

Representing variation in reference grammars

In this paper I discuss the challenges for including variation in reference grammars, illustrated by case-studies from my field data. Dorian (2010) claimed that intra-speaker and inter-speaker variation (with no social weighting) is a lot more common than traditionally described by linguists “socialised in standard-language environments”. The problem is how to handle variation in descriptive grammars. On the one hand, the linguist will not always have enough time to go through the corpora (especially, for previously undocumented languages) to check for inter- and intra-speaker variation at all levels. Moreover, publishing processes have traditionally favoured simplified presentations of the data for the sake of readability (sometimes in an attempt to make it ‘more accessible’ to the community (Riesberg 2018:153)). On the other hand, oversimplification results on many interesting phenomena being overlooked (Rice 2011: 197) and also contribute to perpetrating the “belief that a ‘language’ must exist in some authoritative, invariant form” (Dorian 2010: 287). Most grammar writers live in this dialectic tension of how much should be shared without overwhelming the reader (Payne 2014).

The first case presented here is related to overabundance, suppletion and variation in West Polesian (East Slavic, Belarus). First, in most Slavic varieties the nouns ‘person’ and ‘year’ have a two-stem suppletive paradigm, if there is suppletion at all (e.g. Russian [NOM SG] *god* [GEN PL] *let*; Polish [NOM SG] *rok* [GEN PL] *lat* ‘year’). Those nouns are well-known in the Slavic literature and people expect to find a similar paradigm for other Slavic varieties. However, in West Polesian (WP) there are three-stems for each of these nominal paradigms, which overlap with each other in some cells (i.e. there is overabundance (Thornton 2019)). Moreover, corpus data from naturalistic speech show that speakers use the stems very heterogeneously, even within the same village and household (see *Figure 1 & 2*). Thus, these examples would contribute to the ongoing research on sociologically-neutral-free-variation.

The second case comes from Chamalal (Nakh-Dagestanian), where for some speakers of the same village there are five classes of nouns (e.g. GENDERS), whilst for others there are only four. Nouns from the missing v-class are redistributed among classes III-IV (for those who have a four-class-system), creating an incredible amount of variation in the use of inflectional morphology. This is challenging for glossing, but also for creating dictionaries.

The results are simply too complex to be summarised in a few sentences or a table. But providing too much corpus data could get the series editors to object as it is not what most readers would look for in a grammar. Moreover, this could cause some traditional linguists to dismiss (and even reject) such ‘inelegant and poor’ analyses. Thus, as an early scholar trying to build their career and trying to be a reliable field linguist, I will discuss the middle-ground solutions on which I am working for my descriptions.

In sum, in this paper, based on data from fieldwork, I explore ways of respecting variation whilst keeping the description sufficiently succinct and accessible.

Keywords: corpora; free-variation; overabundance; paradigm representation; suppletion.

Speaker	'ONE'	LOWER NUMERALS	HIGHER NUMERALS	BEFORE THE NUMERAL	QUESTIONS, QUANTIFIERS	WITH ORDINALS
B6	fɔd	'fɔdɪ	fɔd	fɔ'dɪv	fɔd	fɔd
T3	fɔd	'fɔdɑ	lit	'fɔdɑ		ɾɪk
						fɔd
HL4	fɔd	'fɔdɑ	lit	lit		fɔd
Tor1	fɔd	'fɔdɑ	ɾɪk	ro'kɪv	'ro'kɪv	fɔd
			'ro'kɪv			
			fɔd		fɔ'dov	
			fɔ'dov			
TL6		'fɔdɑ	ljet	lit	fɑ'dov	fɔd
		'ro'kɪ	ro'kɪv			
HL3	ɾɪk	fɔ'dɑ	lit	fɑ'dov		fɔd
		'ro'kɪ	fɔ'dov			

Figure 1 Example of heterogeneous suppletion patterns for 'year' in based on corpus hits (WP).

Speaker	'ONE'	LOWER NUMERALS	HIGHER NUMERALS	BEFORE THE NUMERAL	QUESTIONS, QUANTIFIERS	NOM PL
Z10	tʃɔlovɪk		duʃ			
Z4			tʃɔlovɪk			
			duʃ			
TL1	tʃɪlovjek	'duʃɪ		duʃ		du'ʃɪ
T1	tʃɔlovɪk	tʃɔlovɪkɪ	tʃɔlovɪk	tʃɔlovɪk	ludej	ludɪ
		tʃɔlovɪkɑ	duʃ			
HL2		tʃɔlovɪkɑ				ludɪ

Figure 2 Example of heterogeneous suppletion patterns for 'person' based on corpus hits (WP).

References

- Dorian, N.C. (2010). *Investigating Variation. The Effects of Social Organization and Social Setting*. New York/Oxford: Oxford University Press.
- Payne, T. E. (2014). Toward a balanced grammatical description. In T. Nakayama & K. Rice (Eds.) *The art and practice of grammar writing. Special publication of Language Documentation and Conservation*, (pp 91-108). Honolulu: University of Hawai'i Press
- Rice, K. (2011). Documentary Linguistics and Community Relations. *Language Documentation & Conservation*, 5, 187–207.
- Riesberg, S. (2018). Reflections on descriptive and documentary adequacy. In B. McDonnell, A. L. Berez-Kroeker & G. Holton. (Eds.) *Reflections on Language Documentation 20 Years after Himmelmann 1998*. Language Documentation & Conservation Special Publication no. 15. (pp 151-156). Honolulu: University of Hawai'i Press.
- Thornton, A M. (2019). Overabundance: a Canonical Typology. In Rainer, F., Gardani, F., Luschützky, H. & Dressler, W. U. (Eds.) *Competition in inflection and word-formation*, (pp. 189-224). Dordrecht: Springer.