

Frequency Distributions as Faithfulness Targets: Or, Why Bulgarians Feminized Turkish Nouns

The substantial literature on loanword adaptation focuses primarily on changes to source language forms driven by mismatches in the phonological systems of a pair of languages. However, recent work illustrates that loan adaptation may also involve changes that appear not to be motivated by this factor directly. Rather, borrowings may involve ‘unnecessary repairs’ to structures which are unobjectionable in the borrowing language, in order to yield a lexical distribution which better matches that of the borrowing language (Walter 2011).

This study documents one such case, involving Bulgarian loanwords borrowed from Turkish. I demonstrate that Turkish forms which would be unproblematically borrowed as neuter-gendered nouns in Bulgarian are instead borrowed with a final vowel change (/e/ to /a/) which yields feminine-gendered nouns. Moreover, such changes occur with precisely the frequency to yield a statistical distribution among noun genders in Bulgarian that remains unchanged despite the large lexical influx.

Therefore, in loan adaptation it appears that adults deploy their knowledge of distributional generalizations over the lexicon (Frisch and Zawaydeh 2001; Hudson Kam and Newport 2005) and are motivated by faithfulness to such generalizations at least as much as by faithfulness to individual phonemes or derivational transparency (Becker et al. 2011). An optimality-theoretic account is given as one means of formalizing this insight.

Gender in Bulgarian

Bulgarian nouns may be either neuter, feminine, or masculine gender. In general, neuter nouns end in /e/, feminine nouns end in /a/, and masculine nouns end in consonants or other vowels. Loanwords normally also follow these gender assignment rules, but a substantial class of exceptions is exemplified below. Here, potentially neuter nouns are borrowed with a vowel change to final /a/, yielding feminine forms.

<u>Bulgarian</u>	<u>IPA</u>	<u>Turkish</u>	<u>Gloss</u>
чешма	tʃeʃm <u>a</u>	çeşm <u>e</u>	fountain
тенджера	tendʒer <u>a</u>	tencere	cooking pan
махала	mahal <u>a</u>	mahalle	neighborhood
механа	mexan <u>a</u>	meyhan <u>e</u>	tavern

The Corpus

Two separate corpora were assembled of Bulgarian words of Turkish origin, ending in either /a/ or /e/ in the original Turkish. The first corpus is assembled from various sources on the Turkish-speaking minority in Bulgaria (n= 63; Georgieff 2012, Sakareva 2005, Kramer 1992). The second corpus consists of forms from Gadjeva’s (2009) study of Turkish loanwords in Bulgarian (n=144).

Results

Proportions of gender assignments as shown by final vowel in the two loanword corpora were compared to general proportions in the lexicon, as instantiated in the Xeba dictionary (2012) of approximately ten thousand head words. The dictionary data reveals that in the overall lexicon of Bulgarian, masculine and feminine nouns are of roughly equal number (each roughly 40%), while neuters are only 20%. Therefore, the number of neuter nouns is about half that of the number of feminine nouns (Row 1).

		Neuter (-e)	Feminine (-a)	neut/fem
1	Dictionary %	19	39	0.49
2	Corpus 1 (Gadjeva): Turkish	64	80	0.8
3	Corpus 1: Bulgarian	48	97	0.49
4	Corpus 2: Turkish	26	36	0.72
5	Corpus 2: Bulgarian	19	41	0.46

Row 2 gives the distribution of /e/-final versus /a/-final forms in the original Turkish, showing that the /e/-final forms are more prevalent. However, this changes when the final vowel of the resulting Bulgarian loanword is taken into consideration, rather than the source form. Row 3 demonstrates that the vowel changes are overwhelmingly in one direction, yielding a proportion of neuter-to-feminine forms that exactly matches the pre-existing ratio in Bulgarian. Rows 4 and 5 illustrate precisely the same pattern for the second loanword corpus.

Conclusions

Speakers appear to violate theoretical assumptions such as lexicon optimization and a general assumed desire for derivational faithfulness and transparency, in favor of maintaining consistent statistical distributions over the lexicon over time. This can be captured in an optimality-theoretic model of gender assignment (Rice 2006) that incorporates gradient constraint ranking (Boersma and Hayes 2001), following Walter (2011).

References

- Becker, Michael, Nihan Ketrez, and Andrew Nevins. 2011. The surfeit of the stimulus: Analytic biases filter lexical statistics in Turkish laryngeal alternations. *Language* 87:1, 84-125.
- Boersma, Paul and Bruce Hayes. 2001. Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32.45-86.
- Frisch, Stefan and Adnan Bushra Zawaydeh. 2001. The psychological reality of OCP-Place in Arabic. *Language* 77. 91-106.
- Gadjeva, Snejana. 2009. Les turcismes dans la langue bulgare contemporaine/Turcisms in the contemporary Bulgarian language. PhD thesis, Paris INALCO.
- Georgieff, Anthony (ed.). 2012. The Turks of Bulgaria: History, traditions, culture. Vagabond Media.
- Hudson Kam, Carla L. and Elissa L. Newport. 2005. Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development*. 1. 151-195.
- Kramer, Christina. 1992. The use of Turkisms in Balkan languages as reflected in Konstantinov's novel Baj Ganjo. *Zeitschrift für Balkanologie* 28. 44-60.
- Manova, Stela and Wolfgang Dressler. 2001. Gender and declensional class in Bulgarian. *Wiener Linguistische Gazette*.
- Rice, Curt. 2006. Optimizing gender. *Lingua* 116.1394-1417.
- Sakareva, Ivanka. 2005. Changes in the Bulgarian language during the centuries: Impact of different cultures on the language in the past. *TRANS. Internet-Zeitschrift für Kulturwissenschaften* 16.
- Walter, Mary Ann. 2011. Probability matching in Arabic and Romance morphology." In *Perspectives on Arabic Linguistics*, Ellen Broselow and Hamid Ouali (Eds.). Amsterdam: John Benjamins.
- Xeba Angliiski a-b, b-a dzhoben rechnik. 2012. (Хеба Английски джобен речник).