to explore deviations between corpus data and a standard/target: two suggestions". *Corpora* 9 (1): 109-136.

Jaeger, T.F. 2010. "Redundancy and reduction: Speakers manage syntactic information density". *Cognitive Psychology* 61: 23-62.

Torres Cacoullos, R. and Walker, J.A. 2009. "On the persistence of grammar in discourse formulas: a variationist study of *that*". *Linguistics* 47 (1): 1-43.

Wulff, S. and Gries, St.Th. to appear. "Prenominal adjective order preferences in Chinese and German L2 English: a multifactorial corpus study". *Linguistic Approaches to Bilingualism*.

Wulff, S. under review. "A friendly conspiracy of input, L1, and processing demands: *that*-variation in German and Spanish learner language". Submitted to: A. Tyler, L. Ortega and M. Uno (eds.) *The usage-based study of language learning and multilingualism* (Proceedings of GURT 2014). Georgetown: Georgetown University Press.

Recent changes in word formation strategies in American social media

Jack Grieve
Aston University
j.grieve1@aston.ac.uk

Andrea Nini
Aston University
a.nini1@aston.ac.uk

Diansheng Guo
University of South Carolina
guod@mailbox.sc.edu

Alice Kasakoff
University of South Carolina
kasakoff@mailbox.sc.edu

1 Introduction

Current linguistic research is focusing more and more on social media such as Facebook or Twitter. Even though many studies on language variation and change have been carried out using large corpora of social media texts (e.g. Eisenstein et al. 2012; Doyle 2014), not many studies have focused on the analysis of new words emerging from these social media and on their characteristics. This paper begins to fill this gap by presenting results of a study on the emerging new trends of word formation in a large corpus of Twitter messages produced in America.

2 The corpus

The corpus used for this study consists in 6 billion word tokens of geo-coded American tweets collected between January and September 2013 using the Twitter API. The collection of tweets involved only those tweets that were produced within the contiguous United States and that contained both a timestamp and a geocode with the longitude and latitude for the location where the tweet was sent.

3 Methods

As a first step, the 60,000 word types that occurred in the corpus at least 1,000 times were extracted. Through this step it was possible to remove those word types that occurred rarely in the corpus. After this step, the relative frequency of occurrence of the remaining types was calculated for each day represented in the corpus. A Spearman rank-order correlation coefficient was then calculated between the relative frequency of each of these 60,000 word types and the day of the year. By ordering the word types by the value of Spearman rho it was possible to observe which word types increased and which word types decreased in frequency in American tweets during 2013.

4 Results

As an example of the results, the 10 strongest positive and negative Spearman correlation coefficients are presented in Table 1.

Increasing	Decreasing
rn (.978)	*wat* (-.976)
selfie(s) (.965)	*nf* (-.962)
tbh (.960)	*swerve* (-.956)
fdb (.952)	*shrugs* (-.956)
literally (.948)	*dnt* (-.956)
bc (.943)	*wen* (-.948)
ily (.940)	*rite* (-.947)
bae (.934)	*yu* (-.946)
schleep (.932)	*wats* (-.946)
sweg (.932)	*yeahh* (-.945)

Table 1: Top increasing and decreasing words

Among the increasing words, we observe new coinages, such as *selfie(s)* (a photo of oneself), *schleep* (sleep), *sweg* (swag) and *bae* (babe). The word *literally* was also found on the increase and among the several likely explanations for this increase it is possible to propose the rise of a new meaning of *literally* or the increase in the formality of tweets over time. In general, however, acronyms were on the rise in the corpus. Examples of acronyms are *rn* (right now), *tbh* (to be honest), *fdb* (fuck dem bitches) and *ily* (I love you).

Among the decreasing words we observe a number of creative spellings of already established word types, such as *wat* (what), *nf* (now following), *dnt* (don't), *wen* (when), *rite* (right), *yu* (you), and *wats* (what's).

The distributions of these frequencies over time were also explored through scatterplots. As an example of the general patterns observed, the distribution of the top two increasing words, *rn* and *selfies*, and of the top two decreasing words, *wat* and *nf*, are reproduced in, respectively, Figure 1, Figure 2, Figure 3 and Figure 4.

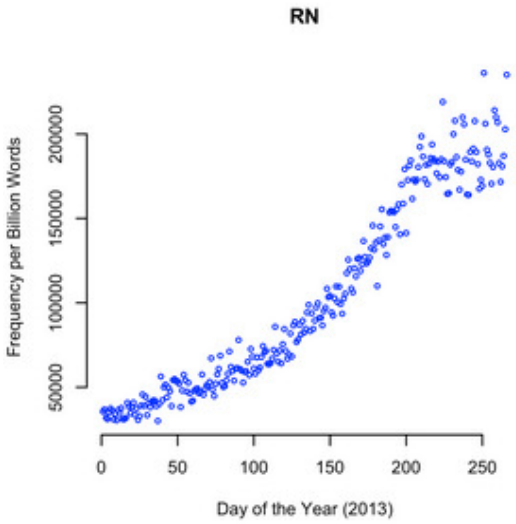

Figure 1: Relative frequency of *rn* over time

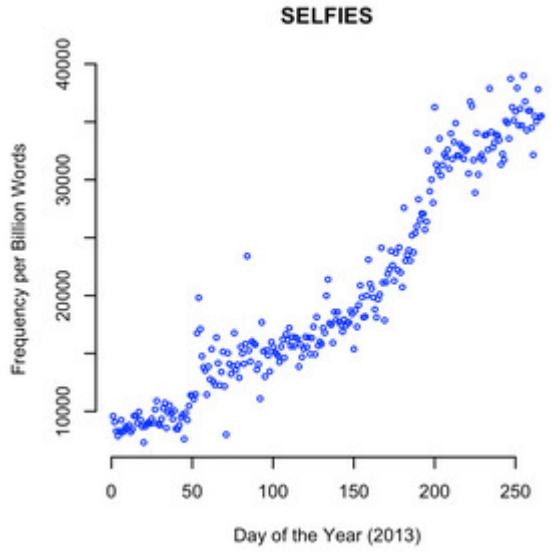

Figure 2: Relative frequency of *selfies* over time

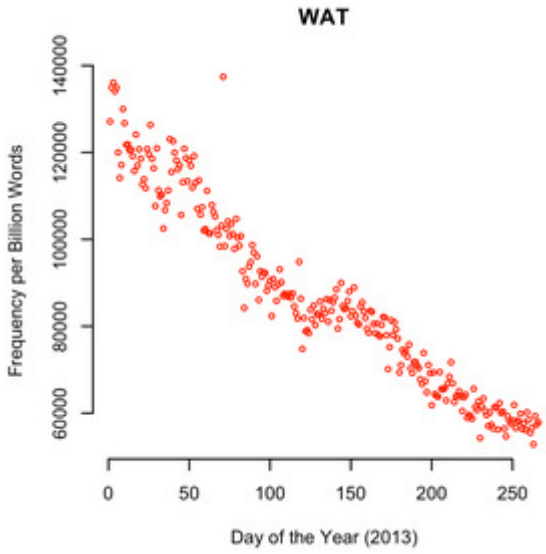

Figure 3: Relative frequency of *wat* over time

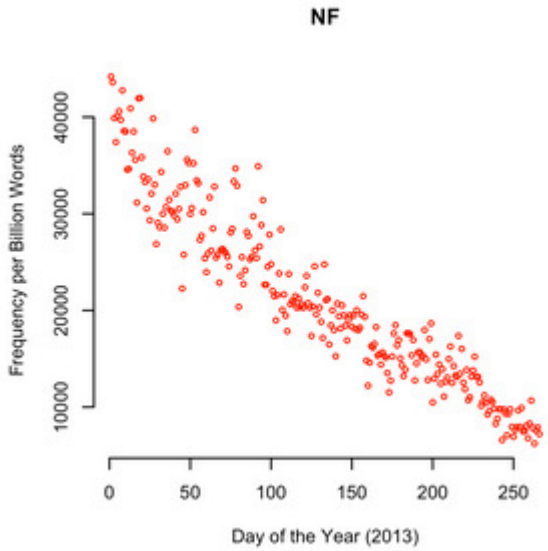

Figure 4 : Relative frequency of *nf* over time

Visual inspection of the examples above and of the other scatterplots suggests that the emerging of new words follows an s-shaped curve of diffusion (Rogers, 2003) whereas the decline of old words follows a steadier and almost linear decrease.

5 Discussion

The results of the study show that the use of creative spellings decreased in American tweets produced during the year 2013, while the use of acronyms increased. The constraints of social media in general are likely to push their users to produce shorter forms and, consequently, to innovate language by introducing new forms that are able to express meaning in few characters. It is clear that both creative spellings (e.g. *wat*, *rite*, *wen*) and acronyms (*rn*, *tbh*, *ily*) achieve this purpose. However, the analysis suggests that acronyms are substantially more popular strategy in contemporary micro-blogging, presumably the 140 character constrain imposed by Twitter has forced the users to contract meaning even more by using acronyms that can reduce common multi-word fixed or semi-fixed phrases in few characters. If this hypothesis is correct, the two trends reflect a change of habit that developed through time and that was determined by the medium.

Apart from the findings on the type of word formation, the examination of the scatterplots suggests that these changes in word formation patterns mirror other types of linguistic change, such as phonological and syntactic changes by following an s-shaped curve typically found in sociolinguistics (Labov, 1995). So far, however, the s-shaped curve has been found in cases of alternation variables in which either the presence or absence of a change is recorded. The present study has produced initial findings that suggest that the same s-shaped curve can be found when frequency variables are considered, with the limit of growth of the curve consisting in the upper limit of that meaning being discussed within the speech community under analysis. A similar s-shaped curve is often found in the diffusion of new symbols or behaviours in various other spheres, such as technology, news, fashion and other aspects of cultural phenomena that represent innovations (Rogers, 2003). The exploration of the similarity between the mechanisms of the diffusion of innovative linguistic items and the diffusion of innovations in other aspects of society is important to be pursued in the future.

6 Conclusions

The present paper reports on an analysis of six billion of tweets for change in patterns of word formation. The results of the study are two-fold. Firstly, it was found that in 2013 American tweets started a shift from a creative spelling word formation trend to an acronym word formation trend. The 140 character medium constrain of Twitter as well as the increasing degree of information used in Twitter has been proposed as main explanation of this phenomenon. Secondly, the words on the increase show an s-shaped pattern that is typical of linguistic changes. If replicated in future studies, these findings can have significant implications for the understanding of the effect of the medium on language evolution and change.

Acknowledgements

This research is funded by the Economic and Social Research Council, the Arts and Humanities Research Counties, and JISC in the United Kingdom and by the Institute of Museum and Library Services in the United States, as part of the *Digging into Data Challenge*.

References

Doyle, G. (2014) Mapping dialectal variation by querying social media, In Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics.

Eisenstein, J., O'Connor, B., Smith, N. and Xing, E. (2012) Mapping the geographical diffusion of new words, arXiv:1210.5268 [cs.CL], pp. 1–13, Available from: http://arxiv.org/abs/1210.5268 (Accessed 13 June 2014).

Labov, W. (1995) Principles of Linguistic Change. Volume I: Internal Factors, Oxford, Blackwell.

Rogers, E. M. (2003) Diffusion of Innovations, New York, Free Press.

Smith, A. and Brenner, J. (2012) Twitter use 2012, Pew

Internet & American Life Project, Available from: http://www.looooker.com/wp-content/uploads/2013/05/PIP_Twitter_Use_2012.pdf (Accessed 21 July 2014).

Transgender identities in the UK mainstream media in a post-Leveson context

Kat Gupta
University of Nottingham
kat@mixosaurus.co.uk

1 Introduction

In this paper I examine the media representation of two trans women, Lucy Meadows and Chelsea Manning. Both women were widely reported in the UK mainstream press in 2013, a period coinciding with Part 1 of the Leveson Inquiry. This Inquiry examined the culture, practices and ethics of the press, with a particular focus on the relationship between the press and the public, the police and politicians.

The Inquiry heard evidence from "organisations representing minority, community and societal groups alleging that individuals within those groups, or the groups themselves, have attracted inaccurate and discriminatory press interest" (Leveson 2012: 448); among these were two written and one oral submission from the organisation Trans Media Watch. In these, they described patterns of negative media representation of trans people, including routine use of previous names, routine use of "before" photos, demeaning and intimidating language for comic effect, and misgendering.

I argue that press misgendering can take more subtle forms than the reporter's use of "quotation marks to dismiss the veracity of the subject's identity inappropriate pronouns or placing the person's identity in" (Trans Media Watch 2011: 11). I examine press usage of pronouns in direct quotations and repetition to investigate how these can be used to undermine trans people's identities.

2 Context

The term "transgender" or "trans" is used as an umbrella term to cover a wide range of gender identities including those of trans women, trans men and people with non-binary, genderfluid and agender identities. Trans people usually experience a sense of misalignment with the sex they were assigned at birth and the gender they identify as; this is in contrast to "cisgender" or "cis" people whose assigned sex and gender identity are aligned. I focus on the experiences of trans women – people who were assigned a male sex at birth, but who identify and/or live as women (Serrano 2007: 11). Serrano (2007: 12) argues that trans women face a complex