of words and non-words in psycholinguistics, cf. word superiority effect.
- The productive combinations are recognized more accurately and faster than the unproductive combinations.

SUFF1 for objects 1 (all combinations)

SUFF1	Syntactic category of SUFF1	SUFF1 semantics	SUFF2	Example of SUFF1-SUFF2 combination in a word	Translation
-ina	N	Object	ADJ: -owy N: -arnia (1) (place) N: -arz (1) (person) N: -ówka (1) (object)	wykładz-in-owy okle-ini-arnia okle-ini-arz okle-in-ówka	flooring- veneer workshop veneer producing worker veneer cutter
-nik	N	Object	ADJ: -owy N: -ka (object) N: -arz (person)	grzej-nik-owy zapal-nicz-ka dzien-nik-arz	heater- lighter journalist
-idlo, -ydło -adło	N	Object	ADJ: -any (3) ADJ: -asty (1) ADJ: -isty (1) ADJ: -owy (default) N: -arz (person) N: -arnia (1) (place) N: -nica (1) (object)	krop-idl-any my-dl-asty zwierci-adl-isty wah-adl-owy abec-adl-arz my-dl-arnia my-del-nica	aspergillum- soapy mirror- pendular stupid teacher soap shop soap dish
-nia	N	Object	ADJ: -owy	przekład-ni-owy	gear-
-Vnie	N	Object	ADJ:-ny (4) ADJ: -owy (default)	sklepi-en-ny mieszk-ani-owy	vault- housing-
-ak	N	Object	ADJ: -owy	leż-ak-owy	deckchair-

SUFF1 for objects 2 (semantic rules)

SUFF1	Syntactic category of SUFF1	SUFF1 semantics	SUFF2	Example of SUFF1-SUFF2 combination in a word	Translation
<i>-ina</i>	N	Object	ADJ: -owy	<i>wykładz-in-owy</i>	flooring-
<i>-nik</i>	N	Object	ADJ: -owy N: -ka (object) N: -arz (person)	<i>grzej-nik-owy</i> <i>zapal-nicz-ka</i> <i>dzien-nik-arz</i>	heater- lighter journalist
<i>-idło,</i> <i>-ydło</i>	N	Object	ADJ: -owy N: -arz (person)	<i>wah-adł-owy</i> <i>abec-adl-arz</i>	pendular stupid teacher
<i>-nia</i>	N	Object	ADJ: -owy	<i>przekład-ni-owy</i>	gear-
<i>-Vnie</i>	N	Object	ADJ: -owy	<i>mieszk-ani-owy</i>	housing-
<i>-ak</i>	N	Object	ADJ: -owy	<i>leż-ak-owy</i>	deckchair-

SUFF1 for places 1 (all combinations)

SUFF1	Syntactic category of SUFF1	SUFF1 semantics	SUFF2	Example of combination within a word (and its translation)
<i>-(n)ica₂</i>	N	Place	ADJ: - <i>ki</i> (2) ADJ: - <i>ny</i> (4) ADJ: - <i>owy</i> N: - <i>nik</i> (3)	prądn- <i>ic-ki</i> ‘of generator’ kamien- <i>icz-ny</i> ‘of tenement’ dźwign- <i>ic-owy</i> ‘of crane’ kamien- <i>icz-nik</i> ‘landlord’
<i>-ina</i>	N	Place	ADJ: - <i>owy</i> (6) ADJ: - <i>ny</i> (6) N: - <i>ość</i> (6)	równ- <i>in-owy</i> ‘flatlands-like’ dol- <i>in-ny</i> ‘valley-like’ dol- <i>in-ność</i> ‘the quality of valley’
<i>-nia</i>	N	Place	ADJ: - <i>owy</i> (6) ADJ: - <i>ny</i>	kawiar- <i>ni-owy</i> ‘café like’ pracow- <i>ni-any</i> ‘studio-’
<i>-Vnie₁</i>	N	Place	ADJ: - <i>owy</i> N: - <i>ec</i> N: - <i>ówka</i> (3)	siedz- <i>eni-owy</i> ‘of seat’ mieszk- <i>ani-ec</i> ‘resident’ mieszk- <i>ani-ówka</i> ‘housing industry’

SUFF1 for places 2 (semantic rules)

SUFF1	Syntactic category of SUFF1	SUFF1 semantics	SUFF2	Example of SUFF1-SUFF2 combination within a word (and its translation)
- <i>(n)ica</i> ₂	N	Place	ADJ: -owy ADJ: -ny (4)	dźwign-ic-owy ‘of crane’ kamien-icz-ny ‘of tenement’
- <i>ina</i>	N	Place	ADJ: -owy (6) ADJ: -ny (6)	równ-in-owy ‘flatlands-like’ dol-in-ny ‘valley-like’
- <i>nia</i>	N	Place	ADJ: -owy (6) ADJ: -ny	kawiarni-owy ‘café like’ pracowni-any ‘studio-’
- <i>Vnie</i> ₁	N	Place	ADJ: -owy N: -ec	siedz-eni-owy ‘of seat’ mieszkani-ec ‘resident’

Summing up



- Our research shows that suffix combinations:
 - ▣ are fixed and predictable
 - ▣ are rote-learned
 - ▣ are semantically motivated (a limited number of semantic concepts derives all suffix combinations in a language)
 - ▣ often derive up to 10 types
 - ▣ exist without bases

Suffix combinability and language learning

- We believe that foreign language learners can profit from the native-speaker strategies for word processing.
- Paying attention to the morphological structure of words may facilitate foreign language learning (cf. He and Deng 2015).
- The combinations of the derivational suffixes in a language are a relatively limited number in comparison to the number of words formed with those combinations; therefore suffix combinations should be easier to memorize than lists of words.
- The fact that the combinations of the suffixes in a language are semantically determined can serve as an additional cue when memorizing suffix combinations.
- Productive combinations should be learnt as such; unproductive combinations (such that derive up to 10 lemmas) should be learnt in whole words.

Corpus-based research on word-formation / Morphological annotation of a corpus

A lemmatized and POS-tagged corpus of ≤ 3 M tokens (see Manova & Talamo 2015); lists of derivational and inflectional suffixes; SUFF1 and SUFF2 are any two neighboring derivational suffixes

- 1) Search for SUFF1 (=string of letters), e.g. *ak*
- 2) List all lemmas with SUFF1, ignore lemmas of the type SUFF1+INFL, e.g. a portion of the list for *ak*:
 - *śpiew-ac-two* ‘all singers’
 - *ryb-acz-ówka* ‘fisher’s house’
 - *pływ-ac-ki* ‘swimming’
 - *prost-ak-owaty* ‘boorish’
 - *cud-acz-ny* ‘peculiar’
 - *cud-acz-nieć* ‘become weird’
- ...

Corpus-based research on word-formation / Morphological annotation of a corpus

- 3) Order the results according to: 1) POS-tag; 2) exact form of SUFF1-SUFF2; and 3) type frequency of SUFF1-SUFF2, e.g.:
- POS Verb
 - *cud-acz-nieć* ‘become weird’ (1 lemma)
- POS Noun
 - *śpiew-ac-two* ‘all singers’, *ryb-ac-two* ‘fishing industry’ ... (more than 10 lemmas)
 - *ryb-acz-ówka* ‘fisher’s house’ (1)
- POS Adjective
 - *pływ-ac-ki* ‘swimming’, *ryb-ac-ki* ‘fisher-’ ... (more than 10 lemmas)
 - *prost-ak-owaty* ‘boorish’ (3 lemmas)
 - *cherl-ak-owaty* ‘weakling-’
 - *dziw-ak-owaty* ‘weird-’
 - *cud-acz-ny* ‘peculiar’ (3 lemmas)
 - *mani-acz-ny* ‘compulsive’
 - *dziw-acz-ny* ‘bizarre’

Corpus-based research on word-formation / Morphological annotation of a corpus

- 4) List all fixed combinations (SUFF1 combines with only one SUFF2 of a major category, N, A, V)
 - Establish the fixed combinations based on the POS-tags, e.g.:
cud-acz-nieć ‘become weird’
- 5) All other combinations are predictable
 - Control for type frequency of a particular SUFF1-SUFF2 combination, e.g.:
śpiew-ac-two ‘all singers’
 - ...
 - ryb-acz-ówka* ‘fisher’s house’ (1)
 - ...
- 6) List all fixed and predictable combinations
- 7) Assign semantics to the combinations (either manually or automatically, e.g. based on list of meanings)

Algorithm for a corpus search for derivational suffix combinations (for Polish; SUFF1 & SUFF2 are any two neighboring suffixes)

Search a lemmatized, POS-tagged corpus (≤ 3 M tokens) for lemmas with *SUFF1*

Filter out lemmas with *SUFF1+INFL (based on the list of inflectional suffixes)

Order the lemmas derived by *SUFF1-SUFF2 according to:

- 1) POS-tag, 2) exact form of *SUFF1+SUFF2 and 3) type frequency of *SUFF1+SUFF2

List all lemmas with a fixed suffix combination (SUFF1 is followed by only one SUFF2 of a major lexical category, N, A, V):

*SUFF1+SUFF2_N, *SUFF1+SUFF2_A *SUFF1+SUFF2_V

List all other lemmas, i.e. those with predictable combinations

List all fixed and predictable suffix combinations as two sets of elements

Assign semantics to the combinations (either manually or automatically)

Conclusions



- SUFF1-SUFF2 combinations are independent structures with a status of their own
- SUFF1-SUFF2 combinations are not only listed in the mental lexicon but also semantically-motivated
- Paying attention to suffix combinability can facilitate vocabulary learning in L2
- Regularities related to suffix combinability can be used for morphological parsing of lemmatized, POS-tagged corpora



Thank you!

stela.manova@univie.ac.at

bbrzoza@wa.amu.edu.pl

Selected references

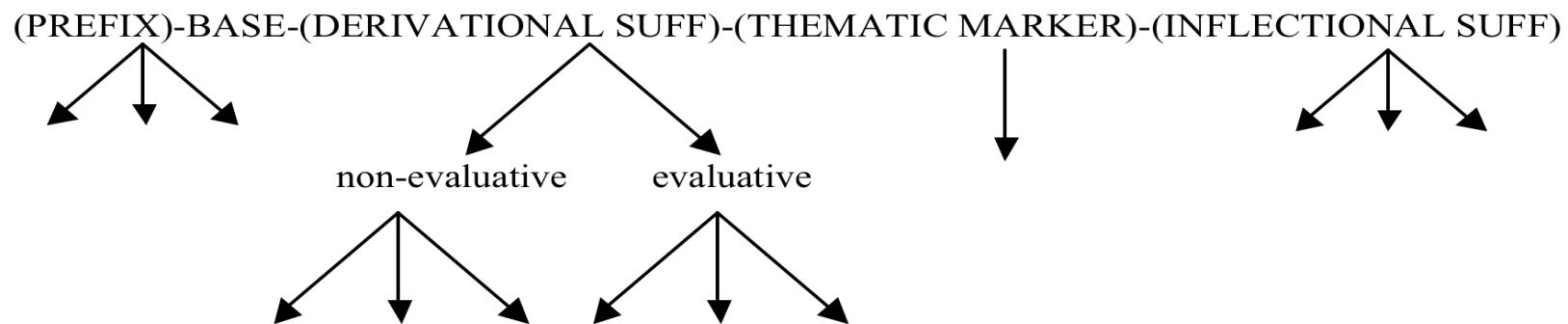
- Aronoff, M. (1994). *Morphology by itself*. Cambridge, Ma: MIT Press.
- Aronoff, M. & N. Fuhrhop (2002). Restricting Suffix Combinations in German and English: Closing Suffixes and the Monosuffix Constraint. *Natural Language & Linguistic Theory*, 20(3), 451-490.
- Baayen, H. R. (2015). Experimental and Psycholinguistic Approaches. In *The Oxford Handbook of Derivational Morphology*, eds. R. Lieber and P. Štekauer, 95-117. Oxford University Press.
- Booij, G. (2010). *Construction morphology*. New York: Oxford University Press.
- Crepaldi D., L. Hemsworth, C. J. Davis & K. Rastle. 2015. Masked suffix priming and morpheme positional constraints. *The Quarterly Journal of Experimental Psychology*. Published online: 11 Mar 2015. DOI:10.1080/17470218.2015.1027713
- Croft, W. (2001). *Radical construction grammar: Syntactic theory in typological perspective*. New York: Oxford University Press.
- Diependale, Kevin, Jonathan Grainger, and Dominiek Sandra. 2012. Derivational morphology and skilled reading: An empirical overview. In *The Cambridge handbook of psycholinguistics*. Edited by Michael J. Spivey, Ken McRae, and Marc F. Joanisse, 311–332. Cambridge, UK: Cambridge Univ. Press.
- Halle, M. & A. Marantz (1993). Distributed morphology and the pieces of inflection. In *The view from building 20*. In Hale K. and S. J. Keyser (eds.), 111-176. Cambridge, MA: MIT Press.
- Hay, J.. 2003. *Causes and Consequences of Word Structure*. London: Routledge.
- Hay, J. & I. Plag. 2004. "What Constrains Possible Suffix Combinations? On the Interaction of Grammatical and Processing Restrictions in Derivational Morphology." *Natural Language and Linguistic Theory* 22: 565–596.

- He, Huaqing and Yunfei Deng. (2015) “The mental lexicon and English vocabulary teaching”. *English Language Teaching* 8(7): 40–45.
- Langacker, R. (1987). *Foundations of Cognitive Grammar, Volume I, Theoretical Prerequisites*. Stanford University Pres
- Lázaro, M, V. Illera & J. Sainz (2015). The suffix priming effect: Further evidence for an early morpho-orthographic segmentation process independent of its semantic content. *The Quarterly Journal of Experimental Psychology*. Published online: 20 May 2015. DOI:10.1080/17470218.2015.1031146
- Lieben, G. (2015). Word-formation in psycholinguistics and neurocognitive research. In *Word-Formation: An International Handbook of the Languages of Europe*, eds. Müller, P. O. et al, 203-217. Berlin: De Gruyter.
- Manova, S. & M. Aronoff (2010). Modeling affix order. *Morphology* 20(1): 109-131.
- Manova, S. (2011). A cognitive approach to SUFF1-SUFF2 combinations: A tribute to Carl Friedrich Gauss. *Word Structure* 4(2): 272–300.
- Manova, S. (2013). Affixation. *Oxford Bibliographies in Linguistics*. New York: Oxford University Press.
- Manova, S. (2015). Affix order and the structure of the Slavic word. In *Affix ordering across languages and frameworks*, ed. S. Manova, 205-230. New York: Oxford University Press.
- Muysken, P. (1986). Approaches to affix order. *Linguistics* 24. 629-643.
- Plag, I. & H. Baayen (2009). Suffix ordering and morphological processing. *Language*, 85(1): 109-152.
- Rice, K. (2000). *Morpheme Order and Semantic Scope*. Cambridge: Cambridge University Press.
- Rice, K. (2011). Principles of affix ordering: an overview. *Word Structure* 4(2): 169-200.
- Saloni, Z., W. Gruszczyński, M. Woliński & R.Wołosz. 2007. *Słownik gramatyczny języka polskiego*. Warszawa: Wiedza Powszechna.
- Schneider, W., A. Eschman & A. Zuccolotto. 2002. *E-Prime reference guide*. Pittsburgh: Psychology Software Tools, Inc.

Word domains

(Manova 2010, 2011b, in press)

Slavic word



Affix ordering and word domains

- Inflection: ABC
pod-pis-yv-a-t'
- Evaluative suffixes: AA
kartina ‘picure’ → DIM1 *kartin-ka* → DIM2 *kartin-oč-ka*
- Non-evaluative derivation: ABAB
lico ‘face’ → *lič-n-yj* ‘personal’ →
→ *lič-n-ost’* ‘person, personality’ →
→ *lič-n-ost-n-yj* ‘related to personality’ →
→ *lič-n-ost-n-ost’* ‘(greater) personality’

Verb INFL



- cud-acz-nieć 1sing. cudacznieję, 2sing.
cudaczniejesz, 3sing. cudacznieje, 1pl.
cudaczniejemy, 2pl. cudaczniejecie, 3pl. cudacznieją

Noun INFL



- śpiew-ac-two: Gen.śpiewactwa, Dat.śpiewactwu,
Loc.śpiewactwie, Acc.Voc.śpiewactwo,
Inst.śpiewactwem

- ryb-ac-two Gen.rybactwa, Dat.rybactwu,
Loc.rybactwie, Acc.Voc.rybactwo, Inst.rybactwem

- ryb-acz-ówka Gen.rybaczówki, Dat. rybaczówce,
Loc.rybaczówce, Acc.rybaczówki, Voc.rybaczówko,
Inst. rybaczówką

Adjective INFL

- płyn-ac-ki Fem. płynacka, Neut. płynackie, Pl. płynackie, płynacci
- ryb-ac-ki ki Fem. rybacka, Neut. rybackie, Pl. rybackie, rybacci
- prost-ak-owaty ki Fem. prostakowata, Neut. prostakowate, Pl. prostakowate, prostakowaci
- cherl-ak-owaty ki Fem. cherlakowata, Neut. cherlakowate, Pl. cherlakowate, cherlakowaci
- dziw-ak-owaty ki Fem. dziwakowata, Neut. dziwakowate, Pl. dziwakowate, dziwakowaci
- cud-acz-ny ki Fem. cudaczna, Neut. cudaczne, Pl. cudaczne, cudaczni
- mani-acz-ny ki Fem. maniaczna, Neut. maniaczne, Pl. maniaczne, maniaczni
- dziw-acz-ny ki Fem. dziwaczna, Neut. dziwaczne, Pl. dziwaczne, dziwaczni

