Analyzing elderly behavior in social media through language use

Paola Monachesi[1] and Tigris de Leeuw[1]

[1] Utrecht University, Trans 10, 3512JK Utrecht, The Netherlands
P.Monachesi@uu.nl; tigrisdeleeuw@live.nl

Abstract. Increasingly more elderly make use of social media in the Netherlands, being at the forefront compared to other European countries. We analyze the language use of Dutch (older) adults in Twitter in order to get insights in their behavior. Previous studies have identified several linguistic features that change as life progresses including pronouns. We don't find this change and we claim that this might be due to the fact that the users tend to adapt to the language of the platform. This is not the case for the topics addressed since we find differences in the use of hashtags, while the elderly use hashtags mainly in relation to leisure and politics, the younger users employ them mainly in relation to their working life.

Keywords: Elderly, Social Media, language use, behavior.

1 Introduction

Information and language technology can play an important role in assessing the changing behavior of individuals and can function as a means to transform cities into environments that support green and healthy lifestyles. In the context of the European project Grage (Gray and Green in Europe: elderly living in urban areas - www.grageproject.eu), we are carrying out a behavioral analysis of Dutch users based on social media data extracted from Twitter since increasingly more older adults make use of social media in the Netherlands, being at the forefront compared to other European countries.

Our goal is to achieve a better understanding of people's behavior on the basis of language use since previous literature has shown that there are several linguistic features that change as life progresses. In addition, we consider the use of hashtags that are a social media specific feature.

2 Previous Literature

In researching ageing behavior, it is important to define age, which is usually assumed to be the time someone has lived from one's birthdate up until the moment of inquiry

(i.e. chronological age). The studies that consider language use often focus on chronological age [1], [2], [3]. However, individuals of the same age can use language in completely different ways, for example because they take different places in society. Therefore, the focus of various researchers has shifted to grouping subjects on life stages rather than on chronological age [4].

Previous studies have analyzed the linguistic features that change in combination with age. In particular, [1] found that aging is associated with a decline in negative emotion words and at the same time an increase in positive ones. Pronouns are quite revealing of social relationships and identity: subjects younger than 14 and older than 70 use first person plural most often, while individual in between these two age groups used it less. Another significant finding was that the use of self-references (meaning first-person singular pronouns) decreased dramatically with age, with an extreme decline for subjects older than 70.

On the other hand, [2] found that younger people's vocabulary contains a wide variety of slang words, such as swear words or non-derogatory slang. The domain of personal pronouns yielded as result that younger people use the pronouns 'I', 'my', 'myself', 'me', and 'you' significantly more than the older group. The older group showed a higher frequency for third-person singular and plural, as well as first person plural.

While these papers focused mainly on spoken and written resources for their analyses, there is a growing number of studies relying on a computational analysis of social media. They also focus on the use of pronouns as a variable and they suggest a tendency for younger people to use more self-references in the form of first-person pronouns in their language [3]. As people grow older, they tend to use more first person plural pronouns [3], [4].

As for social media specific features such as hashtags, [4] found that hashtags are used more often by older Twitter users: low usage in teens, a steep climb in the 20s, the highest and continuous use through the years up until the oldest participants category (over 60 years of age). Furthermore, [5] noticed that in Instagram the group of teens (13-19 years) posted fewer photos but added more hashtags to them than the adults (25-39 years) do. With respect to content the adult group displays a wider range of interests in topics and are very diverse: arts/photos/design, locations, mood/emotion, nature, social/people while the majority of the teens' hashtags concern mood/emotion and follow/like.

3 Social media use among the elderly

The Netherlands are at the forefront with respect to internet and social network use, this is the case also for the elderly [6]. Social networks use, at European level, in 2013 can be seen in fig. 1, which shows that The Netherlands with 55% of users between 16 and 75 years old belong to the top positions. They are behind the Scandinavian countries (with Denmark scoring 63%) and England (less than 60%) and score above European average in use (43%).
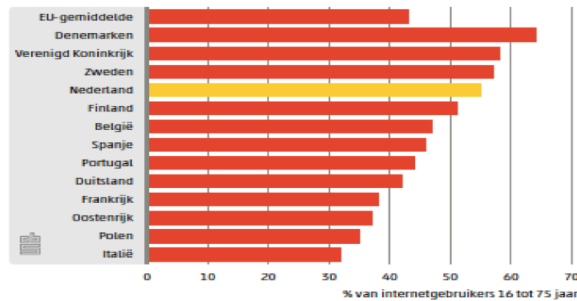
Fig. 1. Social Media Use in EU Countries in 2013

If we consider the social media use (this includes online chatting, writing or reading weblogs, e-mailing and use of social and professional platforms) of the Dutch population in the last years (cf. fig. 2), we notice a clear increase in all age groups. Almost 60% of those between 65 and 75 are active on social media while those above 75 years old are 20%
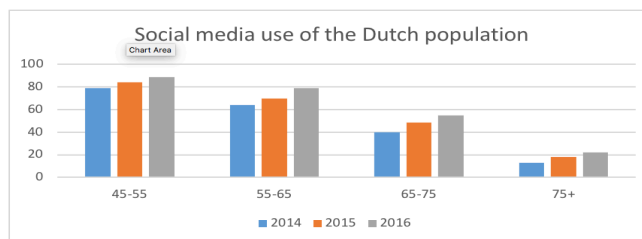
Fig. 2. Social Media Use 2014-2016

The development in the use of the various social networks for different ages in the years 2012 and 2016 is illustrated in fig. 3. where the two years are represented over each other and they partially overlap. It is possible to see that younger people use platforms such as Facebook and Twitter less, while the older population is using social networks more, this is especially the case for instant messaging. Furthermore, the use of professional social networks has increased in the age group 22-35 years old.

Fig. 3. Development of social network use

4   Language analysis

Since social media are being used by an increasingly higher number of people, especially in The Netherlands, it seems relevant to employ social network data to analyze their behavior through language use. Users on a social media platform write spontaneously about their interests, providing information about their needs and their behavior.

In our work, we make use of chronological age of users, but we group them in three classes that reflect three life stages that are related to the active working life of the individuals (i.e. below 55), the pre-retirement stage (between 55 and 67) and post retirement (above 67). In this way, we overcome possible criticism in the use of chronological age ([4], [7]). The users we selected from Twitter had to meet several requirements: their tweets should be publicly accessible and written in Dutch; they should have tweeted at least 400 tweets 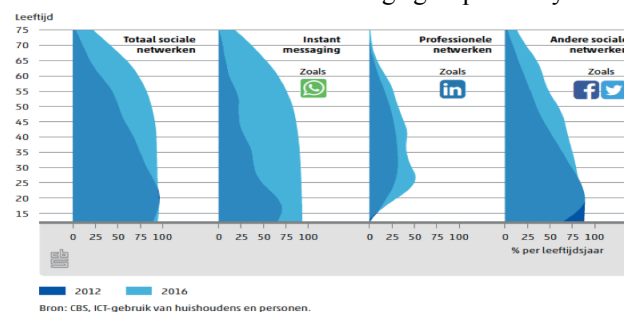and they should show social activity, that is users with a number of followers higher than 300. The classification of users in age groups was carried out manually on the basis of their profile information and the picture posted, the same with respect to gender. Two people have verified the selection in order to retain only those users for which there was agreement in the classification.

We have carried out a preliminary analysis of the most frequent words in the tweets of the selected users and we have focused initially on two age groups. We have carried out a pilot based on 18 users below 55 and 10 users above 67 in order to increase the chances of identifying different behaviour in language use.

Fig. 4. Frequent words of under-55   **Fig. 5**. Frequent words of above-67

The word clouds reveal that Dutch articles '*de*' ('the'), '*het*' ('the') and '*een*' ('a') are very common in the text in both age groups, as well as the use of prepositions: '*in*' ('in'), '*van*' ('of', 'from'), '*voor*' ('before', 'in front of'), '*op*' ('on') and to a lesser degree '*over*' ('over') and '*bij*' ('by') are all used frequently. However, to assess behaviour, the use of pronouns is the most interesting to analyse and we notice that in the case of the group under-55 the pronoun '*je*' ('you' sing.) is more frequent than '*ik*' ('I'). In the above-67 group, the presence of the pronoun '*ik*' ('I') is attested as well as that of the first-person plural '*wij*' ('we'). To assess whether one group uses a word more frequently than the other, a statistical metric suggested by [8] has been used:

$P_{ov67}(w)$ = relative frequency of word w in tweets of subjects over 67
$P_{un55}(w)$ = relative frequency of word w in tweets of subjects under 55

$$\text{variance} = \sqrt{(Pov67(w)/Nov67 + Pun55(w)/Nun55)}$$

$$t = (P_{ov67}(w) - P_{un55}(w))/\text{variance} \tag{1}$$

If the score t is positive, the word is used more often in the tweets written by people older than 67, if negative, the word is more frequent amongst the under 55 group.

Our analysis, revealed that the overall pronoun use increased with age and the pronoun category and the highest frequency in both groups was first-person singular. This usage might imply that the reason for people to use social media is to convey one's own interests and experiences, even as people become older. In addition, as can be seen in fig. 6, the two age groups make a very similar use of the pronouns contrary to what assumed in previous literature.

Pronoun	Translation English	> 67		< 55		T-score
		Times used	P(w)	Times used	P(w)	
Ik	I	663	0,01271	1204	0,01168	0,00958
Je	You,your (sing.)	791	0,01517	1199	0,01163	0,03281
Jij	You (sing.)	68	0,00130	141	0,00137	-0,00172
We	We	110	0,00211	327	0,00317	-0,01887
Wij	We	214	0,00410	64	0,00062	0,13978
Jullie	You,your (plur.)	31	0,00059	118	0,00114	-0,01626

Fig. 6. Pronoun use in the two age groups

5 Hashtags Analysis

Our preliminary analysis of language use has showed that the two extreme age groups share a very similar behaviour in this respect, more specifically in the case of pronoun use. It is thus relevant to assess whether this is the case also with respect to the topics being addressed and an analysis of hashtags use might be revealing. They are a social media specific feature and a relevant source of information given that they are used to index keywords or topics in Twitter, they are indicative of the debate that is being carried out in the platform. We have conducted a preliminary analysis based on the two age groups, that allows for a manual investigation given the limited number of users. Out of the 103.097 words of the tweets belonging to the under-55 group, 5318 were hashtags, that is 5,2%. For the over-67 years old users, it was about 1% (i.e 532 hashtags out of a total of 52.150 words). One possible explanation for this higher usage of hashtags in the younger group could be that Dutch citizens over-67 are usually retired,

whereas those in the younger group still have a working life. Many of their hashtags have some relation to professions or the labour market and they might use hashtags to target specific audiences while this might not be the goal of the older group. We have selected the top 100 most frequent hashtags in each group which we have divided into categories, loosely based on the categories illustrated by [5], as can be seen in fig. 7.

There are clear differences in the topics being addressed by the two groups. The most popular ones amongst elderly users are politics, locations, entertainment and news. Together, these categories comprise 75% of the hashtags. The under-55 group focuses more on work- and world-related tweets, with news, occupational terms, entertainment and companies as the most popular topics. The most noticeable difference between the two groups in terms of topics is that in the over-67 group, none of the topics of the hashtags relate to occupations or companies, and the references to congresses or fairs were much lower than for the under-55 group. Users under 55 write more tweets related to their profession while this is obviously not the case any more for the old adults.

Differences are even more obvious when we look at the theme of *sustainability* where hashtags related to it are only present amongst the younger age group where users belonging to it tweet about this subject in a work-related context. Similar difference can be noticed with respect to the hashtags related to nature that are used by the younger group more than by the group of the older adults. The over-67 group shows an extensive use of location-tags, more than three times as much as the younger group, and entertainment. One might conclude that for those over-67, Twitter is a means to communicate about leisure activities and about their interests where nature and sustainability do not occupy an important place in their communication (cf. [9] for further details).

Table 1. Distribution of the 100 most frequent hashtags into content categories.

Category	Examples in English	% over-67	% under-55
Arts/photos/design	Photography	4.7	1.2
Companies	[names of companies]	0	11.6
Economics	Financing	0.9	3.1
Entertainment	top2000	18.2	12.3
Events/conferences	congresIC	0.6	4.1
Sustainability	40dayssustainable	0	1.3
Locations	Iceland	20.2	6.6
Mood/emotions	Proud	3.2	1
Nature	Bees	0.6	4.3
News	Brexit, [names of Dutch newspapers]	9.4	29.8
Occupationalterms	website, qualitofhealthcare	0	14
Politics	[Names of political parties and laws]	30.8	0.8
Research/univers	RadboudUMC	0.3	0.8
Social/people	AbeltAsman	2.6	0
Sports	NedMex, OS2012	2.9	2.3
Twitter-tags	dtv (*durftevragen*, daretoask)	0.3	4
Other	Wastepaper	5.3	1.2

Fig. 7. Hashtag categories

If we extend this preliminary analysis to the three age groups and additional users (i.e. 100 users), this behavior is confirmed, as can be seen in the word-clouds in fig. 8 and fig. 9.

Fig. 8. Hashtag use for the groups under-55 (left) and between 55-67 (right)

We notice that the two groups below-55 and in between 55 and 67 exhibit a similar behavior with respect to Twitter use in that their discussion evolves around work related and technology related themes, this is even more the case for the 55+. The hashtag use is different for the group above-67 that is especially interested in politics, which is to be expected since this is a group that was more confronted with politics in their youth, as can be seen in fig. 9. Leisure activities and television programs are also mentioned.

Fig. 9. Hashtag use for the group above-67

6 Conclusions

Previous literature has identified several linguistic features that change as life progresses: the use of pronouns being one important discriminator. We do not find this difference among the two age groups analyzed and we claim that this might be due to the fact that since more elderly are active in social media, they might tend to uniform to the language that is used in the platform regardless of age [10]. In a subsequent analysis, it would be relevant to assess when users joined the platform to evaluate whether membership age plays indeed a role.

We have also considered the use of hashtags and we noticed that here are differences in the topics addressed by the groups analyzed, while the elderly use hashtags mainly in relation to leisure and politics, the younger ones use them mainly in relation to their working life. Another important difference is that while the younger group uses hashtags related to nature and sustainability, this is not the case for the elderly who, however, use location tags indicating an interest for the places they live in, but maybe less for the environment. We claim that social media could play an important role in changing this attitude and behavior. We have observed that these differences are attested also when the analysis is extended to additional users and to the three age groups.

Acknowledgements

References

1. Pennebaker, J. W., & Stone, L. D.: Words of wisdom: language use over the life span. Journal of personality and social psychology, 85(2), 291-301 (2003).
2. Barbieri, F.: Patterns of age-based linguistic variation in American English. Journal of sociolinguistics, 12(1), 58-88 (2008).
3. Rosenthal, S., McKeown, K.: Age prediction in blogs: a study of style, content, and online behavior in pre and post-social media generations. In: ACL 2011 (2011).
4. Nguyen, D. P., Gravel, R., Trieschnigg, R. B., Meder, T.: How old do you think I am? A study of language and age in Twitter. In: Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media, pp. 439-448. (2013).
5. Jang, J. Y., Han, K., Shih, P. C., Lee, D.: Generation like: Comparative characteristics in Instagram. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp. 4039-4042. (2015).
6. Centraal Bureau voor de Statistiek (CBS). Retrieved from https://opendata.cbs.nl/ Den Haag, the Netherlands: CBS.
7. Eckert, P.: Age as a sociolinguistic variable. In The handbook of sociolinguistics, pp. 151-167. Blackwell. Oxford (1997).
8. Church, K., Gale, W., Hanks P., Hindle D.: Using Statistics in Lexical Analysis.' In: Zernik U. (ed.), Lexical Acquisition - Exploiting On-Line Resources to Build a Lexicon, Lawrence Erlbaum Associates, pp. 115-164 (1991)
9. Carella, V., Monachesi P.: Greener through Grey? Boosting Sustainable Development through a Philosophical and Social Media Analysis of Ageing. Sustainability 10(2), 499 (2018).
10. Danescu-Niculescu-Mizil, C., West, R., Jurafsky, D., Leskovec, J., Potts, C. No country for old members: User lifecycle and linguistic change in online communities. In: Proceedings of the 22nd international conference on World Wide Web. pp. 307-318. ACM. (2013).