

Productivity in syntactic change: Polish and Czech impersonals

Roland Meyer

meyerrol@hu-berlin.de

<http://www.slawistik.hu-berlin.de/member/meyerrol/ichl-21.pdf>

Institut für Slawistik, Humboldt-Universität zu Berlin



August 6, 2013

ICHL-21 • Oslo • August 5–9, 2013

Argument

- Apparently, Polish (Pol) and Czech (Cz) **impersonal constructions** only occur with a few verb lemmas early on (XIV-XV), but with a larger number later
- this is an instance of **argument-structural change** and a **change in productivity**
- corpus-linguistic methods for analyzing **synchronic productivity in morphology** can be used to model it
- using of a Large Number of Rare Events (LNRE) model remedies the **differences in sample sizes** over time and avoids the **pitfalls of simpler productivity measures**

Overview

- 1 Preliminaries
 - Argument-structural change and productivity
 - Concepts of productivity
- 2 Polish and Czech reflexive impersonals
 - Synchronic description
 - Diachronic description
- 3 Diachronic productivity study
 - Data preparation
 - Results
- 4 Discussion

Argument-structural change and productivity

argument-structural change – following Roberts (2007, 153ff) – is

- either due to **syntactic reanalysis** – “essentially ‘catastrophic’ and driven by a parametric change”
- or to relatively slow **diffusion** of a pattern through the verbal lexicon

different expectations at productivity:

- reanalysis => sudden jump in productivity, inclusion of verb classes as a whole
- diffusion => slow, long-term increase in productivity

preconditions for an evaluation:

- conceptually: adequate productivity measure, applicable to diachrony
- empirically: enough data points, categorization and (hyper-)lemmata

Realized and expanding productivity

(cf. Baayen 2009, 901)

realized productivity

- number of types *or* type/token ratio
- here: how many different verb lexemes occur in indefinite impersonals in the given subcorpus?
- problem (Baayen 2008, 223): strongly dependent on corpus size, even when defined as type/token ratio (!)

expanding productivity

- number of types occurring only once (*hapax legomena*)
- here: how many verb lexemes in an indefinite impersonal occur only once in the given subcorpus?
- problem (Lüdeling and Evert 2005): Why ignore those occurring twice, three times, ...? – Moreover, single occurrence may well be due to too small subcorpus sizes.

Potential productivity

potential productivity

- number of *hapax legomena* relative to subcorpus size (number of tokens)
- here: what is the expected rate of vocabulary growth for the category of indefinite impersonals, based on the given subcorpus?
- Baayen (2009): *“estimates the growth rate of the vocabulary of the morphological category [...] the extent to which the market for a category is saturated”*

how operationalized?

- a **vocabulary growth curve** (vgc) represents the number of hapax legomena for various (sub)corpus sizes (Evert and Baroni 2007; Baayen 2008)
- vgc modeled through Everts (2004) Zipf-Mandelbrot model (fZM) (<http://zipfr.r-forge.r-project.org/>), doing justice to the logarithmic nature of type/token (**“Large number of rare events” – LNRE**)
- fZM predicts vocabulary growth for various corpus sizes; parameter α represents the gradient of the curve
- Lüdeling and Evert (2005) for diachronic morphology: model synchronic **productivity for several “time slices”** and compare the different α -s

Reflexive impersonals (RIs)

- reflexively marked structures are ambiguous, including an impersonal meaning, cf.

(1) In Italia ieri si è giocato male
 in I. yesterday *si aux-3sg* play-*cvb* badly
 'Yesterday one (*lit.* self) played badly in Italy.' (It., Chierchia 1995, 108)

- Pol and Cz reflexives convey, among others

- decausative:** event without an agent, with transitive (non-refl.) counterpart
- passive:** transitive verb with patient argument realized as subject, agent optionally realized as *by*-phrase
- middle:** modal meaning (possibility, permission), usually with manner adverb – Pol *książka czyta się dobrze* 'the book reads well'
- impersonal:** agent role semantically blocked / fixed to an indefinite human referent – no *by*-phrase possible

(cf. Fehrmann et al. (2010) for a detailed synchronic analysis)

RIs from intransitives

- from intransitive verbs: no argument realized – cf. (**unergatives** (2) and **unaccusatives** (3)-(4)):

- (2) Tańczyło się (Pol) / Tancovalo se až do rána. (Cz)
 dance-*pt.n.sg refl* dance-*pt.n.sg refl* up until morning-*gen*
 ‘There was dancing (until the morning).’ (Kibort 2006, 198 (42) / Růžička 1986, 249)
- (3) Wtedy umierało się na tyfus.
 then die-*pt.n.sg refl* on typhus
 ‘Back then one died from typhus.’ (Pol, Fehrmann et al. 2010, 213 (34))
- (4) Pst, na jevišti se právě umírá (*hrdinou / *od hrdiny)!
 sh on stage *refl* just die-*3sg hero-ins* by hero-*gen*
 ‘Sh, one is just about to die on the stage!’ (Cz, Hudousková 2010, 20 (38-a))

RIs from transitives

- from transitives: *by*-phrase excluded, thus not categorized as passive

(5) Ta/Tę powieść się czyta (dobrze) (*przez dzieci).
 this story-nom/acc.sg.f refl read-pt.n.sg well by children
 'This story is/can be read (well) (*by children).'

(Pol)

(6) Diamantový prsten se našel (*bratrem).
 diamond- ring-nom.sg.m refl find-3sg.m brother-ins.sg
 'The diamond ring was found (*by the brother).'

(Cz, Siewierska 1988, 259 (35))

- anaphora binding** in Pol, but not in Cz

(7) a. Teraz się myśli tylko o sobie.
 now refl think-3sg only about self-prp.sg
 'Nowadays, people only think about themselves.'

(Pol, Rivero and Milojević-Sheppard 2003, 106)

b. *Mluvilo se tam jen o sobě.
 talk-pt.n.sg refl there only about self-prp.sg (Cz, Růžička 1986, 255)

- Pol, but not Cz**, reflexive impersonals contain an **invisible syntactic subject**

RIs with accusative object (Polish)

- reflexive impersonals **with accusative object only in Pol** – clearly impersonal active construction (Siewierska 1988; Kibort 2004)

(8) a. Tu się pije wódkę.
 here *refl* drink-3sg Vodka-acc
 'One drinks Vodka here.'

(Pol, Rothstein 1993, 172)

b. *Tady se pije vodku.

(Cz)

- preferred version in modern Pol – nominative reflexive impersonal favors middle interpretation and is becoming less acceptable with events or actions

Reflexive passive vs. impersonal

- expression of the agent (= *by*-phrase) well attested in Cz (Trávníček 1939; Štícha 1988)

(9) Tohoto Pavla Prvního schvaluje se od mnohých kronykářů k
 this-*gen P.* first-*gen* praise-3*sg refl-acc* by many chronicler-*gen.pl* to
 chudým zvláštní přívětivost
 poor-*dat.pl* particular cordiality
 ‘This Pavel I.’s particular cordiality towards the poor is praised by many chroniclers.’
 (Glatouinus 1565, ČNK)

- but very weakly attested in our Pol corpus (*contra* Szlifersztejnowa 1968, 133)

(10) [Ta sztuka niedługo trwała, – ‘This feat did not last long,’]

bo przez oszukanych prędko się po Warszawie rozgłosiła.
 because by/via deceived quickly *refl* over Warsaw propagate-*pt.f.sg*
 ‘because it was quickly propagated over Warsaw by/via the deceived’

(Kitowicz, 18.Jh., PolDi)

- development from truly passive to impersonal – earlier in Pol, later in Cz

Diachronic description

- few verb lexemes in the beginning, increasing number later on

(11) Czcie się tako we Księgach Sędzkich: [...]
 read-3sg refl so in book-prp.pl judges-

'Thus it can be read in the book of judges: ...'

(Pol, KazŚw, XIV)

- research by "first mention method" (cf. Meyer 2010, 2011) leads to

Overview of development of RIs:

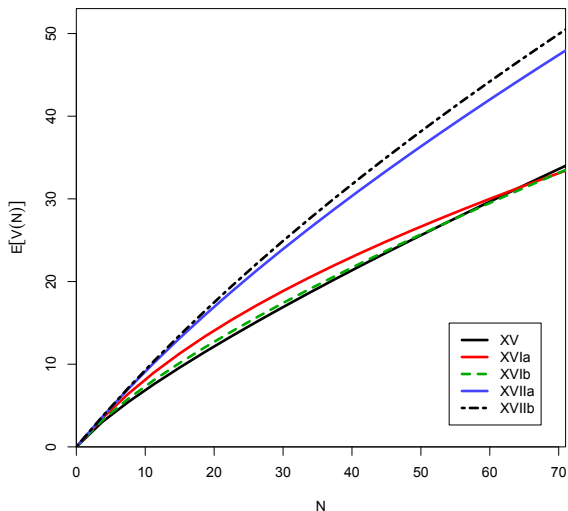
	from transitives	>	from intransitives	>	with accusative object
Cz:	ca. 1400		ca. 1550		–
Pol:	ca. 1500		ca. 1600		ca. 1700

Data preparation and modelling

- **corpora:**
 - Polish Diachronic Research Corpus PolDi – 2.820.759 tokens < 1850
 - diachronic part of the Czech National Corpus (CNC) and corpus of *Vokabulář Webový* – 1.227.925 / 996.209 tokens < 1850
- stratified random samples (per 50y): 7322 (Pol) / 8916 (Cz) reflexives
- categorization as passive, impersonal, \pm accusative object etc. by hand: 641 (Pol) / 805 (Cz) instances of **reflexive passives and impersonals**
- **(hyper-)lemmatized** and annotated for **\pm transitive** base verb
 - variation in phonology (Old Cz. *dieti (sě)* – Cz. *dít (se)* ‘happen’) and orthography, despite standardized transcriptions
 - merger of purely aspectual pairs of verbs (= semantically identical and morphologically derived) – *not* e.g. iteratives in Cz (*psávati* ‘to write’ (*iter.*))
- modelling of **potential productivity per 50y slice** in R using zipfR (<http://zipfr.r-forge.r-project.org/>) – Evert and Baroni (2007)

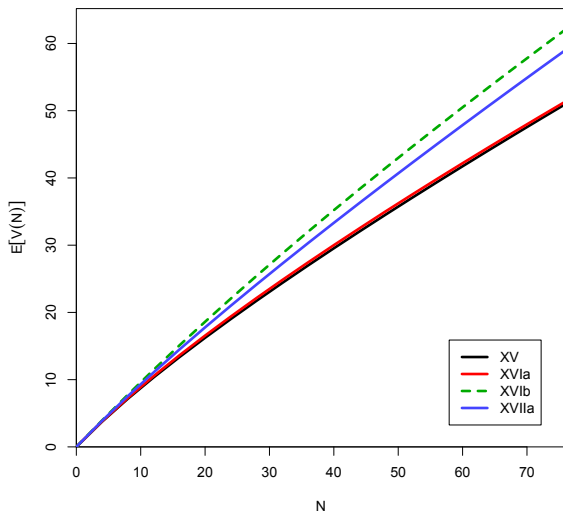
Polish RIs

Productivity leap in Pol between 1550-1600 (XVIb) and 1600-1650 (XVIIa)



Czech RIs

Productivity leap in Cz between 1500-1550 (XVIa) and 1550-1600 (XVIb)



Implications

- at first sight, nice alignment with the table of “first mentions”: productivity leaps ~ extension to intransitives

But: Nothing changes in the above diagrams when the intransitives are excluded – there are just too few instances to have an effect!

- productivity leap => reanalysis?
 - rather **diffusion**: reflexive passives spread through the verbal lexicon and begin to affect also intransitives
 - but don't we *need* a reanalysis here?
 - yes: combining an intransitive with a “defective” *v* would leave the uNom features of T unchecked
 - unaccusative impersonals involve the blocking of a *patient* argument
- reanalysis => productivity leap?
 - not always – cf. appearance of reflexive impersonals with accusative object in Pol: syntactic change with the same transitive verbal lexemes

Outlook

- inclusion of the Pol participial (*-no/-to*) subtype of impersonals
 - however, quite a few verb lexemes attested with this type early on (1400-1450): *mordowano* 'to murder', *biczowano* 'to whip', *swłoczono* 'to undress', *żżono* 'to burn', *rzezano* 'to cut', *wodzone* 'to lead', *morzone* 'to tantalize', *pomszczono* 'to avenge', *dano* 'to give', *dziano* 'to make', *trzymano* 'to hold'
- expansion of the diachronic frame towards XIX and early XX
 - loss of agentive *by*-phrases in Cz
 - dominance of (modal) middle meanings in Pol

References

- Baayen, H. (2008): *Analyzing Linguistic Data*. Cambridge University Press.
- Baayen, H. (2009): Corpus linguistics in morphology: Morphological productivity. In: *Corpus Linguistics. An International Handbook*, eds. Anke Lüdeling and Merja Kytö, Mouton de Gruyter, Berlin, New York, vol. 2, pp. 899–919.
- Chierchia, Gennaro (1995): The variability of impersonal subjects. In: *Quantification in Natural Languages*, eds. E. Bach; E. Jelinek; A. Kratzer; and B. Partee, Kluwer, vol. 1, pp. 107–143.
- Evert, Stefan and Marco Baroni (2007): zipfR: Word frequency distributions in R. In: *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, Posters and Demonstrations Session*, Prague, Czech Republic.
- Fehrmann, D.; U. Junghanns; and D. Lenertová (2010): Two reflexive markers in Slavic. *Russian Linguistics*, 34:203–238.
- Hudousková, A. (2010): *Reflexive Clitics in Czech*. Ph.D. thesis, Univerzita Palackého, Olomouc.
- Kibort, A. (2006): On three different types of subjectlessness and how to derive them in LFG. In: *Proceedings of the LFG06 Conference, University of Konstanz*, eds. M. Butt and T. King, CSLI, Stanford, pp. 289–309.
- Kibort, Anna (2004): *Passive and passive-like constructions in English and Polish*. Ph.D. thesis, University of Cambridge.
- Lüdeling, Anke and Stefan Evert (2005): The emergence of productive non-medical *-itis*: corpus evidence and qualitative analysis. In: *Linguistic Evidence. Empirical, Theoretical, and Computational Perspectives*, eds. Stephan Kepser and Marga Reis, Mouton de Gruyter, Berlin, pp. 351–370.
- Meyer, R. (2010): Reflexive passives and impersonals in North Slavonic languages: a diachronic view. *Russian Linguistics*, 34(3):285–306.
- Meyer, R. (2011): *The History of Null Subjects in North Slavonic. A Corpus-based Diachronic Investigation*. Habilitationsschrift, University of Regensburg.
- Rivero, María Luisa and Milena Milojević-Sheppard (2003): Indefinite Reflexive Clitics in Slavic: Polish and Slovenian. *Natural Language and Linguistic Theory*, 21:89–155.
- Roberts, Ian (2007): *Diachronic Syntax*, vol. 8 of *Oxford Textbooks in Linguistics*. Oxford University Press, Oxford, UK.
- Rothstein, R. A. (1993): Polish. In: *The Slavonic languages*, eds. B. Comrie and G. G. Corbett, Routledge, London, pp. 666–758.
- Růžička, Rudolf (1986): Typologie der Diathese slavischer Sprachen in parametrischen Variationen. *Die Welt der Slaven*, 21:225–274.
- Siewierska, Anna (1988): The Passive in Slavic. In: *Passive and Voice*, ed. M. Shibatani, John Benjamins, Amsterdam, pp. 243–289.
- Šticha, F. (1988): K vývoji vzratného pasíva ve spisovné češtině. *Listy filologické*, 111:22–29.
- Szlifersztejnowa, Salomea (1968): *Bierne czasowniki zaimkowe (reflexiva) w języku polskim*. Wydawnictwo Polskiej Akademii Nauk, Wrocław.
- Trávníček, F. (1939): Pasívum ve spisovné češtině. *Slovo a slovesnost*, 5:13–24.