

Lexicalization and Analogy in Polish o-Raising

Maciej Baranowski and Eugene Buckley
University of Pennsylvania

LSA Annual Meeting, Atlanta
January 2–5, 2003

1. Modern Standard Polish shows an alternation between [o] and [u] in approximately 700 words. The traditional generalization is that /o/ **becomes [u]** before a **voiced word-final consonant**, even if the final consonant is an obstruent that is voiceless on the surface (cf. Bethin 1978, Kenstowicz and Kisseberth 1979, Gussmann 1980, Kenstowicz 1994).

<i>pol-e</i>	<i>pól</i>	[pul]	‘field’
<i>bob-a</i>	<i>bób</i>	[bup]	‘bean’

The raised vowel is **written <ó>**, and is pronounced identically to <u>. Other spellings used here: *a, c, ch, cz, ć, ę, ł, ń, rz, sz, ś, w, y, ź, ż* = /oN, ts, x, ɕ, tɕ, eN, w, ñ, ʒ, ʃ, ɕ, v, i, z, ʒ/.

2. But some vowels followed by voiceless consonants **do raise**, and some native words ending in a **voiced consonant do not** undergo Raising.

<i>stop-a</i>	<i>stóp</i>	‘foot’
<i>tor-y</i>	<i>tor</i>	‘track’

Such irregularities are exactly what we should expect if Raising has become a **property of specific vowel tokens** (or stems) rather than a fact about the grammar as a whole. We argue that in the modern grammar many of these aspects are best treated as **lexicalized** (cf. Sanders 2001). Experimental results from nonce forms support the role of **analogy** with existing words in determining whether a new word will show raising or not.

3. **Borrowings** into the language confirm the idiosyncratic nature of the process. Some do not undergo the alternation, especially masculine nouns borrowed in the citation form without a suffix, i.e. as a surface violation of the Raising generalization (Bethin 1978:49). Some loanwords show the alternation, especially feminine and neuter nouns in which the borrowed citation form has a suffix, and the raised vowel occurs in the derived suffixless genitive plural (Steele 1973, Bethin 1978).

<i>snob-em</i>	<i>snob</i>	‘snob’
<i>mod-a</i>	<i>mód</i>	‘fashion’

4. Based on the data in Tokarski (2002), of more than 4,000 nouns, there is a **very wide range** in the degree to which stems ending in particular voiced consonants show raising.

	b	d	g	dz	dź	w	z	ż	ł	j	l	r
ó	48	142	104	3	8	54	40	15	53	90	6	129
o	20	38	201	0	1	1	130	4	43	14	170	817
	70%	79%	34%	100%	89%	98%	24%	79%	55%	88%	2%	14%

To some degree these differences are due to a preponderance of borrowings ending in that consonant, such as Latin or Greek words in *-or* (*aktor*), *-ora* (*metafora*), *-log* (*katalog*), *-ol* (*metanol*), and *-oza* (*glukoza*). These, like most borrowings, do not undergo raising.

5. Even with **non-native words (largely) excluded**, the rate of raising ranges from 21 to 100 percent. Future work will also have to take into account word **frequency**.

The somewhat chaotic situation reinforces the position that o-Raising has **lost its status as a general phonological process** — at least insofar as the stem-final consonant is referenced — and has become a fact about specific stems (or their vowels).

Some child data

6. Exploratory experiments with **children aged 5.5–9.5**, eliciting unsuffixed forms of actual words presented in a suffixed form containing [o], provide evidence that the traditional generalization regarding the stem-final consonant is not of central importance.

7. There are too few forms to make strong generalizations, but some items stand out. An interesting example is *szopa* ‘shed’, which does not raise for adults but which 40% of the children gave as *szóp*, presumably by **analogy** with common *stopa* ~ *stóp* ‘foot’. Note that the **exceptional status of /p/ as a raising consonant** did not prevent this extension. Also, *osa* ‘wasp’ showed 20% raising, which is not typical of adult speech.

8. With /r/, children generally follow the **idiosyncratic adult patterns**, but in some cases they don’t, e.g. below 50% raising for *bór* ‘woods’, *zapór* ‘dam’ and more than 30% (incorrect) raising for *zmora* ‘nightmare’, *nora* ‘den’. Further work is required to determine the likely analogs or other principles responsible for the deviations, as well as **frequency** of specific forms in the **child’s speech environment**.

9. **Accuracy** is, as one would expect, generally good for words **well-known to most children**. For forms with active raising, words such as ‘foot’ and ‘chin’ have over 80% match with adult forms, while others such as ‘soda’ and ‘virtue’ are around 40%. Again, more data on actual frequency of such words is necessary. Since lack of raising has multiple explanations, such as paradigm leveling, it’s difficult to evaluate the status of words with stable [o].

Adult data

10. We focused on final liquids since they show considerable exceptionality in the native vocabulary. For **monosyllabic words**, similar to non-raising *tor* ‘track’ and raising *dwór* ‘court’, there is suggestive evidence for analogy to these actual words and their behavior. But nonce monosyllables are generally **too short** to know actual analog(s).

11. Similar results for monosyllables resembling *gol* ‘goal’ (masc.) and *sól* ‘salt’ (fem.). Here the **gender** of the noun seems to play a role in establishing an analog, but there is also apparent influence across classes when the **phonological** similarity is very strong.

12. More extensive data were collected for nonce words similar to *gondola* ~ *gondol* and *topola* ~ *topól* ‘poplar’. Overall there was **much more raising** for words resembling *topola*. Both are **feminines** where the genitive plural is the unsuffixed context for raising.

Individuals varied in the degree to which they applied raising to nonce forms (one speaker never did), but for each speaker, raising in *topola* forms was at least as frequent as in *gondola* forms, sometimes much more so.

13. Similar results for two **native** words, **masculines** where the unsuffixed form is the citation. Nonce words resembling *potwór* ‘monster’ raised dramatically more than those resembling *kaczor* ‘drake’, with just two raised tokens.

The liquid examples illustrate the importance of **similarity to existing lexicalizations** in determining the behavior of a new form.

14. We also looked at exceptionality in the other direction, with **voiceless stops**. Here there was **less raising overall** — since only a handful of voiceless stems in the language follow this pattern — but there was still a significant difference between words similar to *zatoka ~ zatok* ‘bay’ (just one raised token, *zatóp*) and those similar to *sobota ~ sobót* ‘Saturday’ (and *robota ~ robót* ‘work’), where a few tokens with **final /k/** were raised (e.g. *sobók*) even though the Polish lexicon has zero examples of raising before /k/.

15. The only exceptionless natural-class generalization about final consonants is that (in the standard dialect) there are **no existing examples of raising before a nasal**. But even this generalization did not prevent some subjects from producing raised vowels in nonce forms ending in a nasal: In an experiment with 6 speakers, **all but one raised at least one nasal-final stem** that was similar to a raising stem ending in /r/, e.g. *potwón*.

These results indicate that the blocking of raising before a final nasal is not an “island of reliability” in the sense of Albright and Hayes (2001). In other words, there is **no motivation to formalize a separate generalization about nasals**; the absence of raising before nasals is an overwhelming fact of the lexical entries, but this “gang effect” in the sense of Skousen (1989) can be trumped by close similarity to a raising form (“proximity”).

A synchronic formalization

16. Following Buckley (2001), we assume that the grammatical representation of the alternation is **categorical**; analogy functions only to assign the new word a particular UR which determines its participation. The specifics are secondary, but a prespecification analysis can handle the facts well:

Vowels that are **stably [u] or [o]** have an explicit value for [high] (or a value that is reliably provided by default); vowels that **alternate between [u] and [o]** are unspecified, and that value is determined by rule or constraint.

a.	stable /u/	[+high]	e.g. <i>gó-r-a – gór</i> ‘mountain’	=	/gur/
b.	stable /o/	[–high]	e.g. <i>nor-a – nor</i> ‘den’	=	/nor/
c.	alternating /O/	[high]	e.g. <i>por-a – pór</i> ‘time’	=	/pOr/

These representations are combined with a **feature-filling process** that prefers [+high] in a word-final syllable, otherwise [–high].

17. For historical reasons, underspecified /O/ tends to be located in **particular contexts**, i.e. before an underlyingly voiced (non-nasal) consonant. But a lexical approach correctly predicts exceptions to this tendency, and that nonce forms (as well as borrowings) will be treated in various ways, partly determined by resemblance to existing words and other factors unrelated to voicing.

The **overall similarity of the word** seems to be what matters, rather than simply what the final consonant may be. It remains for future work to calculate the cumulative analogical effect of the lexicon on nonce forms, but our results support the general role of analogical classification of the new word as “raising” or “non-raising”, rather than a synchronically productive rule referring crucially to the voicing (and nasality) of the final consonant.

References

- Albright, Adam, and Bruce Hayes. 2001. Rules vs. analogy in the English past tense: A computational/experimental study. Ms., UCLA.
- Bethin, Christina. 1978. *Phonological rules in the nominative singular and genitive plural of the Slavic substantive declension*. Dissertation, University of Illinois at Urbana.
- Booij, Geert, and Jerzy Rubach. 1987. Postcyclic versus postlexical rules in Lexical Phonology. *Linguistic Inquiry* 18, 1–44.
- Brooks, Maria Żagorska. 1975. *Polish reference grammar*. Mouton.
- Buckley, Eugene. 2001. Polish o-Raising and phonological explanation. LSA Annual Meeting, Washington, DC, January 4–7.
- Gussmann, Edmund. 1980. *Studies in abstract phonology*. MIT Press.
- Grzegorzczkova, Renata, and Jadwiga Puzynina. 1973. *Indeks a tergo do Słownika języka polskiego pod redakcją Witolda Doroszewskiego*. Państwowe Wydawnictwo Naukowe.
- Kenstowicz, Michael. 1994. *Phonology in generative grammar*. Blackwell.
- _____, and Charles Kisseberth. 1979. *Generative phonology: Description and theory*. Academic Press.
- Sanders, Nathan. 2001. Preserving synchronic parallelism: Diachrony and opacity in Polish. GLiP-3, University of Warsaw.
- Skousen, Royal. 1989. *Analogical modeling of language*. Kluwer, Dordrecht.
- Steele, Richard D. 1973. *The segmental phonology of contemporary standard Polish*. Dissertation, Harvard University.
- Stieber, Zdzisław. 1973. *A historical phonology of the Polish language*. Carl Winter.
- Tokarski, Jan. 2002. *Schematyczny indeks a tergo polskich form wyrazowych*. Edited by Zygmunt Saloni. Warszawa: Wydawnictwo Naukowe PWN.

Dept of Linguistics
619 Williams Hall
University of Pennsylvania
Philadelphia, PA 19104-6305

maciej@ling.upenn.edu
gene@unagi.cis.upenn.edu