Processes and variations in language economisation

Jonathan R. White
Högskolan Dalarna, Falun, Sweden

ABSTRACT

This article analyses the processes of reducing language in textchats produced by non-native speakers of English. We propose that forms are reduced because of their high frequency and because of the discourse context. A wide variety of processes are attested in the literature, and we find different forms of clippings in our data, including mixtures of different clippings, homophone respellings, phonetic respellings including informal oral forms, initialisms (but no acronyms), and mixtures of clipping together with homophone and phonetic respellings. Clippings were the most frequent process (especially back-clippings and initialisms), followed by homophone respellings. There were different ways of metalinguistically marking reduction, but capitalisation was by far the most frequent. There is much individual variation in the frequencies of the different processes, although most were within normal distribution. The fact that non-native speakers seem to generally follow reduction patterns of native speakers suggests that reduction is a universal process.

1. Introduction

This article investigates processes of reducing language in online discourse, specifically textchat. Previous studies have identified many such processes, but they have been analysed to different depths, and few have looked at comparative frequencies of use; also, none have looked at individual variation to our knowledge. These are therefore the main focuses of this study: to see what processes appear in our data, how much they are used compared to one another, and how much variation there is between individuals.

Another feature of this study is that we focus on the language of inexperienced Internet users. These users have very limited Internet experience, from which we can assume that they have not been much exposed to the conventions of computer-mediated communication (CMC) both in their native languages and in English. Therefore, we wanted to investigate if they used the processes typically described in the literature, which would reflect on the universal nature of reduction processes.

The data come from textchat discussions in academic settings. They comprise different seminar discussions, both with and without a native English-speaking teacher being present. The students and the course the data was taken from are presented in more detail in the second section. The amount of reduction and the sorts of word-classes that are reduced are analysed first, and this is explained as a consequence of the context in which reductions are possible and the frequency of the items in question. After discussing the data, the processes we find in our data are presented, as are the metalinguistic markers for showing that a form has been reduced, such as capitalisation and the use of full stops. Individual variations in the processes are considered finally to identify any particular tendencies individuals show in their reductions, including those which deviate from the general behaviour of the cohort. We begin, though, by presenting previous work on reduction processes.

2. Background on reduction

Reduction is a process that has long been recognised in literature on computer-mediated communication, including texting. CMC discourse is typically described as using a simplified language which has the effect of making communication more efficient (Murray, 2000). Thurlow (2003: §4.2.1) refers to the sociolinguistic maxims of CMC, which are: brevity and speed, paralinguistic restitution (making up for the lack of body language and intonation), and phonological approximation. Thus, reduction is one clear manifestation of the need for efficient, fast communication. Herring and Zelenkauskaite (2009) identified more functions for such reductions than simply efficient and fast communication, but the exact functions of reduction are not the focus of this work.

Many authors have identified the strategies used to reduce language in texting and computer-mediated discourse (texting has been investigated by Hård af Segerstad, 2002; Thurlow, 2003; López Rúa, 2005; Kul, 2007a,b; Crystal, 2008, among others). Two
general types of process have been identified: syntactic reduction, on the one hand, and morphological or orthographical reduction, on the other. Syntactic processes include the deletion of subjects and modal auxiliaries (Murray, 2000) and ellipsis (Lee, 2002). Our focus here, though, is on morphological or orthographical reduction, and there are three basic classes of reduction process that are recognised in the literature: clipping, phonetic respellings and homophone respellings.

The term abbreviation is often used by authors to refer to clipping processes in general (by Murray, 2000; Herring, 2012, for example), although for Werry (1996) the term refers to reduction in general. Typical processes include clippings (info from information), acronyms (NASA from National Aeronautics and Space Administration) and initials (PC from personal computer; we use initialism to avoid the potential ambiguity of abbreviation which is often used as a term for the same phenomenon). López Rúa (2005) classifies clippings (tom from tomorrow) separately from abbreviations (pls from please). Lee (2002) distinguishes clippings of individual items from clippings of sentences. Thurlow (2003) sees a difference between shortenings (back-clippings like the example above), contractions (mid-clippings like pse from please) and G-clippings (go in goin), following the classification in Shortis (2001).

Homophone respelling is mentioned by most authors. Lee (2002: 8–10) goes into more detail into the different types of respellings:

(1) letter homophone (u meaning “you”) number homophone (4 meaning “for”) combination of letter and number homophone (b4 meaning “before”) combination of letter initial and letter homophone (oic meaning “oh I see”)

Rebus writing is sometimes mentioned as a different category from homophone respellings. Lotherington and Xu (2004: : 314ff.) include this under the category hybridised codes:

(2) abbreviations (these include homophone respellings in English, like 4 meaning “for”) hybridised codes (rebus writing like b4 homophones (mostly in Chinese)

Hård af Segerstad (2002) also distinguishes homophone respellings from rebus writing.

Phonetic respellings are mentioned in particular by Yus (2011: 176–179) who categorises them as follows:

(3) phonetic spellings (cos meaning “because”) colloquial spellings (wanna meaning “want to”) homophone spellings (every I meaning “everyone”)

These are all included under the category phonetic orthography (other sub-categories relate to prosodic spellings, regiolectal forms, etc.). Androutsopoulou (2000) distinguishes phonetic from colloquial spellings (non-standard orthography, like wuz for was vs. reductions typical of colloquial speech, like wud for would), and gives homophone respelling as a separate category. Thurlow (2003) also mentions a difference between non-conventional spellings (like sumtime) and accent stylisation (like wivout). Al-Sa’di and Hamdan (2005) distinguish g-dropping from other colloquial and phonetic spellings.

To summarise, we can see a very similar set of basic categories that are recognised by all these authors. These are: clipping processes of different types, phonetic respellings including informal oral respellings, and homophone respellings. The exact classification of particular processes is up for debate, though; for example, should rebus writing be recognised as a separate category from homophone respellings; and what is the status of informal forms like goin and yeah? We recognise the very broad categories for now, and will develop a detailed classification later. We will now present our data and informants in detail.

3. The data

The data analysed was produced by learners of English (28 in total) who were all students on a distance MA programme in English Linguistics run by a university in Sweden. From a survey the author carried out to collect metadata on the learners, we see that their ages range from 25 to 55 (although most were between 25 and 35). Admission onto the programme was contingent on a documented IELTS average score of 7.0, with no lower than 6.5 in each component. These students are mostly novice Internet users even in their native languages—one is a speaker of Bangla, and the rest Vietnamese. Thus, we can suppose that they had not been very much exposed to the discourse norms of computer-mediated communication (CMC) in general, and of CMIC communication in English in particular. This was also the first time any of them had taken a distance course.

The data consists of seminars on different topics from an introduction to core linguistic and sociolinguistic topics run in Autumn 2007. There were nine sessions: a general introduction, language and the media, language and politics, language and gender, phonetics, phonology, morphology, syntax, and semantics/pragmatics. Students divided themselves into four groups, and for four of the sessions (media, politics, gender and morphology) the groups arranged a pre-seminar where they discussed the reading on the relevant topics and data analysis alone without the teachers being present. All these pre-seminars took place through Skype textchat. The chatlogs were sent to the teachers, which helped focus the seminars, and these also took place through Skype textchat. Unfortunately, the logs from the introduction session and final session on semantics and pragmatics had not been saved, and therefore were not available to the author for analysis.

At the start of the course, students were informed about research conducted by their teachers, and were asked to give their consent for material they produced on the courses on the programme to be used in research. Only students who gave their permission were included in this study. All students have been made anonymous in the presentation of the data, and are referred to as, e.g. Student 15, including as address forms in the contributions. Typographical errors have been preserved in the extracts.

The data was analysed by reading through the transcripts and identifying the reduced forms. It was only possible to identify mistypings if there was an explicit correction on the part of a student, so otherwise the forms were treated as deliberate spellings. As much as possible, deliberately generated forms were identified, mainly by whether students repeated their use of particular forms. Repetition of a form, in particular by another student, was taken as confirmation that the form was deliberately generated. The analysis was gone through multiple times to ensure accuracy, and identified forms were searched for in the documents to ensure a correct count. The AntConc freeware concordancer was used to calculate frequencies and to search the corpus in general (available from http://www.antlab.sci.waseda.ac.jp/antconc_index.html).

Now we look at how many reductions there were in the data and what was reduced.

4. What do the informants reduce and Why?

Let us begin by looking at the number of reduced forms that appear in each session of the course. Table 1 shows the division according to session and pre-seminar/seminar (the three sessions without pre-seminars have an n/a, not applicable, in that section of the table):

In the table, we present the reduced forms attested per session of the course. Red refers to the tokens of reduced forms attested; Wrds refers to the number of words in that session; and %Red refers to the percentage of reduction in that session calculated from those
numbers. Thirty transcripts were analysed which contained a total of 93,923 words. The number of contributions was 21442, or an average of 714.73 per transcript. A total of 2024 tokens of reduced forms were identified in the data, which gives an overall reduction percentage of 2.15% (2024 tokens of reduced forms/93,923 words). However, if we focus only on expressions where reduction was attested, the number of full plus reduced forms was 24,851, and then the percentage of reduction is 8.14% (2024 tokens of reduced forms/24,851 words).

The reductions were divided very unevenly between the sessions. Most instances of reductions appeared in the pre-seminars, with 1145 tokens (1145/2024, or 56.57% of the total number of reductions), and the sessions with the pre-seminars, namely language and the media, language and politics, language and gender, and morphology, even have most reduced forms in the seminars (660 out of the 879 tokens of reductions in the seminars, or 75.09%). Even though the pre-seminars contained the fewest amounts of words, they contained the most tokens of reduced forms, and therefore the highest percentage of reduction (3.36%). When considering each session, the numbers are distributed quite evenly across the pre-seminars, with a variation of 0.2% between the sessions. There is more variation across the seminars, from 0.8% in phonetics to 2.29% in language and gender. Three seminar sessions had reduction percentages of over 1.5%, namely, language and gender, and morphology. There was a clear increase in reduction over the first three sessions, both in pre-seminars and seminars. This can be explained as the students getting used to the textchat medium and discourse, and also which enablesthe interpretation of a particular reduced form. Students are discussing a political speech, and are looking for rhetorical devices such as three-part statements. Student 19 refers (incorrectly) to this technique, and it is picked up by Student 13, using the reduction 3PS. It is this context of the full form that enables the interpretation of 3PS. As we see from the next contribution, though, a context does not necessarily lead to a shared understanding, and Student 19 questions what PS refers to.

Function items, on the other hand, are found much more frequently. Consider the following:

**Extract 3**

Student 1 says: Everyone, please join this room
Student 3 says: We R all online?
Student 7 says: ok i’m here.
Student 2 says: let discuss the function of Media
Student 3 says: yes, do U find me, Student 5?

If we now consider the linguistic contexts of reduction, there are two main contexts, as discussed in White (2011), which roughly correspond to the content/function item distinction (item is used here as it is not just individual words but also phrases that are reduced). Content items are reduced when there is a specific discourse context to help with interpretation, and usually when there is a full form earlier in the discourse. Consider the following:

**Extract 1**

Student 23 says: Autumn has discussed Chapter 4 in Lang, So, and Power

**Extract 2**

Student 19 says: We are discussing whether they are three-statement or not
Student 15 says: yes
Student 17 says: Yes. “In a whirlwind of change and hope and peril, our faith is sure, our resolve is firm, and our union is strong
Student 13 says: I mean, he uses 3PS twice, in one sentence
Student 19 says: What does PS mean?

In Extract 1, we have an early contribution in a seminar which refers to the title of the textbook, **Language, Society and Power**. It can only be the context of discussions on a course using this book which enables the interpretation of So in Lang, So and Power as Society, and also which allows such an interpretation without a full form appearing earlier. In Extract 2, a full form allows readers to interpret a particular reduced form. Students are discussing a political speech, and are looking for rhetorical devices such as three-part statements. Student 19 refers (incorrectly) to this technique, and it is picked up by Student 13, using the reduction 3PS. It is this context of the full form that enables the interpretation of 3PS. As we see from the next contribution, though, a context does not necessarily lead to a shared understanding, and Student 19 questions what PS refers to.

Table 1

<table>
<thead>
<tr>
<th>Session</th>
<th>Pre-seminars</th>
<th>Seminars</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Red</td>
<td>Wrd</td>
<td>%Red</td>
</tr>
<tr>
<td>Media</td>
<td>251</td>
<td>7281</td>
<td>3.45</td>
</tr>
<tr>
<td>Politics</td>
<td>279</td>
<td>8579</td>
<td>3.25</td>
</tr>
<tr>
<td>Gender</td>
<td>350</td>
<td>10190</td>
<td>3.43</td>
</tr>
<tr>
<td>Phonetics</td>
<td>n/a</td>
<td>52</td>
<td>6461</td>
</tr>
<tr>
<td>Phonology</td>
<td>n/a</td>
<td>78</td>
<td>6827</td>
</tr>
<tr>
<td>Morphology</td>
<td>265</td>
<td>8045</td>
<td>3.29</td>
</tr>
<tr>
<td>Syntax</td>
<td>n/a</td>
<td>89</td>
<td>6748</td>
</tr>
<tr>
<td>Totals</td>
<td>1145</td>
<td>34095</td>
<td>3.36</td>
</tr>
</tbody>
</table>

If we consider the frequencies of function items, using Leech et al.’s (2001) work on the British National Corpus as reference, we find that you comes at a very high frequency of 6954 tokens (with you at an additional 1391) and ranks 14th overall, and 3rd for

---

**Table 1**

**Number of reduced forms per session.**

<table>
<thead>
<tr>
<th>Session</th>
<th>Pre-seminars</th>
<th>Seminars</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Red</td>
<td>Wrd</td>
<td>%Red</td>
</tr>
<tr>
<td>Media</td>
<td>251</td>
<td>7281</td>
<td>3.45</td>
</tr>
<tr>
<td>Politics</td>
<td>279</td>
<td>8579</td>
<td>3.25</td>
</tr>
<tr>
<td>Gender</td>
<td>350</td>
<td>10190</td>
<td>3.43</td>
</tr>
<tr>
<td>Phonetics</td>
<td>n/a</td>
<td>52</td>
<td>6461</td>
</tr>
<tr>
<td>Phonology</td>
<td>n/a</td>
<td>78</td>
<td>6827</td>
</tr>
<tr>
<td>Morphology</td>
<td>265</td>
<td>8045</td>
<td>3.29</td>
</tr>
<tr>
<td>Syntax</td>
<td>n/a</td>
<td>89</td>
<td>6748</td>
</tr>
<tr>
<td>Totals</td>
<td>1145</td>
<td>34095</td>
<td>3.36</td>
</tr>
</tbody>
</table>
spoken texts. Are comes next with a token frequency of 4707 (22nd overall, and 38th in spoken texts), then about with a frequency of 1524 (71st overall), yes with 606 tokens (153rd overall, and 47th in spoken texts), something with 526 tokens (170th overall), please with 133 tokens (292nd overall, and 13th among all nouns), question with 390 tokens (368th overall, and 45th among all nouns), woman with 232 tokens (404th overall, and 13th among all nouns), language with 221 tokens (516th overall, and 150th among all nouns), and finally example with 196 tokens (818th overall, and 191st among all nouns). Especially man and question are frequently used, with both within the top 400 most frequent items overall (man lies just outside the top 400 at 404th). Thus, the contention that frequency of items is an important factor regarding their reduction is supported. We can expect the teachers’ names to be frequent in discussions as address forms as well, and the contention that frequency of items is an important factor overall comes next with a token frequency of 4707 (22nd overall, and 38th in spoken texts), then about with a frequency of 1524 (71st overall), yes with 606 tokens (153rd overall, and 47th in spoken texts), something with 526 tokens (170th overall), please with 133 tokens (292nd overall, and 13th among all nouns), question with 390 tokens (368th overall, and 45th among all nouns), woman with 232 tokens (404th overall, and 13th among all nouns), language with 221 tokens (516th overall, and 150th among all nouns), and finally example with 196 tokens (818th overall, and 191st among all nouns). Especially man and question are frequently used, with both within the top 400 most frequent items overall (man lies just outside the top 400 at 404th). Thus, the contention that frequency of items is an important factor regarding their reduction is supported. We can expect the teachers’ names to be frequent in discussions as address forms as well, and also the technical term, presupposition, which was explicitly discussed in two discussion sessions, language and the media, and language and politics.

In terms of word-classes of reduced forms, we see the vast majority of lexical item reduction are nouns, with a few examples of adjectives like difficult or different and verbs like see. Regarding functional categories, there is more variation, and we find the following categories, divided among the different sessions (but not divided according to pre-seminar and seminar) as given in Table 3:

Pronominal reductions are the most popular, followed by interjection reductions. As textchat is a highly oral discourse type, it will be characterised by much addressee-addresser interaction, as noted by Werry (1996), and so we expect pronominals to appear at a high frequency generally, and consequently be more likely to be reduced. Interjections are also likely to be present in an oral discourse type like textchat very frequently as well, and can also subsequently be expected to be reduced more often. The numbers for the media, politics, gender and morphology are naturally higher since there are pre-seminars in those sessions. The morphology session is also marked by a jump in the frequencies of reductions, and the amount of analytical discussions can explain this result, we propose, as the students discuss one another’s suggestions and address one another much. Thus, the results above are not unexpected.

5. Reduction Processes in the Data

Now let us consider the types of reduction processes that are present in our data. We see the following frequencies, focusing on the following general categories only, namely, clipping, homophone respelling, phonetic respelling, and combinations of clipping plus homophone or phonetic respellings respellings as given in Table 4:

Clipping is by far the most frequent reduction process followed by homophone respellings. The two together make up over 85% of all reductions.

Now turning to the sub-categories, clippings of various types are found, both back-, fore-, and middle-clippings like the following:

**Extract 4**
Student 1 says: U know the net is huge and this net offers u suggestions on the correct way to interact w’ Internet users. E.g: u must make sure the Subject lines of your message are detailed enough so they explain what your mess is. I’ve got this idea from Panorama—the coursebook I have to deal with at my uni.

**Extract 5**
[13:46:15] Student 20 says: yes, ‘cause it a speech to review what they has served for country

**Extract 6**
[9:20:02 PM] Student 2 says: Student 5, plse tell us the next question

**Extract 7**
Student 1 says: so plz, our group has that Q

**Extract 8**
Student 7 says: I agree w Student 5 because whenever i drink I want to go to sing Karaoke.

---

### Table 2
15 most frequently reduced phrases.

<table>
<thead>
<tr>
<th>Phrase reduced</th>
<th>Number of tokens</th>
<th>Percentage of total</th>
</tr>
</thead>
<tbody>
<tr>
<td>you/your</td>
<td>370</td>
<td>18.28</td>
</tr>
<tr>
<td>yes</td>
<td>149</td>
<td>7.36</td>
</tr>
<tr>
<td>example</td>
<td>137</td>
<td>6.77</td>
</tr>
<tr>
<td>language</td>
<td>112</td>
<td>5.53</td>
</tr>
<tr>
<td>thanks</td>
<td>110</td>
<td>5.43</td>
</tr>
<tr>
<td>please</td>
<td>97</td>
<td>4.79</td>
</tr>
<tr>
<td>question</td>
<td>93</td>
<td>4.59</td>
</tr>
<tr>
<td>about</td>
<td>55</td>
<td>2.72</td>
</tr>
<tr>
<td>Teacher 1’s name</td>
<td>43</td>
<td>2.12</td>
</tr>
<tr>
<td>woman</td>
<td>40</td>
<td>1.98</td>
</tr>
<tr>
<td>man</td>
<td>36</td>
<td>1.78</td>
</tr>
<tr>
<td>presupposition</td>
<td>36</td>
<td>1.78</td>
</tr>
<tr>
<td>are</td>
<td>35</td>
<td>1.73</td>
</tr>
<tr>
<td>something</td>
<td>34</td>
<td>1.68</td>
</tr>
<tr>
<td>Teacher 2’s name</td>
<td>30</td>
<td>1.48</td>
</tr>
</tbody>
</table>

### Table 3
Frequencies of functional category reductions per word-class and per session.

<table>
<thead>
<tr>
<th>Pronominal</th>
<th>Interjection</th>
<th>Preposition</th>
<th>Connective</th>
<th>Copular</th>
<th>Auxiliary</th>
<th>Negation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Media</td>
<td>85</td>
<td>61</td>
<td>19</td>
<td>5</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>Politics</td>
<td>103</td>
<td>91</td>
<td>30</td>
<td>12</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Gender</td>
<td>118</td>
<td>97</td>
<td>40</td>
<td>11</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td>Phone.</td>
<td>6</td>
<td>18</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Phonol.</td>
<td>11</td>
<td>22</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Morph.</td>
<td>75</td>
<td>69</td>
<td>24</td>
<td>5</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>Syntax</td>
<td>13</td>
<td>28</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>411</td>
<td>386</td>
<td>114</td>
<td>38</td>
<td>34</td>
<td>4</td>
</tr>
<tr>
<td>(988)</td>
<td>(41.60%)</td>
<td>(39.07%)</td>
<td>(11.54%)</td>
<td>(3.85%)</td>
<td>(3.44%)</td>
<td>(0.4%)</td>
</tr>
</tbody>
</table>
Combinations of the different types of clippings can be employed to produce forms like the following reduction of *problem*:

*Extract 9*

Student 1 says: Hi Student 27. We all have the same *prb*  
[Phonetics seminar]

In this case, we have a middle-clipping of *o* and a back-clipping of *lem*. Finally, we see initialisms of phrases:

*Extract 10*

Student 13 says: too ambiguous what does *NL* mean  
[Language and the media pre-seminar, Winter group]

*Extract 11*

Student 19 says: To protect people who have used *NW*  
[Language and the media pre-seminar, Winter group]

*NL* stands for *nuclear language* and *NW* for *nuclear weapons*. There are no instances of acronyms in the data, unusually for text chat. We find the following frequencies at which these clipping processes were used (see Table 4b):

Nearly half of clippings are back-clippings. As described in Hamans (1996, 1997), and Fisiak and Hamans (1997), the vast majority of clippings in most Romance and Germanic languages and also Polish are back-clippings. Initialisms are also frequent (together with back-clipping, they make up 75% of all tokens of clipping), but there are very few instances of fore-clippings in particular. This can be explained by the work of Johnson and Eisler (2012) who demonstrate experimentally that it is the first letters of words that are most important when it comes to recognition. Thus, the high frequency of processes like back-clipping and initialisms are evidence of this importance (thanks to an anonymous reviewer for suggesting this work).

Next, we turn to sub-types of phonetic respellings. There are many examples of like the following which we classify as oral stylisation:

*Extract 12*

Student 9 says: and the way they talked about things was very simple but *thru* their use of lang, they exploited all the positive meanings of words  
[Language and the media pre-seminar, Autumn group]

This is an example of a reduction of *through*. Finally, there is an informal oral form of yes:

*Extract 13*

[Language and politics pre-seminar, Autumn group]

The processes appear at the frequencies given in Table 4c:

Informal oral forms make up the majority of instances, with oral stylisation a distant second. Since there were so few examples of the latter, it was decided it would not be meaningful to divide this category any further.

Turning now to homophone respellings, such respellings using numbers appear in the corpus only as examples involving 4, as in Extracts 14 and 15:

*Extract 14*

Student 5 says: Wait *4* other a little bit, then we’ll discuss, right?  
[Language and gender pre-seminar, Spring group]

*Extract 15*

Student 20: it will last *4*ever, ladies  
[Language and gender pre-seminar, Autumn group]

We see some cases of respelling by letter in Extract 16, where Student 5 has reduced *you* and *are* to *u* and *r*, respectively:

*Extract 16*

Student 5 says: hi, everyone, plz join this room, n tell me whether *u r* able to see me or not  
[Language and politics pre-seminar, Spring group]

We find the frequencies of homophone respellings as given in Table 4d:

The forms *u*, *ur*, and *r* make up most of the 433 letter homophones (399 instances, or 92.15%). Regarding number homophones, we find exclusively forms with 4, such as reductions of the preposition *for* and the adverb *forever*. This is a clear difference from usage mentioned in previous literature, as there is far more variety in the homophone respellings found. Further research is needed to see if this is a feature of non-native speaker Internet discourse, or simply a peculiarity of this group of informants.

The final set of forms we will look at contains combinations of techniques. We have the following combination of a homophone respelling and clipping:

*Extract 17*

Student 3 says: *gb & g9* everyone  
[Language and gender pre-seminar, Spring group]

This is a common way of ending a chat, especially by Student 3. The phrase stands for *goodbye* and *goodnight*. The use of 9 relates to the Vietnamese way of pronouncing both night and nine as /να/.—this form is not counted among the number homophones in Table 4d. This is combined with the clipping of *good* to just g.
6. Metalinguistic markers of reduction

We find examples of different processes where the reduction has been marked explicitly. Consider the following:

Extract 20
Student 1 says: yeah, i agree
[Language and the media pre-seminar, Spring group]

Extract 21
Student 18 says: you read carefully the example about the different
[b/t] impairment and disability, you will understand more.
[Language and politics pre-seminar, Winter group]

The use of the slash mark, /, is common for certain reductions, such as the ones above of the prepositions with and between. An alternative marking is illustrated below:

Extract 22
Student 1 says: one of the way lang can influence our thought is via
the media. D’ y think so?
[Language and the media pre-seminar, Spring group]

Extract 23
Student 1 says: Surely, D’s writing is very impressive ‘cos of his
achievements
[Language and gender seminar]

The apostrophe marks the reduction of do in the first case, and because in the second. We also find the use of full-stops, such as in the reduction of something below:

Extract 24
Student 13 says: By the way, are we bound to write on MEDIA,
POLITICS, and GENDER or can we choose S.t else for the essay
[Language and gender seminar]

The capitalisation of initialisms, such as for NW, repeated from Extract 11, is another method of marking the reduction:

Extract 25
Student 19 says: To protect people who have used NW
[Language and the media pre-seminar, Winter group]

The frequencies these markers appear in are given in Table 5:

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Capitalisation</th>
<th>Apostrophe</th>
<th>Full stop</th>
<th>Slash</th>
</tr>
</thead>
<tbody>
<tr>
<td>168</td>
<td>21</td>
<td>20</td>
<td>13</td>
<td></td>
</tr>
</tbody>
</table>

6. Individuals’ use of reduced forms

In this final section, we will look at individual variation in the use of particular reduction processes. Table 6 presents information on all the students’ use of reduced forms:

Recall that clipping and homophone spellings are the most frequent processes in the data followed by phonetic respelling. Larsen-Freeman and Cameron (2008) and Larsen-Freeman (2012) argue that variability is not something to be discounted, but is rather a basic part of a complex system. As Larsen-Freeman (2012) states it, following van Geert (2003), an increase in variability
especially far from the average for the cohort. For clipping, 18
us to see any students whose percentages stand out as being ±
clipping of 70.53. The standard deviation for these percentages
each student. The middle lines show the average frequency for
with clippings (see Fig. 1):
will give a clearer picture of the amount of variation within the
to the average, respectively. This enables us to see any students whose percentages stand out as being especially far from the average for the cohort. For clipping, 18
is a sign that learners are restructuring their language (see
and others discuss individual and group norms, and conclude that even though there may be variations, individuals mostly do follow the same norm-development process as the group they belong to. We will focus first on those students that stand-out regarding these processes, going through them in number order.
We have Student 1 first. She is by far the leader in terms of
the numbers of reduced forms used (549 of the overall 2024
tokens of reduction, or 27.12%). As with most students, she
produced clipped forms the most, and especially back-clippings. Almost half of her reduced forms were back-clippings (216 out of her 549 forms, or 39.34%). Homophone respellings are also very popular for her (133 out of her 549 forms, or 24.22%), while in common with most other students there are no examples of for-clippings or a combination of clippings and homophone respelling. Unusually, she preferred the combination of phonetic respelling and clipping over pure phonetic respelling. Student 4 produced an equal number of clippings and homophone respellings. Student 7 produced more homophone respellings than clippings. Student 13 only produced clippings and homophone respellings. Students 20, 21 and 25 produced more phonetic respellings than homophone respellings. These were some individuals who stood out in terms of the processes they favoured.
Now, though, we turn to variations within each process which
will give a clearer picture of the amount of variation within the
cohort. We will look at figures for each process in turn, starting with clippings (see Fig. 1):
The points on the diagram show the frequency percentages for
each student. The middle line shows the average frequency for
clipping of 70.53. The standard deviation for these percentages
was 17.70, and so the lines above and below the average represent ±1 standard deviation from the average, respectively. This enables us to see any students whose percentages stand out as being especially far from the average for the cohort. For clipping, 18
students lie with the range of ±1 standard deviation. Six lie over
+1 standard deviation (Students 2, 13, 18, 19, 24 and 27), and the rest lay under −1 standard deviation (Students 4, 7, 21 and 25). Those students are therefore unusually frequent or infrequent in their use of clippings. As stated above, though, the percentages are only relative, and so some of these may be attributed to the small or large set of reductions individual students produce. Student 2
had 92% clippings because she only produced them and very few
homophone respellings, and the same was true for Students 13 and
19; similarly, Student 18 produced those plus a single phonetic respelling. Students 24 and 27 produced only clippings, and so had a percentage of 100% for that process. Thus, we can conclude that
these percentages are a result of the relative distribution of the
students’ reductions, and not true individual variations. On
the below −1 standard deviation side, Student 4 produced even
numbers of clippings and homophone respellings, and they formed nearly 90% of her reductions, although she produced tokens of all
processes. Student 7 produced more homophone respelling than
clippings, but just like Student 4, they made up nearly 90% together, plus she produced tokens of all except one process. Students 21 and 25 were more evenly distributed between the processes, but
clippings were simply under 50%. Given that there was more
even distribution between the processes, for these cases we can conclude that the percentages are not simply results of relative distribution, and therefore that these latter four were true cases of variation for this process. Student 25 in particular had one of the highest number of tokens of clipping at 62, and so the percentages for him could not be a result of a low number of tokens. In terms of actual tokens of clipping, Students 1, 3, 17 and 5 had the highest
numbers, and they all lie within normal distribution.
We see the variations for homophone respellings below (see
Fig. 2):
In the case of homophone respellings, the average percentage
was 18.44, and the standard deviation was 13.89. 17 students lie
within the ±1 standard deviation range. Students 4, 7, 10, 14 and
16 lie above +1 standard deviation, while Students 6, 13, 19, 23, 24

<table>
<thead>
<tr>
<th>Student</th>
<th>Clipping</th>
<th>Homophone respelling</th>
<th>Phonetic respelling</th>
<th>Homophone + clipping</th>
<th>Phonetic + clipping</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>342</td>
<td>62.30</td>
<td>133</td>
<td>24.23</td>
<td>27</td>
</tr>
<tr>
<td>2</td>
<td>46</td>
<td>92.00</td>
<td>4</td>
<td>8.00</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>122</td>
<td>87.14</td>
<td>12</td>
<td>8.57</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>40</td>
<td>43.48</td>
<td>40</td>
<td>43.48</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>91</td>
<td>68.42</td>
<td>35</td>
<td>26.32</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>54.55</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>29</td>
<td>40.85</td>
<td>35</td>
<td>40.85</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>32</td>
<td>78.05</td>
<td>8</td>
<td>19.51</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>44</td>
<td>62.86</td>
<td>20</td>
<td>28.57</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>16</td>
<td>61.54</td>
<td>9</td>
<td>34.62</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>42</td>
<td>59.15</td>
<td>22</td>
<td>30.99</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>14</td>
<td>60.87</td>
<td>6</td>
<td>26.09</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>29</td>
<td>96.67</td>
<td>1</td>
<td>3.33</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>28</td>
<td>59.57</td>
<td>18</td>
<td>38.30</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>11</td>
<td>78.57</td>
<td>2</td>
<td>14.29</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>25</td>
<td>64.10</td>
<td>14</td>
<td>35.90</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>95</td>
<td>78.51</td>
<td>26</td>
<td>21.49</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>56</td>
<td>93.33</td>
<td>3</td>
<td>5.00</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>25</td>
<td>96.16</td>
<td>1</td>
<td>3.84</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>57</td>
<td>51.82</td>
<td>15</td>
<td>13.64</td>
<td>36</td>
</tr>
<tr>
<td>21</td>
<td>28</td>
<td>45.90</td>
<td>13</td>
<td>21.31</td>
<td>20</td>
</tr>
<tr>
<td>22</td>
<td>7</td>
<td>58.33</td>
<td>3</td>
<td>25.00</td>
<td>0</td>
</tr>
<tr>
<td>23</td>
<td>25</td>
<td>83.33</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>24</td>
<td>37</td>
<td>100.00</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>25</td>
<td>62</td>
<td>48.44</td>
<td>17</td>
<td>13.28</td>
<td>29</td>
</tr>
<tr>
<td>26</td>
<td>17</td>
<td>73.91</td>
<td>2</td>
<td>8.70</td>
<td>0</td>
</tr>
<tr>
<td>27</td>
<td>1</td>
<td>100.00</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>28</td>
<td>6</td>
<td>75.00</td>
<td>6</td>
<td>12.50</td>
<td>1</td>
</tr>
<tr>
<td>Totals</td>
<td>1333</td>
<td>440</td>
<td>148</td>
<td>97</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 6
Types of reductions per student.
and 27 lie under $-1$ standard deviation. As already discussed, the reductions by Students 4 and 7 were distributed over most of the categories, and so we can see these as true variations. They also produced some of the highest number of tokens of homophone respellings. Students 10 and 14 produced mostly clippings, but all but one instance of the rest of their reductions were homophone respellings; Student 16 produced only clippings and homophone respellings. Thus, the latter three are more consequences of uneven distribution across the categories, and are the result of relative percentages. Of those under the $-1$ standard deviation range, Students 6, 23, 24 and 27 produced no homophone respellings at all; and Students 13 and 19 produced just a single token. Even though these four cases are unevenly distributed, still the numbers of tokens are extremely low, and therefore all can be seen as true cases of variation.

Moving on to phonetic respelling, we see the following variations (see Fig. 3):

In these cases, the average percentage of phonetic respelling was 6.15, with a standard deviation of 9.03, which lead to a negative $-1$ standard variation line. A total of 25 students lay with the $\pm 1$ standard deviation range, and those outside, Students 20, 21 and 25, all lie above $+1$ standard deviation. Phonetic respelling for all of them came in second place behind clipping, and all produced examples of most of the categories, and so these results were significant examples of individual variation.

Next, we will look at the combination of homophone respelling and clipping (see Fig. 4):

Here, the average was 0.16, with a standard deviation of 0.49. This again leads to a negative $-1$ standard deviation line. All 24 students within the $\pm 1$ standard deviation range had 0 instances of homophone respelling plus clipping. The four who produced tokens all lie above the $+1$ standard deviation range, Students 3, 4, 5 and 23, hence the fact that their results were so divergent from their fellow students, and thus constitute examples of true individual variation.

Finally, we turn to variations in phonetic respelling plus clipping (see Fig. 5):

The average here was 4.61 with a standard deviation of 8.24. This also leads to a negative $-1$ standard deviation range. Out of the 24 students within the $\pm 1$ standard deviation range, 15 scored 0. Four students scored above the $+1$ standard deviation line, Students 6, 22, 25 and 26. Students 6 and 22 had a high percentage since they produced very few reduction tokens in the first place (11 and 12, respectively); thus, these are the result of relative percentages, not true variation. Students 25 and 26 were clearer examples of variation, as their distribution over the processes was more even. This was especially the case for Student 25 for whom phonetic respelling plus clipping came in third place.

As we have seen, most individuals do lie within the $\pm 1$ standard deviation range for each process. Many of those results outside could be explained as consequences of using percentages of frequencies, but many could not. Clipping and homophone respelling in particular were subject to greater variability, but we have seen it was rare to find individuals who used the processes radically differently from the cohort average. This supports Regan’s (1994) conclusion that individuals generally follow group norms.

8. Conclusions

After this analysis of textchat data, we can make the following conclusions. Frequency of items has been seen to be a strong factor in whether forms can be reduced. Those with high frequencies
are reduced more frequently, it appears, and without full forms to help the interpretation of the reduction. A wide variety of classifications of reduction processes are found in the literature on this topic. We found a similar range of processes to what has been identified before, and this is in itself an important conclusion, we feel, given that the informants are non-native speakers of English with little Internet experience, as described in the background to the data. Thus, we can argue that these reduction processes appear to be very much cross-cultural and universal, as Crystal (2008) demonstrated for reductions in texting. In terms of frequencies, our results differed from the literature, in that clipping and homophone respelling were the most frequent processes by far. There was a large amount of individual variation in the frequencies of the different processes, with a complex and dynamic picture of
variation emerging. Clipping and homophone respelling seemed to be subject to much more individual variation than the others, but in general most individuals followed the norms of the cohort as a whole.

Future research in this area could look more at differences between native and non-native speakers in English and other languages. Regarding non-native speakers, the issue of whether there are cultural differences in reduction tendencies is a major issue. Are there universal or culture-specific reduction tendencies? Similarly, is genre important in reduction? The data analysed here came from educational contexts, but are there differences in other less formal genres?

We hope that this work, and especially the conclusions about individual variations in frequencies, contributes to an understanding of reduction and how it is used in non-native speaker Internet language.

Acknowledgements

This research was made possible with a personal stipend from Birgit och Gad Rausings Stiftelse for Humanistisk Forskning, and funding from Åke Wibergs Stiftelse and the Culture, Identity and Representation research profile at my university.

Appendix. Full table of frequencies of reduced phrases

See Table A.1.

The following items had reduction frequencies of less than 6:

Frequency of 5: allomorph, answer, for, goodnight, Ha, oh my god, Phuong

Frequency of 4: conversation, do you, female, inflection, message, October, parallelism, Saddam Hussein, Thursday

Frequency of 3: auxiliary, computer-mediated communication, education, especially, hospital, September, somebody

Frequency of 2: biography, communication, dictionary, different, everything, Friday, from, government, minister, Monday, nuclear speak, pronoun, software, statement, through, Tony Blair, university, vocabulary, Wednesday, what, Yen

Frequency of 1: accept, acronym, affricate, back-channel support, brother, business, city, comparative, congratulations, conjunction, consonant, dysphemism, electricity, eventually, examination, forever, fricative, grammar, Hanoi, Language and Gender, Language and Politics, lecture, Master of Ceremonies, message, middle, military, minimal, minute, modifier, morpheme, not, November, object, online, politics, pre-seminar, president, professor, regular, save, semantic derogation, subject, terminal, transcript, Universal Grammar, use, word.

References


http://dx.doi.org/10.1075/ps.5.1.05whi


Härdf af Segerstad Y. Use and adaptation of written language to the conditions of computer-mediated communication. [Ph.D. thesis], Göteborgs universitet; 2002.


