Quantifying potential: Non-canonical word order through a variationist perspective
Mercedes Durham
Cardiff University

The principle of accountability (Labov 1972), whereby all instances of a variable and not just those of interest are analysed, is rightly one of the main tenets of variationist sociolinguistics. In practice, however, the full envelope of variation can difficult to circumscribe, especially for morphosyntactic and discourse-pragmatic features. What is the best procedure to follow in cases when every sentence could potentially contain an overt variant, but generally has an unrealised form? How do we handle situations where separate but partly related features are all potential variants? It is of course possible to focus on the functions of a single variant rather than what could be there, but this is not suitable in cases where the ultimate aim is to compare rates of use across social categories and across different varieties.

This paper aims to suggest ways to resolve these issues by examining five types of non-canonical word order (Birner and Ward 1998).
   (1) Left dislocation: Chester, he comes over several times a year twice
   (2) Right dislocation: cos I remember we used to be able to buy it from Shaws, this hoop.
   (3) Focus Fronting: Early sixties it started, yeah.
   (4) Inversion: all they ever speak, really, is Welsh.
   (5) Clefting: Oh golly, I was in hospital for- I think it was about a couple of weeks, I think, they kept me in.

Non-canonical word order is a good test case as certain variables are perceived to be more frequent in some varieties than others (e.g. right dislocation in the North of England (Durham 2011) and Wales (Penhallurick 2007), fronting in Yiddish English (Prince 1981)), but without a clear method to compare rates across varieties it is difficult to confirm whether these perceptions are accurate, as well as whether findings related to age, sex and other social factors in one variety are unique or shared across varieties.

Previous researchers have dealt with the issues in various ways: some have coded every sentence (but on a restricted data set), some have done their analysis using the numbers of overt tokens per 1000 or 10000 words, yet others have focused on the functions of the variants.

By examining the tokens of all five variables (as well as the unrealised forms) in a half a million word corpus of interviews from Cardiff stratified by age and sex, this paper will compare the various methods and offer suggestions of how best to deal with such types data in language variation and change research, as well as demonstrate which methods are most suited to cross-variety comparisons. It will also discuss why analyses of the interaction between perceived and actual frequency of features can further our understanding of sociolinguistic processes more generally.

References

Birner, Betty J. and Gregory L. Ward. (1998). *Information status and noncanonical word order in English*. Amsterdam: John Benjamins.

Durham, Mercedes (2011). Right dislocation in Northern England: frequency and use perception meets reality. *English World-Wide* 32 (3): 257 279.

Labov, William (1972). *Sociolinguistic Patterns*. University of Pennsylvania Press, Philadelphia.

Penhallurick, Robert. (2007) English in Wales. In David Britain (ed.) *Language in the British Isles.* pp. 152-170. Cambridge: Cambridge University Press.

Prince, Ellen F. (1981). Topicalization, Focus-Movement, and Yiddish-Movement: A pragmatic differentiation. Berkeley Linguistics Society 7:249-64.