VARIATION AND SPEAKER AWARENESS: INVESTIGATING MORPHOSYNTACTIC CHANGE IN TYNESIDE ENGLISH AND WHAT IT MEANS

Marie Jensen, Northumbria University

Outline

- Study 1 – corpus-based study
  - Data, variables, methodology, and results
- Discussion
  - Levelling, cultural re-generation, enregisterment, and identity formation
- Study 2 – interviewer-led questionnaire study
  - Outline of study investigating speaker awareness of forms and the role in language change

Study 1 – Data

- Corpora:
  - TLS (1960s) - 37 speakers
  - PVC (1990s) - 36 speakers
  - NECTE2 (2007-2009) - 48 speakers

Corpus and years collected	Younger speaker birthdates (age 17-34)	Older speaker birthdates (age 35+)
TLS 1965-1970	**1925- 1968**	**1895- 1934**
PVC 1991-1994	**1954- 1977**	**1911- 1953**
NECTE2 2007-2009	**1967- 1990**	**1923- 1966**

Study 1 – Variables

- 7 variables in 3 categories
  - Category 1: sentential negation with *do* (do + NEG)
    - you know and I divn't **suppose** he ever come back I don't **suppose** he ever showed his face in Newcastle again (PVC02b, male, old, MC)
  - Category 2: personal pronouns (1st pers) and (2nd pers)
    - Keeps us on my toes (NECTE2, 07-08/G/DM/456, young, male, MC)
    - it's just yous were good weren't you oh apart from that time yous collapsed (PVC09a, male, young, MC)
  - Category 3: verbal constructions (can + NEG), (go), (throw), (told)
    - it canna be doing you no good (PVC18b, old, female, WC)
    - we often gan on about it now (TLS03, old, female, WC)
    - you have to hoy the boxes though you see (TLS37, old, female, WC)
    - but you telt me it was a fact (PVC06b, young, male, WC)

Study 1 – Variables, category 1

- (do + NEG)
- Variants included: *do, don't, don-t, div, divn't, divn-t, divn, does, doesn't, doesn-t, dinna, divven't.*

	Standard	Tyneside
1st+2nd person singular	**Do + not**	**Div + not**
3rd person singular	**Does + not**	**(no distinct form)**
1st+2nd+3rd person plural	**Do + not**	**Div + not**

Study 1 – Variables, category 2

- (first person pronoun)
- Variants included: *I, we, us, me, my, our, wor, mi.*

	Standard	Tyneside
Subject singular/plural	**I / we**	**I / us**
Object singular/plural	**Me / us**	**Us / we**
Possessive singular/plural	**My / our**	**Me / wor**

- (second person pronoun)
- Variants included: *you, yous, ye, yees, ya.*

	Standard	Tyneside
Subject singular/plural	**You / you**	**(ye) yous / yous**
Object singular/plural	**You / you**	**You / yous (yees)**

Study 1 – Variables, category 3

- (can + NEG)
  - Variants included: *can not, cannot, can't, canna, cannae, can-nae, can-not, can-na, canne, can-ne.*
- (go)
  - Variants included: *go, goes, goin, going, gan, gans, gannin, ganning.*

	Standard	Tyneside
1st person singular	**I go**	**I gan / gans**
2nd person singular	**You go**	**You gan**
3rd person singular	**He / she / it goes**	**He / she / it gans**
1st person plural	**We go**	**We gan / gans**
2nd person plural	**You go**	**(no occurrences)**
3rd person plural	**They go**	**They gan / gans**

Study 1 – Variables, category 3

- (throw)
  - Variants included: *throw, throws, threw, thrown, throwing, throwin, hoy, hoys, hoyed, hoying, hoyin*

	standard	Tyneside
1st person singular	**I throw**	**I hoy**
2nd person singular	**You throw**	**(no occurrences)**
3rd person singular	**He / she / it throws**	**(no occurrences)**
1st person plural	**We throw**	**(no occurrences)**
2nd person plural	**You throw**	**(no occurrences)**
3rd person plural	**They throw**	**They hoy**

- (told)
  - Variants included: *telt, told*.

Methodology

- Tokens were extracted from the corpus using R.
- Coding was done manually in excel
- Statistical analyses carried out in SPSS 19.0
  - Category 1+2: ANOVA
  - Category 3: non-parametric tests (chi-square and Kruskal-Wallis)

Results – category 1

- Initial search: 3,400 tokens. 10 tokens were randomly selected for each speaker. Out of the 120 speakers in the corpus, 17 speakers produced less than 10 tokens. This left 103 informants (NECTE2=43, PVC=30, TLS=30) and a total of 1030 tokens.

	TLS				PVC				NECTE2				Full corpus			
	N	mean	SD	max	N	mean	SD	max	N	mean	SD	max	N	mean	SD	max
(do+NEG)	30	1.27	2.449	10	30	0.73	1.413	5	43	1.02	1.858	6	103	1.01	1.933	10

- The one-way between-groups ANOVA did not show any effect of time of collection (or corpus) on the frequency of use of vernacular forms ($F(2,100) = 0.568$, $p = 0.568$, $X^2 = 4.280$, $df = 2$, *N.S.*, effect size (partial eta squared) = 0.011 (no effect))

Results – category 2, (1st person)

- Initial search: 40,000+ tokens. Half of this was *I* which is the same in both standard and Tyneside English. All tokens of *I* removed from the data set. 20 tokens randomly selected from each speaker. 7 speakers had produced less than 20 tokens. This left 113 speakers (NECTE2=45, PVC=36, TLS=32) and a total of 2260 tokens.

	TLS				PVC				NECTE2				Full corpus			
	N	mean	SD	max	N	mean	SD	max	N	mean	SD	max	N	mean	SD	max
(1st pers)	32	1.53	1.796	7	36	1.92	2.116	6	45	2.56	2.981	13	113	2.06	2.443	13

- The ANOVA test did not reveal any effect of time of collection on use of vernacular forms ($F(2,110) = 1.761$, $p = 1.77$, $X^2 = 20.737$, $df = 2$, *N.S.*, effect size (partial eta squared) = 0.031 (no effect)).

Results – category 2, (2nd person)

- Initial search: 15,000 tokens. 20 tokens randomly selected from each speaker. 2 speakers were represented by less than 20 tokens so left out. This left 118 informants (NECTE2=47, PVC=36, TLS=35) and a total of 2360 tokens.

	TLS				PVC				NECTE2				Full corpus			
	N	mean	SD	max	N	mean	SD	max	N	mean	SD	max	N	mean	SD	max
(2nd pers)	35	0.09	0.284	1	36	0.19	0.467	2	47	1.60	3.221	17	118	0.72	2.164	17

- The one-way between-groups ANOVA indicated an effect of group on use of vernacular forms ($F(2,115) = 7.082$, $p = 0.001$, $X^2 = 60.070$, $df = 2$, $p < 0.001$, effect size (partial eta squared) = 0.110 (small effect size)). Post-hoc tests (Tukey HSD) indicate a significant difference between NECTE2 and PVC ($p = 0.007$) and between NECTE2 and TLS ($p = 0.004$). There was no significant difference between PVC and TLS ($p = 0.973$).

Results – category 3

- Frequencies (chi-square)
  - (can + NEG): 260 tokens from 65 speakers (NECTE2: 24, PVC: 21, TLS: 20).
  - (go): 4,567 tokens from 114 speakers (NECTE2: 42, PVC: 35, TLS: 37).
  - (throw): 86 tokens from 51 speakers (NECTE2: 19, PVC: 20, TLS: 12)
  - (told): 188 tokens from 76 speakers (NECTE2: 29, PVC: 29, TLS: 18).

Results – category 3

Within group frequencies of vernacular forms

- the figures are proportions of vernacular forms out of the total number of tokens collected for that group. I.e. out of the total number of tokens collected for the speakers in the NECTE2 corpus for the variable (can+NEG), 28.9% were vernacular forms. The Ns are the total number of tokens collected for each variable.

	TLS	PVC	NECTE2	total
(can+NEG) N=260	0% (N=0)	1.2% (N=1)	28.9% (N=33)	13.1% (N=34)
(go) N=4567	11.6% (N=84)	4.2% (N=93)	8.2% (N=132)	6.8% (N=309)
(throw) N=86	44.4% (N=8)	18.9% (N=7)	25.8% (N=8)	26.7% (N=23)
(told) N=188	6.7% (N=2)	14.3% (N=13)	7.5% (N=5)	10.6% (N=20)

Within variety frequencies of vernacular forms

- the figures are proportions of vernacular forms in each corpus out of the total number of vernacular tokens collected for that variable. I.e. out of the total number of vernacular tokens collected for (go), 42.7% came from the NECTE2 corpus. The Ns are the total number of vernacular tokens collected for each variable.

	TLS	PVC	NECTE2	total
(can+NEG) N=34	0% (N=0)	2.9% (N=1)	97.1% (N=33)	100% (N=34)
(go) N=309	27.2% (N=84)	30.1% (N=93)	42.7% (N=132)	100% (N=309)
(throw) N=23	34.8% (N=8)	30.4% (N=7)	34.8% (N=8)	100% (N=23)
(told) N=20	10% (N=2)	65% (N=13)	25% (N=5)	100% (N=20)

Results – category 3, (can + NEG)

- A Chi square test of this variable found the difference between groups to be highly significant ($X^2 = 45.032$, $df = 2$, $p < 0.001$) and Cramer's V test for effect size revealed a highly significant medium effect (Cramer's $V = 0.413$, $p < 0.001$). A Kruskal-Wallis test further confirmed that there are significant differences between the distributions of vernacular tokens across the three corpora ($p < 0.001$). NECTE2 speakers contributed 97.1% of all vernacular tokens for this variable.

corpus	TLS				PVC				NECTE2				Full corpus			
variables	N	mean	median	SD	N	mean	median	SD	N	mean	median	SD	N	mean	median	SD
(can + NEG)	20	0.00	0.00	0.000	21	0.01	0.00	0.031	24	0.29	0.00	0.415	65	0.11	0.00	0.287

Results – category 3, (go)

- Chi square testing of the difference in frequency of vernacular forms between the three corpora showed a highly significant difference with a small effect size ($X^2 = 56.618$, $df = 2$, $p < 0.001$, Cramer's $V = 0.111$, $p < 0.001$). There is a steady increase in the number of vernacular forms used by the speakers with the highest proportion being used by the speakers from the NECTE2 corpus. However, the Kruskal-Wallis test returned a non-significant result ($p = 0.288$).

corpus	TLS				PVC				NECTE2				Full corpus			
variables	N	mean	median	SD	N	mean	median	SD	N	mean	median	SD	N	mean	median	SD
(go)	37	0.11	0.00	0.231	35	0.0498	0.00	0.165678	42	0.10	0.00	0.243	114	0.09	0.00	0.218

Results – category 3, (throw)

- The Chi square test found no significant differences between the three groups but the Kruskal-Wallis test did. The result of the Chi square test was non-significant ($X^2 = 4.049$, $df = 2$, $p = 0.132$, Cramer's $V = 0.217$), however, the result of the Kruskal-Wallis test shows a significant difference between the three groups ($p = 0.034$). It would seem that the use of vernacular *hoy* for standard English *throw* is becoming more frequent again after a dip in the 1990s (PVC data).

corpus	TLS				PVC				NECTE2				Full corpus			
variables	N	mean	median	SD	N	mean	median	SD	N	mean	median	SD	N	mean	median	SD
(throw)	17	0.46	0.33	0.484	20	0.21	0.00	0.386	14	0.08	0.00	0.214	51	0.26	0.00	0.409

Results – category 3, (told)

- The final variable is (told) and neither the Chi square test ($X^2 = 2.482$, $df = 2$, $p = 0.289$, Cramer's $V = 0.115$) or the Kruskal-Wallis ($p = 0.691$) found any statistically significant differences between the use of vernacular forms across the three corpora.

corpus	TLS				PVC				NECTE2				Full corpus			
variables	N	mean	median	SD	N	mean	median	SD	N	mean	median	SD	N	mean	median	SD
(told)	18	0.05	0.00	0.138	29	0.11	0.00	0.310	29	0.06	0.00	0.228	76	0.08	0.00	0.246

Results overview

	(do+NEG)	(1st pers pronoun)	(2nd pers pronoun)	(can+NEG)	(go)	(throw)	(told)
Significant ANOVA	NO	NO	YES				
Significant Chi sq				YES	YES	NO	NO
Significant K-W				YES	NO	YES	NO

- Differences between results for for the variables (go) and (throw)
- Results of ANOVA for (do + neg) different from results obtained for this variable in pilot studies (which used the Chi square test).

Discussion

- Levelling
  - Loss of marked local vernacular items, replacement by regional or standard forms. The outcome is convergence of regional dialects toward each other and/or the standard. (Kerswill 2003, Trudgill 1986).
- Watt (2002): Study of phonological levelling in Tyneside English, variables (FACE and GOAT)
- Cheshire, Kerswill, and Williams (2005): Expectation: levelling in one part (phonology) leads to levelling in another part (morphosyntax), finding: no (substantial) connection
- Cultural re-generation: Major social changes in the region in recent years: change from coal mines and ship yards to 'cultural capital of the North' (Miles 2005)
- Urban development and "culture-led regeneration" emphasising middle class values and a lack (or loss) of what (used to?) characterise Newcastle leads to a linguistic move towards less localised varieties (i.e. levelling).
- Revival of vernacular morphosyntactic forms and construction of Tyneside linguistic identity in the face of external changes

Discussion

- Enregisterment and identity formation:
  - **Enregisterment:** the identification of a set of linguistic norms as a linguistic repertoire differentiable within a language as a socially recognized register which has come to index speaker status linked to a specific scheme of cultural values (Agha 2003, Beal 2009).
  - **Commodification:** social and linguistic stereotype of 'Geordie' entrenched in the community.
    - Embrace of 'commodified' dialect by Tyneside speakers (Beal 2009).
    - Enregisterment of stereotypical linguistic features which become meaningful to the vernacular speakers → performance of local identity.
- 'culture-led regeneration' of the urban townscape → 'linguistic regeneration' of local vernacular forms (through processes of enregisterment of commodified forms)

Study 2

- Interviewer-led questionnaire study
  - Aim of study: to investigate the salience of the morphosyntactic variables tested in the corpus study
  - Structure of study: questionnaires, interviews

- Speaker awareness of forms (salience) and the role in language change: does language change lead to forms becoming salient or does salience of forms lead to language change?

References

References

- Agha, A. 2003. The social life of cultural value. *Language and communication* 23:231-273
- Beal, J. 2009. Enregisterment, commodification, and historical context: "Geordie" versus "Sheffieldish". *American Speech* 84:2. 138-156
- Cheshire, J., P Kerswill, and A. Willliams. 2005. 'Phonology, grammar, and discourse in dialect convergence' in P. Auer, F. Hinskens and P. Kerswill (eds.) *Dialect Change*. Cambridge University Press, Cambridge, UK. 135-167
- Kerswill, P. 2003. 'Dialect levelling and geographical diffusion in British English.' in D. Britain and J. Cheshire (eds.) *Social dialectology. In honour of Peter Trudgill*. Benjamins: Amsterdam, The Netherlands. 223-243.
- Miles, S. 2005. 'Our Tyne': Iconic Regeneration and the Revitalisation of Identity in NewcastleGateshead. *Urban Studies* 42: 5/6, 913-926.
- Trudgill, P. 1986. *Dialects in contact*. Blackwell: Oxford, UK

Online Sources and software

- Buchstaller, I. and Corrigan, Karen P. (2007-present) *The Newcastle Electronic Corpus of Tyneside English 2*. <http://research.ncl.ac.uk/necte2/>
- Corrigan, Karen P., Moisl, H.L. and Beal, J.C. (2000-2005) *A Linguistic Time-Capsule: The Newcastle Electronic Corpus of Tyneside English*. <http://research.ncl.ac.uk/necte/>
- R Development Core Team. 2011. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria.