Adapting the BNC for sociolinguistic research – a case study on negative concord

Susan Reichelt (ESRC Centre for Corpus Approaches to Social Science, Lancaster University, UK)

Introduction

The following paper details the use of a specialized sub-set of the BNC (BNC64) for sociolinguistic purposes. The corpus, consisting of close to 1.5 million words, is based on the demographic part of the main BNC and offers a balanced sample of speakers across social variables such as gender, age, social class, and regional background. The study is the first stage of a comprehensive re-purposing of the BNC.

Negative concord, the feature used as a case study here, is a non-standard and stigmatized variant that appears in contexts where two negations are used within a single phrase. Prescriptive language norms dictate this form as ungrammatical or illogical (where two negatives form a positive) and give preference to forms with a single negative marker. The paper focuses on categories of gender, age, dialect, and social class and adds to reports of negative concord using the BNC (Anderwald (2002), Muntaña (2008)), highlighting in particular sociolinguistic aspects and its methods used.

The paper is thus aimed at providing some insights into social distributions of negative concord using corpus data, as well as providing detailed considerations in preparing and adapting existing larger-scale corpora for sociolinguistic research.

Data and feature background

In order to obtain representative results from this analysis, it was decided to adapt speaker variables given in the BNC64 to allow for better representation across speaker groups. Thus, dialect areas were broadened from an initial 21 areas to five. Further, ambiguous speaker variables (as found in regional backgrounds and social class), were omitted altogether.

The updated corpus encompasses comparable spoken conversations of 48 speakers. The analysis represents social variation of negative concord across the modified corpus. Social categories that are provided in broad and narrow forms (class and dialect) were initially tested for broad distinctions. Following that, variation that was shown to be statistically significant was further tested in narrow distinctions.

Negative concord (or double negation, NC henceforth) appears in various syntactic contexts and, even though relatively easy to define from a given text, needs to be operationalized thoroughly for larger corpus data. The principle of accountability states that variants within a closed set should include not only the tokens that are of interest for the current study (NC), but also instances where the token could have appeared but did not ("zero" tokens). Zero tokens for NC are structures that are considered as Standard English forms – forms that are considered grammatical and do not carry stigma. In order to give an account that includes both

negative concord as well as the Standard English variant, the token extraction was limited to two rather specific types of NC. Both types are defined through a negated auxiliary (such as *can't, won't, don't,* etc.), a following verb, and either an indefinite noun (such as anything, anyone, nothing, none, etc.) or a negative particle (such as any, owt, no, nowt) in combination with a noun. Examples below show tokens as found in the BNC64 (provided with line number) and the alternative variant (given with *) in the two contexts included (structural delimitation).

(1)   They *can't* suggest *anything*. (270)
       *They *can't* suggest *nothing*.
(2)   *Couldn't* hear *nothing* for splashing. (440)
       **Couldn't* hear *anything* for splashing.
(3)   Because Sainsbury *didn't* have *any* yesterday. (51)
       *Because Sainsbury *didn't* have *none* yesterday.
(4)   I *don't* want *no* cake. (19)
       *I *don't* want *any* cake.

Token extraction

Limiting the search terms for NC (and zero forms) to this specific syntactic pattern which can be described as neg.AUX + VERB + ind.NOUN allowed for relatively straightforward token extraction through the corpus analysis toolkit AntConc.

All tokens were coded for actual NC (and zero form) functions, as well as for syntactic elements. Every token was extracted including speaker information so that meta-data that was provided was easily lined up with the analysis data.

In addition to social factors (age, gender, social class), grammatical factors such as the kind of particles that combined into the construct, were also included in the analysis, though will not be the focus of the current paper.

The following table summarizes all tokens that were included in the main analysis. Approximately 20% of all negations following the pattern neg.AUX + VERB + ind.NOUN appeared as NC. This is higher than comparable studies showed (14.3% in Anderwald 2002:105).

Variable	Total	Standard		Negative concord	
	373	**N**	**%**	**N**	**%**
		296	79.36	77	20.64
Age grouping					
Under 35	128	105	82.03	23	17.97
35 and above	245	191	77.96	54	22.04
Gender					
Female	261	199	76.25	62	23.75
Male	112	97	86.61	15	13.39
Dialect					
North	56	42	75.00	14	25.00
Midlands	118	98	83.05	20	16.95
East	82	63	76.83	19	23.17
London	50	42	84.00	8	16.00

South	67	51	76.12	16	23.88
Class					
Lower	207	151	72.95	56	27.05
Middle	166	145	87.35	21	12.65

Results

For the multivariate analysis the Rbrul package that loads into R was used. Multivariate analysis results show that social class is the most telling social category in NC variation, followed by gender. Neither age nor dialect area seem indicative (in terms of statistically significant variation) of this non-standard feature.

Stat. values from the analysis:
Class (0.000686) + Gender (0.0151) + Age (0.196) + Dialect (0.735)

With social class being an important factor here, a following analysis included narrower descriptions for both classes. The lower class group is made up of C2 and DE, while the middle class consists of AB and C1.[1]
  If the use of negative concord is socially stratified with a higher relative use of the standard variant by the higher social class (and vice versa), the preference of negative concord would increase from highest (AB) to lowest (DE) class. The analysis however shows that the class DE (semi-,unskilled) is higher in their relative use of the standard variant than the next higher social class C2 (skilled manual).

Discussion

Previous studies, looking at individual speaker variables, have found there to be stratification in region (Anderwald 2002). It is not clear whether different results are due to different samples from the BNC or due to the inclusion of other factors in the present study.
  A possible interpretation lies with an awareness of the feature by the lowest social class. As was seen with other non-standard features, lower social class members, or those that are on the cusp between two classes, would oftentimes adjust their linguistic behaviour upwards. While the data provided here only gives a limited amount of information about the speakers and their social backgrounds, this could serve as a possible explanation for the social class distribution.
  In terms of how the unusual gender distribution maps onto social class variation shows that the middle class men do not actually use NC and only at relative low rates in lower class contexts. However, the expected decline of non-standard features with increasing social class can be traced. It is the lowest social class for the female speakers that is surprisingly low (which might be a reflection of the class analysis above where it might be the lowest social class female speakers that are consciously avoiding using non-standard stigmatized features). Taken together female speakers have a much higher rate of NC overall, and male speakers only overtake NC frequencies in the lowest social class.

---

[1] AB – Managerial, administrative, professional; C1 – Junior management, supervisory; professional; C2 – Skilled manual; DE – Semi- or unskilled

Further data is needed to investigate this distribution more thoroughly. In terms of age and dialect groupings the present data did not suggest that the use of NC over standard variants is correlating with differences between younger or older speakers, or where they are located within England (contrary to previous research).

Conclusion

Using larger-scale corpora for sociolinguistic studies offer a wealth of (new) information about language feature use, as well as stratification across speaker backgrounds. Unfortunately, efforts into preparing corpora for sociolinguistic analyses are quite high and many features do not lend themselves to corpus-based studies following variationist methodologies. The case study shows that even with well researched features such as negative concord, there is still room for more diverse approaches and methodologies. This is a first step of adapting existing data in order to create new research, enhancing the field of corpus-based variationist sociolinguistics.

Acknowledgement

References

Anderwald, L. (2002). Negation in Non-Standard British English: Gaps, Regularisations and Asymmetries. London: Routledge.

Labov, W. (1972). Negative attraction and negative concord in English grammar. Language, vol. 48, pp. 773-818.

Nevalainen, T. (2006). Negative concord as an English "vernacular universal". Social history and linguistic typology. Journal of English Linguistics, vol. 34, no. 3, pp. 257-278.

Tubau, S. (2008). Negative concord in English and Romance: Syntax-morphology interface conditions on the expression of negation. Netherlands Graduate School of Linguistics.