

Some Observations Regarding the Frequencies of *a* and *the* in Text Messaging

Caroline Tagg¹

1. Introduction

What makes word frequency in text messaging so interesting is that the generally stable distribution of the most frequent words is turned on its head. Unlike most (if not all) other corpora, not only is *a* more frequent than *the*, but *you* more frequent than *I*; *have* more than *is*; *me* more than *my*; and *just* more than *so*. Some variation in the order of these items is of course evident in other corpora—*I* and *you*, for example, creep up to second and third position in spoken English (Leech, et al., 2001); while *of* moves from its more normal position as second most frequent word down to the ninth in romantic fiction (Tribble, 2000). Rarely however, if ever, do the above pairs of words swap as they do in text messaging—in fact, as Sinclair (1991: 31) notes, so rarely does the top of the frequency list alter in a fundamental way, that any noticeable change is likely to be of significance.

This paper seeks to explain these reversed frequencies and explore what they reveal about the language of text messaging, by focusing on the use of two items: *the* and *a*, in a corpus of nearly 11,000 messages. As previous research and our own experience can tell us, text messaging is a private and intimate means of communication and this, along with the abbreviation and unconventional spelling which characterises the medium, raises certain methodological issues when collecting data and conducting word frequency counts. Analysis of *the* and *a*, however, suggests that it is not simply abbreviation that shapes the language of text messaging, but also the use of phrases and discourse markers with which texters construe a language specific to text messaging.

2. Methodological Issues

The text message corpus used for this study comprises 10,903 messages (196,077 tokens) collected from between March 2004 and the beginning of 2007, and is a larger version of one I have already reported on (Tagg, 2007). Features of the medium, however, and of the situation in which text messages are typically sent, not only shape the language used but complicate both data collection (as I discuss in Tagg, 2007) and the conducting of word frequency counts.

Generating a word list for a text message corpus using Wordsmith Tools (Scott, 1996) raises certain issues due primarily to the variation in spelling and unconventional abbreviations which characterise text messaging (and which are described by Kasesniemi and Rautianen, 2002; and Hard af Segersteg, 2002; among others). The following message from my corpus includes number homophones (2, 4) and other phonetic spellings (*u*, *wot*, *sumfing*), as well as clippings (*jus*, *prob*, *bday*) and consonant writing (*frm*):

¹ Department of English, University of Birmingham
e-mail: cxt491@bham.ac.uk

- (1) Hi susie did u decide wot 2 get george 4 his bday if not ill prob jus get him a voucher frm virgin or sumfing xxx²

(rather than *Hi Susie did you decide what to get George for his birthday, if not I'll probably just get him a voucher from Virgin or something xxx*)

These variants can distort word lists and mean that certain decisions must be made. The issue here is not unlike that faced by analysts of Chaucer's *Canterbury Tales*, which is similarly characterised by numerous spelling variants (Barnbrook, 1990), and ultimately the decision may have to be made to 'tokenise' the text message corpus. For this study, only two very frequent and straightforward 'spelling lemmas' need be recognised. *U* is treated alongside *you* to form a lemma on the basis that, while *you* is a more frequent form than *u*, the latter is still significantly frequent; and that it is consistently used interchangeably with the more conventional form *you*. The following concordance lines illustrate the fact that there is no apparent semantic difference between the two:

```

6271          rts again! What time do you get in? If you can s
6272          ello Guy, what time r u thinking about leaving?
6273          ld be fine. What time do you need picking up? 7
6274          ust be a terrible time for you, and everyone's bein
6275          hats cool what time do you get back cos we m
6276          real word? wot time sun u planning leaving or did
6277          x Hey tagg, what time r u guys heading 2 spoon
6278          ey honey! What time do u want me n craig to
6279          t, 2029. What time have you booked? Hope you
6280          ey buddy, what time do you finnish work? just ro
6281          e tights. What time are you thinking? Paul Chee
6282          cornwall. What time do u finish work? Alsn and
6283          yet Boo. What time do you finish work? Hey jus
6284          ress? Hi, what time are you coming in? Musgrov
6285          ust finished wot time do u think you'll be done?
6286          ill alright? What time do u want to go? Ready w

```

Figure 1: Concordance lines for *you / u*

Not to include the two together would distort our perception of the frequency at which the second person pronoun occurs. While a similar case can be made for *are* and the letter homophone *r*, the number homophones 2 and 4 are a little more troublesome, because they are used also with their numerical meaning and, in the case of 2, analysis of concordance lines is needed to determine whether *to* or *too* is intended. For the moment, it is sufficient to acknowledge that although *to* and *for* may occur lower on the word frequency list than they would if 'lemmatised', the order of the words in the frequency list is not affected. As for the various forms in which other words occur (such as *tomorrow* which occurs variously as *tomo*, *tomoz*, and *tmrw* amongst other forms), for the time being they remain unlemmatised.

A second issue is that of alphanumeric sequences such as *lookin4ward 2seein u soon*, which comprise strings of letters and number homophones with spaces between words and numbers omitted. The individual words are therefore not recognised by Wordsmith and the whole string (*lookin4ward* and *2seein*) is treated as one word in the frequency counts.

² Names have been changed in all the messages in this paper.

Another issue is that shared by analysts of spoken or more informal written corpora. Text messaging has not only been described as an intimate and interpersonal medium more similar to conversation than typical writing situations in terms of its functions and participants' relationships (Tagg, 2007), but expressions and discourse markers more typically associated with the spoken mode also occur in text messaging (such as, in the following messages, *blimey*, *yeah*, *kinda*, *hmm*, *oi*, *gonna*, *yay*, *ok*, *Cheers*, and *oh yeah*):

- (2) Well done, blimey, exercise, yeah, i kinda remember wot that is, hmm. Xx
- (3) Oi when you gonna ring
- (4) I'm coming back on Thursday. Yay. Is it gonna be ok to get the money. Cheers. Oh yeah and how are you. Everything alright. Hows school. Or do you call it work now

As well as the issue of spelling variants such as *kinda* and *gonna*, another issue raised here concerns the number of standard contracted forms, seen in (4) with *I'm*, as well as the frequent non-standard contractions without an apostrophe: *Hows*, in (4). The question is whether or not to set WordSmith to recognise strings which include an apostrophe as one word (which lowers the overall frequency of words like *I*, as *I'm* and *I'll* for example are treated as distinct items) or two (which affects analysis of words like *don't*). The current decision to treat contracted forms as two words echoes the practice used in exploring the British National Corpus (BNC) (Leech, et al., 2001) and the Bank of English (BoE), and this must be borne in mind when analysing the text message word frequency list—*I* for example, comprises an '*I* group' including *I'm*, *I've*, *I'd* and *I'll*. For the moment, standard contractions without an apostrophe such as *Ive* are counted as separate words.

The choices which all this implies texters have when composing text messages—between standard and non-standard spelling variants, or between abbreviated or speech-like modes of expression—suggests that texters are actively involved in construing a language particular to texting, and that in doing so they draw both on a perceived need for brevity and their awareness of the interpersonal functions of the medium. As we shall see, the language which emerges is very different from that of other registers.

3. *the* and *a* in Text Messaging

3.1 Word Frequency in Text Messaging

A comparison of the twenty most frequent items in the BNC (Leech, et al., 2001: 120) and in my corpus of text messaging (generated with WordSmith's Wordlist tool) shows just how much the language of texting differs from that of the general language corpus.

BNC word frequency

Word	PoS	Frequency (million wds)
the	det	61847
of	prep	29391
and	conj	26817
a	det	21626
in	prep	18214
to	inf	16284
it	pron	10875
is	verb	9982
to	prep	9343
was	verb	9236
I	pron	8875
for	prep	8412
that	conj	7308
you	pron	6954
he	pron	6810
be	verb	6644
with	prep	6575
on	prep	6475
by	prep	5096
at	prep	4790
have	verb	4735
are	verb	4707
not	neg	4626
this	detp	4623
's	gen	4599
but	conj	4577

Text messaging word frequency

Word	PoS	Raw Frequency 196,077 wds	Rounded Frequency million wds
<i>you/u</i> *	pron	7949	40540
<i>I</i>	pron	6129	31258
<i>to</i> **	p/inf	4272	21787
<i>a</i>	det	3582	18268
<i>the</i>	det	3505	17876
and	conj	2954	15066
in	prep	2384	12158
it	pron	2304	11750
x	sym	2187	11154
<i>have</i>	verb	1869	9531
for	prep	1692	8629
<i>be/b</i> ***	verb	1587	8094
<i>is</i>	verb	1576	8038
<i>me</i>	pron	1556	7936
on	prep	1523	7767
<i>t</i>	neg	1514	7721
<i>of</i> †	prep	1420	7242
at	prep	1392	7099
<i>'s</i> ††	g/v	1341	6839
<i>my</i>	det	1284	6548
<i>that</i> †††	c/detP	1287	6564
but	conj	1259	6421
<i>just</i>	adv	1211	6176
good	adj	1195	6094
we	pron	1195	6094
<i>so</i>	adv	1151	5870

* *u* accounts for 3073 of the total occurrences of *you/u* (39 percent)

** Unlike the BNC wordlist, *to* includes infinitive and preposition

*** *b* accounts for 398 of the total occurrences of *be/b* (25 percent)

†† *of* accounts for 0.72 percent of total word count, compared to over 2 percent in most corpora

†† Unlike the BNC wordlist, *'s* includes genitive and contracted verb

††† Unlike the BNC wordlist, *that* includes determiner/pronoun and conjunction.

Figure 2: Comparison of word frequency in the BNC and in text messaging

As well as the fact that *the* is not the most frequent word in texting, the lists above highlight the way in which the order of the most frequently *texted* words are the

reverse of those of the general corpus. In other words, the words swap positions in the frequency list. It is, for example, perhaps no huge surprise to find that the most frequent items in texting are *you/u* and *I* because, as mentioned above, personal pronouns are also higher in informal and interactive conversation (Leech, et al., 2001). Where texting differs from conversation, however, is that *you/u* is more frequent than *I*—bearing in mind that we have included *I* in contractions but also the alternate spelling *u*. Similarly, *the*, having been knocked off the top spot by *you/u*, can be found just below *a* rather than being three times as frequent as *a* (as *the* is in most corpora). Although these are not the only examples: *me* and *my* are also reversed; as are *just* and *so*, the reversal of *a* and *the* seems more startling not only in their upset of the established order but simply because the two items occur a similar number of times throughout the corpus. The fact that *the* is far less frequent in texting than general language, while *a* retains a more similar frequency, is to a certain extent not surprising: Biber, et al. (1999: 267) note ‘greater differences across registers in the distribution of the definite article than of the indefinite article’. In other words, while *a* is relatively stable across registers, the frequency with which *the* occurs tends to change according to the language situation in which it is used. However, in no other corpus is the frequency of *the* quite so low in comparison with *a*. Why, then, are their frequencies so similar, and what does this reveal about the nature of the language of texting? In other words, what is it about texting that results in such mind boggling frequencies?

One clue is suggested by the word list: namely, the absence of *of* in the top words. This is highly unusual—in 1991, Sinclair noted that *of* is usually ‘over two percent of all the words—regardless of the kind of text involved’. In my corpus of text messages, *of* comprises 0.72 percent, lower even than the figure of 1.72 percent found in romantic fiction (Tribble, 2000). The relevance of this for the infrequency of *the* is that, according to Sinclair’s study, 80 percent of the occurrences of *of* are involved in elaborating nominal groups often pre-modified by *the*. This leads us to consider the effect of the speech-like features of texting. Fewer nominal groups is of course also a feature of speech and so it may be that the construal of a speech-like language in text messaging is responsible in part for the lack of *the*.

3.2 The Speech-Like Features of Texting

In fact, we should perhaps not be surprised that *the* is less frequent in an informal and interpersonal medium such as text messaging: Biber, et al. (1999: 267) point out that ‘[i]n conversation, the frequencies of the indefinite and the definite article are more similar’. Even in conversation, however, *the* remains more frequent than *a*—27,351 occurrences of *the* per million words in the conversational parts of the spoken BNC, for example, compared to 17,056 of *a* (Leech, et al., 2001: 223-239). There are obviously other features, particular to texting, that mark it as different from spoken language and which account for the low frequency of *the* (and the relatively high frequency of *a*) and, as we shall see, more of the phatic and interpersonal phrases which characterise text messaging involve *a* than they do *the*. Before looking at the particularly phrasal use of *the* and *a*, another possibility is, of course, simply that *the* is being omitted from otherwise speech-like structures.

3.3 The Omission of *a* and *the*

To explore the extent to which abbreviation plays a role in reducing the numbers of *the*, the occurrences of both *a* and *the* counted in a random sample of text messages. *the* occurred 75 times, and *a* 86, which reflects the general corpus word frequency in that *a* is more frequent but differs slightly from the general list in the size of the gap between them: *a* is slightly more frequent in the sample.

The number of times both articles were omitted was then counted. Assessing where an article was missing was made intuitively, and in most cases seemed uncontroversial. Examples judged to be omissions included:

- (5) Hiya, probably coming home * weekend after next
- (6) You all ready for * big day tomorrow?
- (7) Reckon need to be in town by eightish to walk from * carpark.
- (8) Cant think of anyone with * spare room off * top of my head
- (9) have got * few things to do. may be in * pub later.
- (10) have * good weekend.

Uncertain cases were disregarded, either where it was not clear whether *the* or another word was needed as in:

- (11) Car passed [the/its] mot;

or where it was not clear whether *a* or *the* was needed:

- (12) Currently test driving [the/an] electric car
- (13) i want her to be able to drink cos she hasn't since having [the/a] car

or where it was debateable whether a word was missing at all:

- (14) Just got back from Exit festival in serbia
- (15) prince of wales beer garden is best
- (16) will call u when at uni
- (17) am in reception
- (18) am off for games night

The number of omissions counted came to thirty-four for *the*, and sixteen for *a*. In other words, *the* was omitted twice as often as *a*, which supports the hypothesis that omission accounts for its infrequency. However, if we add the omitted occurrences together with the realised occurrences, we find that *a* and *the* still 'occur' only to a similar extent: 109 incidences of *a* and 108 of *the*. So, while this suggests that the omission of *the* goes some way towards explaining its relative infrequency, it does not fully explain the discrepancy between its frequency in texting and in general corpora—given the numbers, over 150 omissions of *the* would need to be found before the ratio of occurrences of *a* and *the* matched that of general corpora. However, the question remains as to why, if the low frequency of *the* is because of abbreviation, *a* is not similarly omitted. Indeed, why is *a* (at 18,863 occurrences per million words) slightly *more* frequent in texting than in conversation? What this suggests is that the low frequency of *the* cannot be explained solely through structural abbreviation, just as it cannot be explained simply through a linguistic similarity with speech, but must also be addressed by looking at how both *the* and *a* are actually used throughout the corpus.

3.4 Patterns of *a* in the Construal of Texting

3.4.1 Overview

Analysis of *a* in texting reveals a seemingly text-specific phrasal use of the indefinite article, in that it occurs frequently in formulaic and phatic phrases such as *have a good day* and *give me a ring*, as well as in interpersonal hedges such as *a bit*. Before going on to look more specifically at the collocates of *a* and the phrases in which the article occurs, a random sample of twenty occurrences of *a* in context gives an idea of the proportion of occurrences of *a* in frequently occurring phrases:

```
1796         t outside Hey buddy, us a bell when you finnish
1797         st arrived in and there's a couple of things we n
1798         oman and sounded like a pile of snot and tears.
1799         er Kathy here and i have a new number. This is it
1800         at kav so there might be a bit of a selly oak poss
1801         me posted xx Depends a little whether nick joins
1802         mum company ... Have a good time xx Mystery
1803         nd of person who needs a smile to brighten his
1804         in. We bought a couple a Xmas pressies. Wot u
1805         da throw the night. have a good one x Cool cool.
1806         ing.she's better now stil a bit dazed i think.i had
1807         sumed we were getting a lift... Want me to boo
1808         kered so better not. Hve a gud time tho Yipee i'v
1809         ob than dave! Fab. Have a good week, don't work
1810         ow tomorrow. Did u hav a successful shoppin tri
1811         know I have just bought a fish and chip and mus
1812         ow the address give me a txt. steve Hey hav se
1813         nything about it.x Have a good day love to all Y
1814         oreat girl moving down a class tomo, poss new
1815         u know wot i mean! Wot a slapper eh?! X Happy
```

Figure 3: Random sample of concordance lines for *a*

The phrases above can be divided into four groups: firstly, *give us a bell* and *give me a txt*; secondly, *a couple of*, *a pile of*, *a bit of a*, *a little*, *a couple a*, and *a bit*; then *Have a good time*, *have a good one*, *Hve a gud time*, *Have a good week*, and *Have a good day*; and finally *wot a (slapper)*. These phrases, most of which occur frequently throughout the corpus, account for fourteen of the twenty concordance lines (70 percent). What this suggests is that any attempt to explain the frequency of *a* in text messaging must make reference to the phrases in which it occurs.

3.4.2 Collocates of *a*

If we start by comparing the use of *a* in text messaging to that in spoken English (as represented by 20,000 sample lines of the Bank of English's *brspok* subcorpus), we find that *a* in spoken language collocates with the following:

Quantifiers:	lot, bit, little, few, couple, hundred
Prepositions:	of, as, for
Noun modifiers:	a, good, big
Nouns:	job, week, problem
Verbs:	had, have, was, got, is
Adverbs:	quite
Other:	there

Figure 4: Collocates of *a* in brspok (BoE)

We can see that these include a number of quantifiers as well as the adverb *quite*, which could presumably be fulfilling a similar quantifying function, and a limited number of prepositions, noun modifiers and nouns, and some verbs. A common structure would appear to be *There is a* . The list of collocates in text messaging, which were generated through WordSmith's Concord tool, reveals a few differences in use, which begin to highlight the specific ways in which texting is construed.

Quantifiers:	bit
Prepositions:	for, with
Noun modifiers:	the, good
Nouns:	day, time, night
Verbs:	have, had, was, hope, got, get, see
Adverbs:	just, not
Pronouns:	you, I'm
Conjunctions:	and, but, that

Figure 5: Collocates of *a* in my text message corpus

Firstly, the only quantifier to collocate with *a* is *bit*; and nor does *a* collocate with the preposition *of* which immediately suggests fewer elaborated nominal groups. The nouns which collocate with *a* are distinct from those in spoken language, and are references to periods of time. There are far more verbs in the list than among the spoken collocates, of which the most frequent is *have* while others include *hope* and *see* (which are particularly frequent verbs in texting, as I commented on previously in Tagg, 2007); but not *is* – nor *there*. There are also a number of conjunctions, all of which suggests a significant degree of the clausal complexity identified by Halliday (1985) in spoken language, rather than the nominalisation more typical of writing. *Just* and *not* are collocates, but not *quite* (which, as we shall see with *a bit*, may reflect a tendency to hedge and soften rather than intensify statements), and the personal pronouns also collocate with *a*—*you*, in fact, is the most frequent collocate. As we shall see, most of these collocates combine with *a* in the frequent phrases which characterise texting.

3.4.3 Phrases with *a* in Text Messaging

The phrases in which *a* most frequently occurs were identified by generating and sorting concordance lines in Concord (one place to the right of the node word, *a*), and then manually counting. The most frequent phrase *have a good* occurred 236 times, followed by various references to time:

Have a good (236)	day	(35)
	one	(30)
	time	(30)
	week/wk	(24)
	night/nite	(22)
	weekend/wkend	(21)
	evening	(18)
	trip	(8)

Figure 6: *Have a good* ... phrases in text messaging

The pattern is at once a formulaic and phatic phrase, and also one with some flexibility and creativity, as shown by the more unusual instances in the concordance lines below: ‘*Have a good strike!*’ (in reference to union action, line 254 below) and ‘*Have a good debriefing!*’ (line 257 below).

```

246         lem is, i wasn't alone ..! Have a good weekend a
247         Lunch?steve Hope you have a good night! see
248 waive it just for you. Tit! Have a nice day xxx So
249         t was you. Speak soon. Have a good week. Xxxx
250         ike it not. Haha. Try and have a good day, i kno
251         n sort it as we go along! Have a good evening x
252         it. See you later mate. Have a good l. i use m
253         yeah its all good so far. have a good day aswell
254         . Going to bed now. Have a good strike! Sorry,
255 The jury is still out! Ok. Have a good journey. I
256         re crossword clues and have a good week x Do
257         decent weather for me Have a good debriefing!

```

Figure 7: Concordance lines for *have a good*...

There is also variation in terms of the adjective used: *nice* occurs thirty times to describe *lunch* and a *meal* as well as periods of time; *great* occurs twenty-eight times, most frequently in *have a great time*; *lovely* fifteen times, most often alongside birthday or Christmas wishes, as in *have a lovely birthday*; or *Happy bday 2 u!* *Hope u have a lovely day*; *fab* occurs fourteen times, predominantly in *Have a fab day*. Other less frequent phrases include: *hope you have a spiffing good night*; *a smashing day*; *a wicked time*; and *a wonderful afternoon*. In other words, the ‘prototypical’ and formulaic phrase *have a good day* forms the basis of much variation and creativity throughout the corpus.

Another frequent phrasing is *give me a* (53 occurrences) or *give you a* (21), which refer to future phone calls, text messages or other acts of communication:

give me a (53)	shout	(11)
	call	(10)
	text/txt	(8)
	ring	(6)
	bell	(5)
give you a (21)	ring	(9)
	text	(4)
	call	(3)
	shout	(2)

Figure 8: *Give you/me a ...* in text messaging

Sample concordance lines for *give me a* and *give you a* show that the former are predominantly imperative *Ok give me a text* although they also occur in requests with *Could you*, while the latter are prefaced with *Will* (or *I'll*).

```

31      ound tomo afternoon. i'll give you a ring x All don
32      ell if you choose to then give me a shout and i'll
33      Ok just walking up x Ok give me a text when you
34      t me know when you I'll give you a text later cos
35      ed to know the address give me a txt. sam He
36      w? Hey tagg. Could you give me a call when you
37      e shopping successful. Give me a warning when
38      of gays I mean games! Give me a bell when ya
39      , i've finished work now, give me a shout when you
40      saturday. Not sure yet give me a second Loads
41      ill meeting later and will give you a ring after you
42      xx What's happening?! Give me a missed call

```

Figure 9: Concordance lines for *give you a*

The exception in the above concordances is *give me a second* (which interestingly refers in a text-like way to time but does not represent a frequent pattern in terms of the present phrase). *Give me a missed call* is interesting as a phrasing which has surely come about with the use of mobile phones, to refer to the act of ringing a phone without it being answered in order to pass on your number (but also leading to the convoluted phrase (which I've overheard a couple of times): *I've had a missed call from you*).

While *have a good day* and *give you/me a call* highlight in quite explicit terms the personal and speech-like purposes of texting—such as to keep in touch and maintain relations—the hedge *a bit* shows the interpersonal sensitivity which shapes how texters construe their text messages. As an adverb, *a bit* does not simply serve to specify amount. Firstly, as Carter and McCarthy (2006: 64) state, it is 'deliberately vague and informal' and serves to hedge, or soften, language otherwise deemed by the speaker as direct or overly authoritative, as well as hyperbolic, exaggerated or metaphoric (Carter, 2004). *A bit* can also be described as playing an evaluative role, in that (when used with non-comparative adjectives or determiners) it marks a statement as 'negative or critical', as pointed out by Carter and McCarthy (2006: 65), who give the examples: *It's a bit extravagant, isn't it?* and *He's a bit old to be driving, I think*.

The concordance lines below from my corpus suggest that *a bit* with non-comparative adjectives is used predominantly in text messaging with negatively oriented adjectives such as *dazed*, *lame* (meaning *weak* or *pathetic*, as in *a lame excuse*), *miserable*, *flat*, *bizarre*, *useless*, and *gay* (in its derisive sense, similar to *lame*), as well as *mind boggling*:

```

99         ing.she's better now stil a bit dazed i think.i had
100        hat she did, which was a bit lame, i thought tha
101        e fun! Am ok tho feeling a bit miserable in my bi
102        ws, am ok so will try for a bit of independence an
103        y local unfortunately it's a bit flat, infact it has n
104        ank you and see you in a bit. 10mins is your fri
105        ng anything. See you in a bit. How's it going? G
106        it? Well we are running a bit late... it turns out t
107        rial. But she talks quite a bit. So its ok. Any ad
108        i'll come up seaview for a bit. Xxx Just finished
109        rrow when I hope to feel a bit better Just got a te
110        eed2 ask u something, a bit bizarre just say se
111        en u would hear by? Im a bit use less 2 u really
112        te! Ok maybe 5.30 was a bit hopeful...we've only j
113        ink Jeremy Clarkson is a bit sexy now! hope y
114        u guys later. See you in a bit. Paul Lol! Nah was
115        eleanor for shopping in a bit. You lot? I reckon
116        o play with x Well thats a bit gay.. Make a tin fo
117        offered it to me already-all a bit mind boggling really...
118        avour ... Yep see you in a bit. Ax Might not have
119        an comes tomoz up for a bit of motown madness?

```

Figure 10: Concordance lines for *a bit*

A negative connotation can also be perceived in other, less obviously derogatory messages such as line 113 above: *Ok so maybe 5.30 was a bit hopeful* (that is, they did not make it on time), while the seemingly positive assertion (line 114) that: *Jeremy Clarkson is a bit sexy now!* is explained with reference to the full text message: *Did you watch Top Gear ? It was so funny, even I think Jeremy Clarkson is a bit sexy now!* which marks the positive evaluation as grudging or cautious. As well as a negative evaluation, the above uses of *a bit* also serve to downplay or hedge the assertion. In contrast, *But she talks quite a bit* (see above, line 107) can be interpreted as *But she talks a lot* (Carter and McCarthy 2006: 128). It is much less frequent than the hedge *a bit*, with only one other occurrence: *I've got quite a bit of work 2 do, but the does sound tempting*. This suggests, along with the seemingly reduced frequencies of *quite* in general and quantifiers such as *a lot*, that texters perhaps tend towards softening rather than intensifying their language.

A bit of serves as a hedge, but is less evaluative than either *quite a* or *a bit*. It softens the strength of the texter's ambition in: *so will try for a bit of independence* (along with the verb *try*); and playing down the extravagance of the entertainment promised by *up for a bit of motown madness?* (rather than the more unmarked *some motown madness*). Similarly, *in a bit* (which occurs 41 times) or *for a bit* (18 occurrences) modify the length of time involved and thus construe a more intimate language.

Interestingly, *a bit of a* reverts to the more negative connotation also implied by *a bit*, as can be seen in the following concordance lines:

Well done. Often **a bit of a** challenge to do it
 rd. i know this is **a bit of a** cheek especially
 eems to be. I had **a bit of a** cold but gettin bet
 y touch! feel like **a bit of a** cow. But ta for th
 . I know I've been **a bit of a** drip but I will start
 dy wife are having **a bit of a** get together satur
 at might have been **a bit of a** mean message to
 thing has proved **a bit of a** mistake i think- lo
 il out tonight for **a bit of a** piss up, he's not
 cancel lunch. Have **a bit of a** problem. Call you

Figure 11: Concordance lines for *a bit of a*

The use of *a bit* in texting, then, (rather than *a lot* or *a few*) weaves into the construal of texting a concern for interpersonal considerations—both in softening expressions and in evaluating them, just as the phatic (but potentially creative) phrases: *have a good time* (or *day, week, weekend*) and *give me/you a shout* (*ring* or *text*) illustrate the interpersonal functions which text messaging fulfils. These interpersonal phrases are very frequent in text messaging. As we saw illustrated in the concordance lines at the beginning of this section, as many as 70 percent of total occurrences of *a* in the text message corpus can be described as phrases, and it is this phrasal use of *a* which (along with the relative infrequency of elaborated noun groups and abbreviation) accounts for its high frequency in comparison with *the*. As we shall see below, although the use of *the* in text messaging in many ways parallels that of *a*, the phrases in which it occurs are not as frequent as those we have seen with *a*.

3.5 Patterns of *the* in the Construal of Texting

3.5.1 Overview

A phrasal approach to the description of *the* in text messaging (that is, classifying uses of *the* according to the lexical patterns in which the word occurs) is justified not only because it offers a useful comparison with the more strongly phrasal use of *a* in the text message corpus, but also by the limitations involved in classifying uses of *the* on the basis of its functions (Biber, et al., 1999: 263-5; Kachru 2003; Poesio and Vieira 1998; Quirk, et al., 1985), both textual (anaphoric and cataphoric reference), and contextual (situational reference). Firstly, it becomes necessary to then add other types of category: Kachru, for example, also includes what he vaguely calls ‘unfamiliar noun phrases’ such as *the fact that* or *the OP ED page of the NY Times* (which Biber, et al. classify as cataphoric); as well as the use of *the* in idioms; while the superlatives *the last bus to Denver* and *catch the last bus* he confusingly labels ‘fixed collocations’ (which Quirk, et al. call ‘the logical use of *the*’). Furthermore, as Kachru himself demonstrates, it is not always a straightforward or very objective task to identify whether a particular occurrence of *the* is, for example, idiomatic or whether it is making a situational reference. Kachru cites Poesio and Vieira (1998), who find that native speakers vary in their judgement as to whether particular occurrences of *the* indicate anaphoric reference, and which noun phrase is being referred back to. Finally, as we saw above, researchers have yet to agree on the categories they propose. Instead of this approach, then, a more effective description of *the* starts from analysis of the lexical patterns or phrases in which it occurs, similar to the approach adopted by Sinclair (1991) in his analysis of *of*.

Analysis of the phrases in which *the* occurs in texting shows that it occurs in adjuncts, or prepositional phrases referring to time and place, rather than in elaborated nominal groups in subject or object position, while the frequency of the phrase *by the way* reflects again the importance of interpersonal considerations in text messaging. None, however, of these uses of *the* is as frequent as the phrases in which *a* has been seen to occur. The relative lack of patterning can be seen in the following random sample of concordance lines:

```

1715      t looks like you are ill in the morning ? X  Neath
1716      like cluedo but I'm sure the bishops won't mind.
1717      imbo waiting to cross to the other side. Hence th
1718      he lion says when I roar the whole jungle shakes
1719      at list ..? Can you bring the suitcase to work. X
1720      d little friends in town in the afternoon/evening.. J
1721      little thing maybe? X  Is the cd somewhere i can
1722      u c d little thing i left in the lounge?it was free b
1723      r lives together call u in the morning xxx  am ho
1724      ! And lo, it is printed, n the computer switched o
1725      loads of exercise shit in the summer when are yo
1726      ldlife expert. What job? The one in Belgium? Ok
1727      ondon office for work for the next week.02087550
1728      london, and u shall hav the grand slam! X  I am
1729      r long, me! X  Why dont the txts deliver? Maybe r
1730      g. Would u be watching the rugby? Dunno if its
1731      ow's the house without the chicks. quiet and tid
1732      ook soonish and tell me the date i received my
1733      will look out for dean. All the best! No definately-
1734      after screaming child at the mo! u two gals run

```

Figure 13: Concordance lines for *the*

The above concordance lines represent the corpus as a whole in terms of the frequency of references to time: *in the afternoon*, *in the morning*, *in the summer*, and *at the mo*; and place: *to the other side*; as well as the social formulae *All the best!* The lines are also representative in that a smaller proportion of occurrences can be described as phrases than for *a*: in the above set, seven of the twenty lines are labelled phrases (35 percent), compared to fourteen with *a* (70 percent).

3.5.2 Collocates of *the*

If we start, as with *a*, by comparing collocates of *the* in spoken English (again, as represented by the spoken British English component *brspok* of the Bank of English) we can begin to highlight specific ways in which *the* is used in texting.

Prepositions:	of, in, on, at, from, by;
Nouns:	end, way, thing, time, moment, side, world, one
Pre-modifiers:	same, first, other, whole, main, most
Verbs:	is

Figure 14: Collocates of *the* in *brspok* (BoE)

We can see that *the* collocates with a number of prepositions, most frequently with *of*; several nouns such as *end*, *way* and *thing*; and pre-modifiers including

adjectives, determiners and superlatives. The only verb is *is*. All of this suggests a frequent occurrence of *the* in elaborate nominal groups such as *the end of the world* and in lengthy adjuncts such as *at the time*, as well as certain phrases associated with the spoken mode: *by the way*, perhaps, and *the thing is*. If we compare this list with the collocates of *the* in texting, we see a weaker tendency for *the* to occur in elaborated nominal groups.

Prepositions:	for, with
Pre-modifiers:	the, all, good
Pronouns:	you, I'm,
Conjunctions:	and, that, but
Verbs:	have, was, are, will, get, see, going, can
Adverbs:	just, about

Figure 15: Collocates of *the* in text messaging

Firstly, there are few prepositions and those that feature (*for*, *with*, and *about*) are different from those in *brspok*. Most noticeably, *the* does not collocate with *of*, nor with nouns or many noun modifiers, suggesting a lack of elaborated nominal groups. The collocates instead comprise several verbs (*have*, *was*, *are*, *will*, *see*, *going*, *can*), pronouns (*you*, *I'm*), adverbs (*just* and *all*) and conjunctions (*and*, *but*), rather than the prepositions, nouns and adjectives seen in *brspok*—suggesting a greater use of definite noun phrases in speech compared to texting and, again, a clausal complexity in texting which is normally associated with spoken language. The other observation that analysis of the collocates supports is that the patterns in which *the* occurs are less frequent than those in which *a* occurs: neither *in* nor *morning*, for example, are collocates and yet (as described below) *in the morning* is the most frequent phrase in which *the* occurs.

3.5.3 Phrases with *the* in Text Messaging

The most frequent phrases in which *the* occurs are prepositional phrases to do with time and place, as well as certain spoken discourse markers, and formulaic phrases. This again highlights the social and interactive functions of texting, and marks the speech-like interpersonal sensitivity of the texters. However, no one phrase seems to occur particularly frequently, and not as frequently as those with *a*, and it is this which explains the unusual relative frequencies of the two items. For example, the most frequently occurring phrase, *in the morning*, occurs only 35 times, compared to 236 occurrences of *have a good* (the most frequent phrase with *a*).

ke me up before u leave **in the morning** i need to
 nd. If not give me a call **in the morning** to make
 kay, prob you'll see me **in the morning**. E x Ye
 urs. How about meeting **in the morning** for sandm,
 our lives together call u **in the morning** xxx am
 xt a picture of it to you **in the morning** x OK se
 se possibly pick me up **in the morning**. My car i
 sed u r re lesson! See u **in the morning**. P x I'm
 mselves Yes ok will do **in the morning** We Onl
 e They want you to ring **in the morning** if poss I

Figure 16: Concordance lines for *in the morning*

Other references to time include *at the moment* (with 27 occurrences) and the slightly more frequent *at the mo* (occurring 33 times).

Dying of pneumonia **at the moment** well got alittle cold
 h fine, on holiday **at the moment** just relaxing. I have
 having a shit time **at the moment** but we will have it
 en able to do much **at the moment** as my CRB check has
 get hold of anyone **at the moment**. I'll have to let

Doing my career plan **at the mo**. Thats an interesti
 m hard at college work **at the mo**. I've got so much to
 in hospital. She is ok **at the mo** r u home alone
 orague? I'm in ireland **at the mo** with the in laws! Go
 i is beta. I'm at home **at the mo**, that's why i wanted

Figure 17: Concordance lines for *at the moment/at the mo*

What is also interesting about these phrases is the number of times *the* is omitted. While *at moment* occurs only three times, *at mo* occurs twenty-three times throughout the corpus—over two thirds again the occurrences of *at the mo*:

ng to gym and I'm in town **at mo**.So can we make it for
 run here with my bro but **at mo** got stinking cold. Have
 . Can't return the camera **at mo**. Prob. Be able to make
 chine! And sleeping badly **at mo**. Observation nxt wk -
 Naughty gal! Yes, on bus **at mo**. Are you still in town?

Figure 18: Concordance lines for *at mo*

This phrase does not occur in the BoE (except in reference to staying with Mo Mowlem!) and so its occurrence in texting would appear to illustrate the role of abbreviation in the low frequency of *the*.

The other most frequent phrases are *the end of* (with nineteen occurrences) and *the rest of* (with seventeen) which, as these concordance lines show, refer predominantly to time:

it should b back **by the end of week** - not happy
 issue delivered **by the end of the month**. Deliver
 nt to the run up **to the end of term**! Nothing els
 near coming **near the end of Cornwall** now, so

looking forward to **the rest of the month!** See
to do? What **about the rest of me?!** What day
e thats cool, enjoy **the rest of your day.** Hey y
ice weekend, enjoy **the rest of your night** and
5. Should be in for **the rest of the evening** t
should be nice for **the rest of the week.**

Figure 19: Concordance lines for *the end of* and *the rest of*

Looking first at *the end of*, those that do not refer to time often refer to physical locations: such as *near the end of Cornwall*; while exceptions with *the rest of* include the presumably more light-hearted *the rest of me* in the concordance above. Again, as with *at the mo*, it is interesting to note the frequency with which *the* is omitted from the above phrases. There are eighteen occurrences of *end of* in the corpus, and ten of *rest of*. Although the phrases do occur in the BoE they are by no means as frequent as in text messaging.

nds good! Txt u **towards end of day** re exercise.
spuds r gud! Only **until end of today.** This is fro
ess. Won't find out **til end of july** - another co

'ham. Have a jolly good **rest of week** xx Hello! T
but Sat fine too. Enjoy **rest of hols** x We're her
orting tomo now. Enjoy **rest of wk!** L x Hope w

Figure 20: Concordance lines for *end of* and *rest of*

The other frequent phrases with *the* in texting are largely prepositional phrases which refer to place or to travel: *on the way* (17), *on the bus* (6), *in the car* (10), *to the cinema* (13), *at the pub* (12), *in the pub* (18), and *on the train* (12), among others. As with the time phrases, these highlight the role of texting in coordinating social and mundane arrangements (a function described by Ling and Yttri 2002)—and that texting is about doing things, through verbs, rather than elaborate nominal groups.

The twenty-nine occurrences of *by the way*, discourse marker involved in organising segments of text, are interesting because, while the time and place adverbials explicitly reflect the purpose of texting, *by the way* shows how its interpersonal nature shapes the language in a similar way as with *a bit*. As Carter and McCarthy (2006: 262) point out, the phrase is 'particularly frequent in informal spoken contexts'. Assuming that the phrase is being used in a similar way to that identified in speech, *by the way* also indicates a degree of interpersonal sensitivity on the part of texters. Specifically, *by the way* shows how the text segment it introduces relates to the previous one: namely, that it 'represents a temporary digression from the previous segment or a shift in topic' (Carter and McCarthy 2006: 262)

In 11 messages, *by the way* follows the 'text segment' it introduces, for example:

- (19) At home **by the way**
- (20) Have you booked the hut? And also your time off? How are you **by the way?**
- (21) Ok. Have a good journey. I had an email from nicky **by the way.** She seems fine. Speak soon. Xxxxx

- (22) Am on a train back from northampton so i'm afraid not! I'm staying skyving off today ho ho! Will be around wednesday though. Do you fancy the comedy club this week **by the way**?

In another eight, *by the way* precedes the segment, as in:

- (23) Not yet. Just i'd like to keep in touch and it will be the easiest way to do that from barcelona. **By the way** how ru and how is the house?
- (24) That way transport is less problematic than on sat night. **By the way**, if u want to ask anne n rachel to join my bday, feel free. But need to know definite nos as booking on fri. X
- (25) I feel very very ill- sonya says she'll be in late by the way. I'll start getting up now Robin ross nigel, kevin Mc cloud from grand designs? **by the way** i've left a lap top in my car, can you

There are a couple of *and by the ways*:

- (26) Thank you. **And by the way**, I just lost.
- (27) **Oh and by the way** you do have more food in your fridge! Want to go out for a meal tonight? X
- (28) Indeed **and by the way** it was either or - not both !

In other cases, *by the way* opens the message, suggesting a link to a previous message or to communication through other means:

- (29) **By the way**, you are innately very good - the seven twenty nine was waiting for me in the station. Had flapjack for brekkers ... Had a great weekend and, yes, keep me updated xx
- (30) **By the way**, 'rencontre' is to meet again. Mountains dont....
- (31) **By the way**, i've put a skip right outside the front of the house so you can see which house it is. Just pull up before it.
- (32) **By the way**, make sure u get train to worc foregate street not shrub hill. Have fun night x

A number of the messages relate to texter identification, and this seems to be a more text-specific use of *by the way*:

- (33) Not tonight mate. Catching up on some sleep. This is my new number **by the way**. Mike
- (34) you owe me 50p fucker its craig **by the way**
- (35) Hey u. Howz ur day been, i didnt make it out very long last night was home by 5, did u go out in the end. Xx ow its carol **by the way**
- (36) Hello hun how ru? Its barbs here **by the way**. Im good. Been on 2 dates with that guy i met in walkabout so far. We have to meet up soon. Hows everyone else?
- (37) Okay, see uu then then! **By the way** its me, Nicky, trying to steal back my phone! Im back! Hope u r well, happy christmas! x x

- (38) yeah sure thing mate haunt got all my stuff sorted but im going sound anyway promoting hex for nigel.**by the way** who is this? dont know number. Joke

Again, it is interesting to note that this frequent phrase also occurs in a form (or spelling variant) without *the*: this time, with the abbreviation *btw*, which helps to explain the low frequency of *the*.

```
iff and its cold! What does btw stand for? I'm just  
ian network! All fine here. Btw, am back friday morn  
tty fab thanks 2 me! Hehe btw craig where were  
Tee hee. But ta for invite. BTW, running late and mi  
elly, how r u? Hows work, btw how r they 4 staff  
ws on what ur doin 2nite?! Btw do i nt get x x x's  
monday, probably all day, btw did you know this we  
wot the details Yeh yeh. btw charlottes access have
```

Figure 21: Concordance lines for *btw* in text messaging

So, what does analysis of *the* tell us about the language of texting? It suggests a language characterised not by subjects or objects realised by elaborated nominal groups (including *of*), but by adjuncts (time and place adverbials such as *in the morning* and *at the mo*); and by discourse markers typically associated with the spoken mode, chiefly *by the way*, alongside variants without *the*: *at mo* and *btw*. As with the phrases with *a*, we see how the interactive functions of text messaging produce phatic and formulaic phrases, while the occurrence of phrases such as *and by the way* reveal a degree of interpersonal sensitivity on the side of the texters. And crucially, in explaining word frequency, we see a less strongly phrasal use of *the* than of *a*.

4. Conclusion

The starting point for this paper was the highly unusual, if not unique, way in which the typically most frequent words are distributed in text messaging. The example I chose to explore in this paper was *a* and *the*, the frequencies of which are not only very similar, but reversed—*a* is more frequent than *the*. This can be accounted for only in part by features of text messaging which are more typically associated with the spoken mode (such as the infrequency of elaborated nominal groups suggested by the fewer occurrences of *of*), as well as the frequent omission of *the* or the use of abbreviations such as *btw*. Of greater relevance is the distinctly text-specific phrasal use of *the* and, to a greater extent, *a*. Text messaging comprises a high frequency of interpersonal phrases, which account for a third of the occurrences of *the* and 70 percent of *a*. These include phatic (yet creative) formulae such as *Have a good day* and *give us a ring*, repeated reference to time and place in *at the moment* and *on the way*; and apparent signs of interpersonal sensitivity in *a bit of* and *by the way*. Crucially for understanding the frequencies of *a* and *the*, however, we see that the phrases with *a* are much more frequent than those with *the* and it is this, along with the spoken features and the abbreviation, that explains their ‘mind boggling’ word frequencies.

What this also highlights is that abbreviation is only one feature in the construal of texting. As we have seen, the omission of the articles is neither comprehensive nor consistent, and it appears side-by-side with other features characteristic of texting such as the phatic social formulae, the hedging, and the

almost speech-like performance of certain discourse markers. What we see being construed through texting, then, is the appearance of brevity and the performance of communicative features more typically associated with the spoken mode, as well as the inclusion of features specific to text messaging, and it is the combination of all these features that reverse the frequencies of *the* and *a*.

References

- Barnbrook, G. (1990) The computer analysis of spelling variation in Chaucer's Canterbury Tales. MPhil thesis, English Department, University of Birmingham.
- Biber, D., S. Johansson, G. Leech, S. Conrad, and E. Finegan. (1999) Longman Grammar of Spoken and Written English. Harlow, Essex: Pearson Education.
- Carter, R. (2004) Language and Creativity: the Art of Common Talk. London: Routledge.
- Carter, R. and M. McCarthy. (2006) Cambridge Grammar of English: a Comprehensive Guide. Cambridge: Cambridge University Press.
- Halliday, M.A.K (1985) Spoken and Written English. Oxford: Oxford University Press.
- Hard af Segersteg, Y. (2002) Use and adaptation of the written language to the conditions of computer-mediated communication. Doctoral thesis, Linguistics Department, University of Goteborg. Available online from: http://www.ling.gu.se/%7eylvah/dokument/ylva_diss.pdf (accessed 25th June 2007)
- Harkin, J. (2002) Mobilisation: the growing public interest in mobile technology. London: Demos.
- Kachru, Y. (2003) 'On definite reference in world Englishes' *World Englishes* 22/4: 497–510.
- Kasesniemi, E.-L. and P. Rautianen (2002) 'Mobile culture of children and teenagers in Finland' in J.E. Katz and M. Aakhus (eds) *Perpetual Contact: Mobile Communication, Private Talk, Public Performance*, pp. 10-192. Cambridge: Cambridge University Press.
- Leech, G., P. Rayson and A. Wilson (2001) *Word Frequencies in Written and Spoken English: based on the British National Corpus*. London: Longman.
- Ling, R. and B. Yttri (2002) 'Hyper-coordination via mobile phones in Norway' in J.E. Katz and M. Aakhus (eds) *Perpetual Contact: Mobile Communication, Private Talk, Public Performance*, pp. 139–69. Cambridge: Cambridge University Press.
- Oksman, V. and J. Turtianen (2004) 'Mobile communication as a social stage: meanings of mobile communication in everyday life among teenagers in Finland'. *New Media and Society* 6/3: 319–39
- Poesio, M. and R. Vieira 1998 'A corpus-based investigation of definite description use. *Computational Linguistics* 24/2: 183–216.
- Quirk, R., S. Greenbaum, G. Leech and J. Svartvik (1985) *A Grammar of Contemporary English*. London: Longman.
- Scott (1996) *Wordsmith Tools*. Oxford: Oxford University Press.
- Sinclair, J. (1991) *Corpus Concordance Collocation*. Oxford: Oxford University Press.
- Tagg, C. (2007) 'Corpus-based analysis of SMS text messaging' in Teo, P. and C. Ho (eds) *Discourse in the Modern World: Perspectives and Challenges*. Singapore: McGraw-Hill.

Tribble, C. (2000) 'Genres, keywords, teaching: towards a pedagogic account of the language of project proposals' in Burnard, L. and T. McEnery (eds) *Rethinking Language Pedagogy from a Corpus Perspective: Papers from the third international conference on Teaching and Language Corpora*, pp. 75–90. Frankfurt am Main: Peter Lang.