



Development of SMS language from 2000 to 2010: A comparison of two corpora

Úrsula Kirsten Torrado (kirsten@uvigo.es)

*Language Variation and Textual Categorisation Research Group
University of Vigo*

lvtc

ISLE - Boston, 17th June 2011



Outline

- **Introduction**
- **Description of the corpora**
- **Analysis of data**
- **Concluding remarks**
- **Further research**

lvto



Introduction

- “a quick, cheap and easy to use” (Grinter & Eldridge, 2001:219).
- “young and free but tied to the mobile” (Bryden-Brown, 2001: *The Australian Newspaper*)
- Spelling adaptations caused by “the needs for speed, ease of typing and, perhaps, other symbolic concerns” (Thurlow, 2003a).

lvto



Introduction

- Other classifications:
 - Crystal (2009): pictograms and logograms (rebus abbreviation), initialisms, omitted letters (contractions and clippings), nonstandard spellings, shortenings and genuine novelties.
 - López Rúa (2007): abbreviations, clippings, initializations (acronyms, analphabetisms and abbreviations composed of initials), phonetic respellings, letter and number homophones and symbols and onomatopoeic expressions.
- Texting is not the cause of bad spelling and it can improve the user's literacy (Crystal, 2008a; 2009).

lvto



Classification (adapted from Thurlow 2003)

- Reductions: shortenings (sometimes called ‘back clipping’ or ‘apocopation’), contractions (also called ‘middle clipping’ or ‘syncope’), other types of clippings (also known as ‘fore-clipping’ or ‘aphaeresis’), g-clippings, acronyms, initialisms, misspellings and typos and abbreviations.
- Phoneticised respellings: letter/number homophones, non-conventional spellings, accent stylization and stylish talk.
- Word deletion and punctuation: lack of function words, lack of punctuation, over-punctuation and the usage of capital letters.
- Smiley faces and other symbols, also known as ‘emoticons’ (Crystal 2004: 38).

lvto



Description of the OC

- 202 text messages - free online British SMS corpus at Netting-it.com.
- The corpus was taken from a group of about 50 UK college students and their family and friends in 2000. The students were all 17-18 year olds from a sixth form college in Bristol (St Brendan's Sixth Form College).
- Non-tagged corpus

lvto



Description of MOC

QUESTIONNAIRES:

Three questionnaires.

1. SQ1: personal information and some information about their usage of mobile phones.
2. SQ2: transcription of different text messages.
3. TQ: impressions on the impact that mobile phones have on their students' message writing and on other types of writings.

lvto



Devices

SMS LANGUAGE	Tokens		Types	
	OC (2000)	MOC (2010)	OC (2000)	MOC (2010)
Reductions	379 (30.5%)	2085 (33.53%)	216 (48.97%)	433 (48.5%)
Phoneticised respellings	737 (59.3%)	3928 (63.16%)	201 (45.57%)	436 (48.8%)
Smiley faces and other symbols	126 (10.1%)	206 (3.31%)	24 (5.44%)	24 (2.7%)
Total	1242 (100%)	6219 (100%)	441 (100%)	893 (100%)

Table 1: *Devices used in SMS. Comparing OC and MOC (2000 to 2010).*

- Difference in the total amount of tokens
- Purely linguistic analysis
- the most common adaptations when texting are phoneticised respellings; recurrent device
- Relationship between spelling & pronunciation.

lvto

Reductions

REDUCTIONS	Tokens		Types	
	OC (2000)	MOC (2010)	OC (2000)	MOC (2010)
Shortenings / Apocoptation	124 (32.7%)	341 (16.4%)	64 (29.6%)	90 (20.8%)
Contractions / Syncope	89 (23.5%)	1224 (58.7%)	52 (24.1%)	236 (54.5%)
G-clippings	76 (20.1%)	170 (8.1%)	36 (16.7%)	20 (4.6%)
Other types of clippings / Aphaeresis	5 (1.3%)	67 (3.2%)	5 (2.3%)	14 (3.2%)
Abbreviations	24 (6.3%)	31 (1.5%)	11 (5.1%)	10 (2.3%)
Acronyms	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Initialisms	31 (8.2%)	225 (10.8%)	21 (9.7%)	45 (10.4%)
“Misspellings” and typos	30 (7.9%)	27 (1.3%)	27 (12.5%)	18 (4.2%)
Total	379 (100%)	2085 (100%)	216 (100%)	433 (100%)

Table 2: Devices used for reductions. Comparing OC and MOC (2000 to 2010).



lvtc



Phoneticised respellings

PHONETICISED RESPELLINGS	Tokens		Types	
	OC (2000)	MOC (2010)	OC (2000)	MOC (2010)
Letter / number homophones	467 (63%)	2237 (57%)	83 (41.3%)	98 (22.5%)
Non-conventional spellings	144 (20%)	695 (17.7)	54 (26.9%)	117 (26.8%)
Accent stylisation	126 (17%)	908 (23.1%)	64 (31.8%)	163 (37.4%)
Stylish talk	0 (0%)	88 (2.2%)	0 (0%)	58 (13.3%)
Total	737 (100%)	3928 (100%)	201 (100%)	436 (100%)

Table 3: *Devices used for phoneticised respellings. Comparing OC and MOC (2000 to 2010).*

usage of *-ii* instead of *-i* or *-y*, specially in final position; the usage of *-ee*, when a word finishes in *-e*; the addition of *h* to express attitude and accent; the replacement of *o* by *w* in any position of the word; replacement of *-o* by *-oe* for the diphthong /əʊ/; and the usage of *-sz* for plurals.

lvto



Devices

- Instances of slang words and colloquial expressions

yo ('hello'), *hiya* ('hello'), *hollah* ('hello'), *skeen* ('I see'), *peng* ('sexy/hot'), *boomting* ('very good looking'), *buff* ('good looking' only used in London), *crap* ('rubbish'), *mandem* ('friends/person/men'), *niggah* ('friends/black', no racist word), *blad* ('brother', it comes from the Jamaican pronunciation of blood), *copp* ('buy/win'), *safe* ('thanks'), *don't gas* ('don't lie'), *gwarn* ('going on'), *bare cute* ('very cute'), adding the word *man* to the sentences, replacing the 1st singular personal pronoun ('I') by *me*

- Agrammatical expressions:

innit, *he don't*, *ain't*

lvto



Concluding remarks

- SMS language is presumably for reasons of lack of ambiguity, **brevity?** and speed.
- Adaptations based on the sound-to-spelling relationship are the most frequent ones.
- Intrinsic relationship between English pronunciation and spelling
- Shortening is not the aim of respelling in SMS language.
- Emphasizing accent, slang, and attitude seems to be the aim of SMS language.
- Frequent usage of slang and ungrammatical expressions.

lvtc



Further research

- Comparison of UK/US results.
- Comparison with similar students of similar age in similar context in order to determine:
 - if ‘stylish talk’ is used in the different regions of the UK
 - if there are sudden changes (SMS is a rapidly evolving language)
- Reasons for lengthening words.
- Compilation of a tagged SMS corpus.

lvto



References

- Bryden-Brown, Sarah. 2001. Young and free but tied to the mobile, *The Australian*, 10 August.
- Crystal, David. 2004. *A glossary of netspeak and textspeak*, Edinburgh: Edinburgh University Press.
- Crystal, David. 2008. The joy of txt. *Spotlight*, November 2008, 16-21.
- Crystal, David. 2009. *Txting: the gr8 db8*, Oxford: Oxford University Press.
- Grinter, Rebecca E. & Margery A. Eldridge. 2001. y do tngrs luv 2 txt msg? In W. Prinz, M. Jarke, Y. Rogers, K. Schmidt & V. Wulf (eds), *Proceedings of the seventh European Conference on Computer-Supported Cooperative Work* (pp, 219-238), Netherlands: Kluwer Academic Publishers. Available at: <http://www.cc.gatech.edu/~beki/c14.pdf>
- López-Rúa, Paula. 2007. Teaching L2 vocabulary through SMS language: some didactic guidelines. *ELIA* 7: 165-188.
- Online SMS corpus available at: <http://www.netting-it.com/txt8.html>
- Thurlow, Crispin. 2003. Generation Txt? The sociolinguistics of young people's text messaging. *Discourse Analysis Online (DAOL)*. Available at: <http://extra.shu.ac.uk/daol/articles/v1/n1/a3/thurlow2002003.html>



Development of SMS language from 2000 to 2010: A comparison of two corpora

Úrsula Kirsten Torrado (kirsten@uvigo.es)

*Language Variation and Textual Categorisation Research Group
University of Vigo*

lvtc

ISLE - Boston, 17th June 2011