

Code, mixing: A corpus-based investigation of ELF and netspeak features in European demoscene texts

**Paul Collins** 

**Degree: English Language** 

Supervisor: Dr Clare Wright

March 2016

# Table of Contents

Abstract	iv
Acknowledgements	iv
Chapter 1: Introduction	1
Chapter 2: Literature review and research questions	2
2.1 English as a lingua franca	2
2.2 Netspeak	3
2.3 The demoscene and the ZX Spectrum	4
2.4 Research questions	6
Chapter 3: Methodology	7
3.1 Ethical considerations	7
3.2 Corpus construction	7
3.3 Feature analysis	9
Chapter 4: Results, analysis, and discussion	13
4.1 Introduction	13
4.2 ELF features	15
4.2.1 Grammar at word level	15
4.2.2 Grammar at clause level	16
4.2.3 Semantic phenomena	20
4.3 Netspeak features	23
4.3.1 Abbreviations	23
4.3.2 Capitalisation	24
4.3.3 Non-standard spelling	25
4.3.4 'Time-stretching' of syllables with repeated letters	25
4.4 Other observations	26
4.4.1 Regularisation of noun and verb forms	26

4.4.2 Code-switching and code-mixing	
4.4.3 Attitudes to English	
Chapter 5: Conclusion	
5.1 Summary and implications	
5.2 Limitations	
List of references	
Appendices	
Appendix A: Some representative demo texts	
Appendix B: Custom software tools for text extraction	40
Appendix C: Custom software tools for data analysis	
Appendix D: Innovative lexical items	

# ABSTRACT

The recent paradigm of English as a lingua franca (ELF) suggests that non-native English speakers (NNES), rather than being deficient learners aspiring to a native variety, are often capable of successful endonormative communication in a form of consensus-driven English that would be regarded as erroneous from a native-speaker viewpoint. Most previous ELF studies have looked at speech rather than writing.

Referring mainly to the ELF research of Jenkins, Seidlhofer, and Dewey, and to Crystal's work on the innovative and visually creative 'netspeak' of Internet users, the author builds a grammatically-tagged corpus of English-language texts from the European demoscene – an informal, collaborative youth culture that creates and shares computer artworks – and analyses those texts in terms of the lexicogrammatical feature sets typically associated (i) with ELF and (ii) with netspeak.

The study uncovers (and attempts to explain) similarities and dissimilarities in both categories. The presence of code-switching and varied attitudes toward English are briefly discussed in light of the ELF debate. Several non-standard features are interpreted as simplifying regularisations, and it is argued that the corpus as a whole demonstrates fluent and successful use of non-standard English, predicated on flexibility and group identity rather than prescriptive native standards of correctness and error.

# ACKNOWLEDGEMENTS

I am grateful to Matt Westcott, for supplying an annotated data dump of the Demozoo archive; Piotr Czerski, for providing a machine-readable version of his article *Maszyny, które mogą wszystko* ('Machines that can do anything'); Gleb Albert, for pointers to online demoscene resources; and my supervisor, Dr Clare Wright, for her helpful advice.

# **CHAPTER 1: INTRODUCTION**

The recent paradigm of English as a lingua franca (ELF) suggests that non-native English speakers (NNES), rather than being deficient learners aspiring to a native variety, are often capable of successful endonormative communication in a form of consensus-driven English that would be regarded as erroneous from a native-speaker viewpoint. Most previous ELF studies have looked at speech rather than writing.

One area that has yet to be investigated in relation to ELF is that of the innovative and visually creative 'netspeak' of Internet users, characterised by experimentation and rejection of formal written norms. Both ELF and netspeak are associated with youth cultures, and both represent non-standard forms of English.

The demoscene, introduced in the literature review below, is a European youth subculture involving the creation and exchange of computer art. Its informal language crosses international boundaries and it is therefore likely to involve both ELF and netspeak features; it therefore represents a suitable source of authentic texts (rather than e.g. artificially elicited classroom data) for a novel corpus-based investigation into the presence of these features, the communicative success of the participants in the subculture, and what the implications might be for teaching English to NNES.

# CHAPTER 2: LITERATURE REVIEW AND RESEARCH QUESTIONS

## 2.1 English as a lingua franca

The global English-speaking population may be broadly partitioned into native and nonnative English speakers (NES and NNES), the former category being those for whom English is a first language (L1). Even the NES portion of this apparently simple division is complicated by the multiplicity of worldwide 'New Englishes', or World Englishes, such as those of Singapore and Ghana, which vary in lexicon, grammar, and pronunciation (Crystal, 2012). Investigations into the use of English among NNES are therefore relatively recent, and still somewhat conventionally restricted to limited sociocultural settings such as the classroom and international business-speak. Such studies may also insist on framing NNES as "learners" and thus *a priori* classifying their usage as deficient (Dewey, 2007, p. 55).

In the connected age of the Internet and the European Union, and with NES now outnumbered by NNES (Crystal, 2012), there are many situations where NNES with differing L1s may come into contact and use English as a common language, or lingua franca (ELF). In such scenarios, English is neither 'foreign language' nor learner target, but a convenient tool for *endonormative* communication (i.e. that proceeding in, and upholding the norms of, a culture or cultures that do not traditionally use English):

ELF speakers ... tend to retain their different national, regional, local and individual lingua-cultural identities. [ELF] is no longer owned by the inner circle of English native speakers, rather [it is] a kind of 'pluralized English' that accommodates diverse speakers' needs, norms and values. (House, 2012, p. 173)

The work on ELF is generally seen as having begun with phonology. Jenkins (2000) proposed a Lingua Franca Core (LFC), a set of strategies for pronunciation of English between NNES. In contrast to the traditional view of NNES as deficient (i.e. an imperfect approximation of native English), the LFC focuses on mutual comprehension and avoidance of ambiguity: for example, substituting  $/\theta/$  with /f/ does not tend to impair communication (and, in fact, even occurs in native varieties such as Estuary), whereas conflating long and short vowels can lead to confusion (e.g. *live* versus *leave*). In the LFC model, non-native

accents are not second-class citizens but regional standards, reflecting local sociolinguistic norms.

This approach has now been extended to lexicon and grammar. ELF is not a single "monolithic" variety (Jenkins, 2007, p. 19), but it does exhibit certain characteristic traits. Jenkins in fact suggests that ELF should not be measured against an ENL baseline (p. 2), but since ELF is a form (or forms) of English, such comparison is warranted, in the same way that one might reasonably compare the features of, say, Scottish and Caribbean Englishes – as long as it is not carried out from a deficiency standpoint. In practice, typical lexicogrammatical features have been described by Seidlhofer (2004) and Dewey (2007) and these are outlined in the methodology.

Poppi (2013), one of relatively few researchers to study written ELF corpora, remarks that written ELF "has not received much attention" (p. 15). Cogo and Dewey (2012) also point out that most ELF studies are based on speech corpora (p. 3) and that non-native usage is underrepresented in corpus studies in general, suggesting the need for corpus analyses of non-standard Englishes in order to "redress the imbalance" (p. 21).

# 2.2 Netspeak

For young Europeans, English can symbolise an international youth culture (Cheshire, 2002, p. 31; Pennycook, 2003, p. 11), and one setting for youth interaction is the relatively new sphere of home computing and the Internet.

Computer-mediated communication (CMC), while virtually unknown outside government and academia up until the late 1970s (Herring, 1996, p. 1) has since been the focus of much and varied research. CMC has, for example, been analysed as a social arena for group interactions (Korenman & Wyatt, 1996) where conflicts must be resolved and netiquette observed (Kollock & Smith, 1996); as a tool for education and cross-cultural exchange; and from critical perspectives including those of film studies and feminism (White, 2006). It has even spawned new literary forms such as codework (Sondheim, 2001) and hyperfiction. CMC thus touches upon practically every issue in applied linguistics.

One commentator who has documented many of the overt orthographic and discourse features of CMC is Crystal (2006), who concludes that 'netspeak' is neither spoken writing nor written speech; it is too wide-ranging to be viewed as a single variety, and is rather a

"fourth medium" (pp. 271-272), after speech, writing, and signing. Nevertheless, Crystal isolates a number of features that tend to occur consistently in online textual interactions; these are outlined in the methodology.

The construction of a corpus reorders language so that it can be methodically examined in ways that would otherwise be impossible (Hunston, 2002, p. 3). The corpus approach is particularly well suited to the study of computerised texts (Baker, 2010, p. 13) and has been successfully utilised in diverse areas of CMC.

## 2.3 The demoscene and the ZX Spectrum

One route into studying CMC in the rarely-examined ELF context is through a phenomenon known as the demoscene, an informal community of practice that originated in Europe (Montfort, 2013, p. 5) during the home-computing boom of the 1980s (Lean, 2013). 'Sceners' collaboratively create and exchange computer art, music, and 'demos': self-contained software programs that generate non-interactive audiovisual spectacles somewhat akin to music videos (Montfort, 2013). At international demoparties, sceners meet in order to socialise and enter their works into competitions; this potentially gives them occasional venues for exchanging spoken language as well as written, and the environment is likely to be conducive to the "overtly consensus-oriented, cooperative and mutually supportive" character of ELF talk (Seidlhofer, 2004, p. 218).

Demos are often produced by multi-member collectives, or 'demogroups'. While the demoscene today is organised around community Web sites such as Pouët (www.pouet.net), it formerly operated through local dial-up bulletin board services (BBS), and also by the physical swapping of computer tapes and disks via postal mail; demos, as messages between sceners, therefore often served an epistolary function and could be an innovative way to gauge whether the new medium retained or changed ways of communicating, particularly in the period of upheaval after communism (Czerski, 2014).

Modern demos tend to focus on visual effects, but the earlier ('old-school') productions typically contain a significant amount of text, often in a scrolling message that accompanies the visuals (Marecki, 2015). Such text may include news, jokes, technical commentary, contact details, calls for collaboration, and greetings to other sceners; it may be in any or many languages, by one or several authors (frequently NNES writing in English),

4

and can involve both linguistic and stylistic code-switching, as well as typographical innovations and other visual paralanguage.

A good source of early European demos (and thus texts) is the ZX Spectrum computer. Marketed from 1982 until 1992, the Spectrum was a first computer for many in Britain; it enjoyed similar success in much of Europe (Kelion, 2014) and had an established international demoscene by the early 1990s. The Spectrum also attained popularity in the Soviet Union, in the form of illicit 'cloned' versions built by hobbyists based on technical schematics from the West (Stachniak, 2015).

European Spectrum demos achieved some degree of mainstream recognition in the UK by way of the high-street magazine *Your Sinclair*, which in its latter days – with the games market drying up – began printing commentary on the demoscene (Pillar, 1992) and including demos on its cover-mounted cassette. Today, hundreds of Spectrum demos can be obtained from Web sites such as Demozoo (www.demozoo.org) and Pouët, and the Spectrum, nicknamed the 'Speccy' by users, remains a favourite with retrocomputing enthusiasts, who have, for instance, modified the machine to connect to the Internet (Orlowski, 2010).

The early demoscene pre-dates the World Wide Web and inexpensive consumer Internet access by about half a decade, but it seems plausible that some netspeak features, as identified by Crystal, might have their parallels in the informal, collaborative speech-like writing of the demoscene. (Nor were the two entirely distinct: some Web users came from a BBS background, and Czerski cites the rise of IRC – Internet Relay Chat – as the cause of a gradual reduction in the exchange of demoscene letters and media by postal mail.) Demos, as written, published and swapped by sceners, represent an early and uniquely asynchronous form of CMC; yet, to the author's knowledge, the only research alluding to the text *within* demos has been Marecki's (2015) ethnographic exposition of the Polish demoscene.

The presence of international NNES writing in English suggests that demos could be an untapped source of authentic written ELF that significantly predates the material typically studied in ELF contexts and avoids the 'observer's paradox' that may arise with e.g. group audio recordings artificially elicited for the purpose of study. Given the often local or regional structure of demogroups, there is also the possibility of 'clusters' of nearby sceners spreading and sharing lexicogrammatical features, perhaps influenced by their common L1.

5

Appendix A provides some representative portions of Spectrum demo scrolltexts in English, by way of an introduction to the genre.

# **2.4 Research questions**

- Which of the distinctive features of ELF and netspeak can be found in Spectrum demos (as a representative slice of the early demoscene)?
- Do these features vary in prevalence according to the L1 of the writer, and if so how?

# **CHAPTER 3: METHODOLOGY**

# 3.1 Ethical considerations

While informed consent is typically required for linguistic fieldwork with individuals (Bell, 2010, p. 46), this is not the case with demos, which are voluntarily published works. Likewise, there is no need to conceal creators' names (Bruckman, 2002), which in any case are generally self-selected pseudonyms.

Demo texts often contain sceners' contact details (address, telephone, etc.) for purposes of communication and software swapping; these are redacted from texts presented here. (TagAnt removes them automatically in the case of e-mail addresses and Internet domains.)

# **3.2 Corpus construction**

The corpus analysed here is based on a copy of the Demozoo archive (Demozoo, n.d.) from June 2015, comprising 2,349 Spectrum demos plus metadata including the titles, release dates, authors (often a demogroup), and authors' nationalities. The latter is the amalgamation of all the distinct nationalities (where known) of credited authors, or failing that, the nationality of the demogroup (if known to be from a single country), and, though not guaranteed, is taken to indicate the probable L1 of the writer.

Demos are distributed in various formats and generally employ data compression routines; this means that the raw file, prior to execution of the program stored inside it, rarely contains any recognisable text. Therefore, each demo was manually launched in a Spectrum emulator (Needle, 2015), and a 'snapshot', or record of the entire memory contents, was captured whenever a text appeared on screen. At the end of this process, custom-written software was used to extract the text from the snapshots. Appendix B discusses the text extraction tools in greater detail.

Texts with little or no English were discarded. All texts were then lower-cased (a prerequisite for tagging) and tagged for grammatical parts of speech using TagAnt (Anthony, 2014), an implementation of TreeTagger (Schmid, 1994).

7

Table 1 shows a breakdown of the texts and tokens in the corpus by originating country. Tokens do not precisely correspond to words, due to the nature of tagging: for example, individual punctuation marks are separate tokens.

Originating country	Probable L1	Tokens	Texts	Avg tokens
				per text (0 d.p.)
(unknown or multiple)	-	243549	286	851
Great Britain	English	175767	90	1952
Poland	Polish	54322	83	654
Germany	German	39737	44	903
Russia	Russian	32889	58	567
Czech Republic	Czech	30808	41	751
Slovakia	Slovak	28247	29	974
Belarus	Russian	14946	16	934
Ukraine	Ukrainian	13828	15	921
Netherlands	Dutch	9970	15	664
Norway	Norwegian	9086	17	534
Lithuania	Lithuanian	4552	5	910
Greece	Greek	4130	8	516
Austria	German	3453	8	431
Sweden	Swedish	2272	6	378
Latvia	Latvian	576	2	288
Turkey	Turkish	455	1	455
Argentina	Spanish	415	2	207
Total:		669002	726	921

Table 1: Corpus breakdown by originating country

The mean sentence length is approximately 13 (669,002 tokens in 52,539 sentences, counting non-empty sentences after splitting on punctuation marks . *!* ? ), while, for example, that of the Arts domain of the British National Corpus (BNC) exceeds 22 (Gilquin, 2010, p. 36). Short sentences are characteristic of speech and thus of the speech-like writing of CMC.

## **3.3 Feature analysis**

In the separate cases of ELF and netspeak, a number of features have been identified as characteristic. ELF features have been described by Seidlhofer (2004) and Dewey (2007), and netspeak features by Crystal (2006). Since these features are already defined as differences from standard or native-like English, it is sufficient here to analyse the demo corpus for their presence or absence, and unnecessary to compare it with external sources or to measure actual feature occurrences as a proportion of potential points of occurrence.

A small pilot study of the tagged data was conducted in order to develop plausible search strategies. Since tagging is rarely perfect, especially in the presence of non-standard vocabulary and punctuation (Baker, 2010, p. 11), and there are many ways to phrase an utterance, these searches would not find all occurrences of a feature, and could also produce various 'false positives' requiring manual removal; they should, however, be sufficient to demonstrate whether a given feature exists and to gauge its approximate frequency.

The method consisted of running a set of custom searches on the corpus to produce (for each feature) a list of possible matches and their sources, in the form of a feature concordance file (Appendix C). These files were filtered manually to remove false positives. The remaining instances – and any other cases of note arising from a general study of the corpus – were then discussed as evidence for the feature. To determine whether nationality (i.e. predicted authorial L1) might be a factor in usage, the number of occurrences per nationality was computed, normalised for the number of tokens available in the corpus for that nationality; this is referred to as the *feature density*.

Features and corresponding search strategies (with heuristics to eliminate significant false positives arising from the pilot study) are presented in Table 2. The exact TreeTagger search patterns used can be found in the source code on the accompanying CD. Following Dewey (2007, p. 80), deprecatory terms such as 'confusion' and 'overuse' are avoided in descriptions, even where such terms were used in the source material. ELF has its own communicative norms; these happen to be more flexible and less codified than native English, but this does not make ELF deficient relative to native English, merely different (Jenkins, 2007).

9

Emoticons and non-standard punctuation, while present in some texts, were not measured because of space considerations; since these paralinguistic elements stand apart from the words of the text, they are of lesser communicative significance.

Feature	Search strategy
ELF: Grammar at w	vord level
Third-person zero,	Noun or third-person pronoun, followed by non-third-person present-
e.g. <i>he look</i>	tense verb, omitting common false positives (e.g. enter could be a
	verb but mostly occurs as press Enter).
Insertion of	A/an/the tagged as determiner and not clause-initial (since such
articles	articles are rarely redundant).
Omission of	Preposition followed by noun. Still many false positives, but absence
articles	is hard to search for, and – from observation – articles are frequently
	omitted after a preposition (e.g. arrived at party, all of members).
ELF: Grammar at c	lause level
Exchange of	Who/which tagged as pronoun or determiner, immediately preceded
relative pronouns	by a noun or pronoun.
who and which	
Tag question not	Clause-initial verb other than common false positives (e.g. got it?),
matching main	and with explicit cases for missing apostrophes (e.g. isnt),
verb	immediately followed by optional -n't, personal pronoun, and
	question mark. Also search for or? and no?.
Redundant or	Verb followed by preposition (other than mistagged conjunctions,
innovative	and by, mostly used in author credits), followed by noun or pronoun.
preposition use,	
e.g. study about	
Replacing	Verb followed by <i>that</i> (regardless of tag, since it mostly fails to tag
infinitive-	correctly as a complementiser), omitting common false positives
constructions with	such as believe that. An alternative approach based on a whitelist of
that clauses, e.g. I	verbs that frequently take the infinitive (British Council, 2016) was
want that he	less successful and so discarded.

#### Table 2: Features and corresponding search strategies

Feature	Search strategy
Innovative word	Search for (i) clause-initial adverb (omitting some common false
order with	positives) immediately followed by a personal or possessive
adverbials	pronoun; and (ii) verb (other than be or have) immediately followed
	by adverb and then noun, proper noun, (pre)determiner, or possessive
	pronoun.
	(i) finds e.g. simply I need; (ii) finds e.g. I need also.
ELF: Semantic phe	nomena
Frequent use of	Any form of the verbs do have make put take, not clause-final or
verbs of high	followed by a preposition (both suggesting intransitivity) nor in a
semantic	negation, nor followed by a pronoun (likely to be <i>do you think</i> etc.);
generality	and omitting modal have to.
Overdoing	Adjective (other than commonly mistagged determiners many and
explicitness, e.g.	second) immediately followed by noun or personal pronoun. This
black colour	yields many false positives but there is no obvious alternative.
Transitive verbs	Any of 22 very common, primarily transitive verbs (Linguasorb,
used as if	2016) occurring clause-finally but without a question mark (because
intransitive, e.g. I	of inversion, e.g. what do you want?).
don't want	
Lexical	Generate a list of words that are not in a standard English dictionary
innovation	(see Appendix C for details).
Netspeak	
Abbreviations	Any abbreviation from Crystal's list (2006, pp. 91-92), omitting
	common false positives (e.g. so, which is a common word in the
	corpus but rarely or never stands for 'significant other').
Lower case as	Determine whether each text contains more wholly upper-case tokens
default	or more wholly lower-case tokens (ignoring tokens that contain a
	mixture).
Non-standard	Crystal (2006) gives final $z$ for $s$ as a common netspeak pattern.
spelling	Search for any word ending in $z$ (or a series of $zs$ ) that can be found
	in a standard English dictionary only with an s. Locate other
	examples via the 'lexical innovation' search described above.

Feature	Search strategy
'Time-stretching'	Search for any word containing three of the same letter in a row,
of syllables with	omitting common false positives such as III (number three) and NSSS
repeated letters,	(a particular meetup event).
e.g. yeeeaaah	

# CHAPTER 4: RESULTS, ANALYSIS, AND DISCUSSION

# 4.1 Introduction

This chapter analyses the demoscene corpus of 669,002 tokens in order to address the research questions, i.e. which distinctive ELF and netspeak features can be found in the corpus, and how, if at all, they vary in prevalence by L1. Features are ordered as in Table 2, so that section 4.2 deals with the ELF features (word level, clause level, and semantic) and section 4.3 with the netspeak features, while the concluding section 4.4 addresses some further points arising from observation of the corpus.

Table 3 shows the feature densities per feature and country (for those features that were amenable to numerical comparison). Owing to space constraints, the feature concordance files themselves have been relegated to the accompanying CD. These are in CSV format, viewable in a spreadsheet program.

Demos of British origin are included in the corpus. Jenkins (2007, p. 2) does not consider NES texts to represent ELF, though Modiano (2009) disagrees on the grounds of multiculturalism and the amount of NES input to which many ELF users are exposed. Since NES are deeply familiar with a native variety and thus probably less likely to innovate, the discussion here will focus for the most part on NNES texts.

Texts are generally presented in lower case, which aids human-readability (Butterick, 2014) and was also a requirement for tagging. The exception is section 4.3.2, which explicitly discusses patterns of capitalisation. However, in all cases, the original texts, prior to lower-casing, can be found on the CD.

The corpus analysis is necessarily somewhat exploratory, because the mistaggings resulting from imperfect automated tagging preclude a strictly quantitative analysis. For this reason, stated quantities such as numbers of nouns and verbs are sometimes flagged as approximations, and attempts have been made to identify overall patterns by observation.

13

	AR	AT	ВΥ	CZ	DE	B	GR	LT	LV	NL	NO	Π	RU	SE	SK	TR	NΑ
Third-person zero			2.1	2.3			2.5					2.3	0.7	4.4	0.8		5.8
Insertion of articles	48.2	8.7	13.4	5.6	0.5		4.9	2.2		2.1		5.2	1.9		0.8		9.4
Omission of articles		29	22.8	30.6	7.6	1.1	4.9	4.4	34.8	4.1	5.5	24	12.2	4.4	20.2	99	31.9
Exchange of rel prons 'who' and 'which'				0.4								0.2					
Redundant or innovative preposition use		5.8	0.7	3.6	0.5			6.6			11	3			0.4		1.5
Innovative word order with adverbials			2.1	1.7	0.5	0.1						2.3	0.3		0.8		
Verbs of high semantic generality		8.7	6.7	6.5	2.6		2.5			2.1	1.1	4.5	2.8		4.3		12.3
Overdoing explicimess			0.7									0.4	0.3				
Transitive verbs used as if intransitive				0.4	0.3					-							
Abbreviations		2.9	3.4	3	0.3	0.6	7.3	2.2				1.1	4.3		2.5		0.8
Replacement of -s with -z		20.3	18.8	8.2	6.1	2.3		8.8			4.4	9.4	17.4		2.9		16
'Time-stretching' with repeated letters	24.1	14.5	12.1	13.7	16.9	41.9	2.5	4.4	52.1	16.1	11.1	17.3	9.2	4.4	17		42.7

Table 3: Feature densities compared per feature and country

Each value is a feature density, i.e. the number of feature occurrences from one nationality, divided by the number of tokens available in the corpus for that nationality. Since these values tend to be very small (each feature representing less than 1% of a country's texts), they have been further multiplied by 10,000 for easier readability and comparison.

Texts originating from unknown or multiple countries, and false positives for each feature, are excluded.

Country codes are those defined by ISO (2016).

# **4.2 ELF features**

## 4.2.1 Grammar at word level

#### 4.2.1.1 Third-person zero

Third-person -*s* is communicatively redundant in English and already commonly dropped in some varieties (Dewey, 2007, p. 138); it is not improbable that regularisation will eventually dispense with it altogether (Jenkins, 2013, p. 33). ELF exhibits patterns of dropping akin to those in World Englishes: for example, -*s* is likelier to survive in "prefabricated chunks" such as *it depends* (Breiteneder, 2009, pp. 259-260).

42 instances of third-person zero were found in the corpus, predominantly in Czech, Polish, and Ukrainian texts, and covering both animate and inanimate referents: *she shake her head*; *it take 10 years*. Since there were approximately 3,480 standard -*s*-inflected forms, this represents far less than the 16.5% found by Breiteneder in a spoken corpus (2009, p. 262). It is possible that the prevalence of NES texts in the demoscene (as illustrated in Table 1) has served as a checking influence and helped to preserve the native-like form.

#### 4.2.1.2 Insertion and omission of articles

This feature is overwhelmingly associated with Slavic texts (87 of 168 insertions and 204 of 347 omissions), presumably because of the lack of articles in these L1s:

- *what the machine is this?* (Belarusian)
- my the most favourite music group is the pink floyd (Czech)
- *we're still on* Ø *scene* (Ukrainian)
- please send Ø photo of Ø car along with Ø letter (Slovak)

Indefinite *a* before plurals is not uncommon, and there is some suggestion of *one* emerging as a novel indefinite article: *we are not a graphics makers* (Russian); *some special words to one girl i met* (Polish); *i downloaded one track by vojtech* (Slovak).

Where insertion occurs outside of Slavic texts, it is typically before a mass noun: *how about the teamwork again?* (Austrian); *do a research* (Greek). It would be a mistake, however, to view this as a conversion of mass noun into count noun: Dewey links it with "uncountable and abstract nouns where the reference is generic rather than specific" (2007, pp. 117-118), among them *marriage*, *society*, and *democracy* (which are countable when speaking of specific instances), and suggests that "keyness", i.e. importance of the topic, is a driving factor in article insertion (p. 156).

Dewey also found that NES and ELF speech use articles differently, and points out that articles can be communicatively redundant (2007, pp. 106-110). As with the Slavic examples above, Dewey's Korean and Japanese speakers have article-less L1s; but the prevalence of this feature and its communicative efficacy suggest it is something more significant than L1 transfer. Nevertheless, at least one Czech scener in the corpus expresses doubt: *do you think i often use*  $\emptyset$  *word "the"*?

# 4.2.2 Grammar at clause level

#### 4.2.2.1 Relative pronouns

Only two instances of exchange of relative *who* and *which* were found (out of around 440 possible sites, i.e. 0.05%): *everybody which i forgot* (Czech) and *i remember joystick who worked as any key ya want* (Polish). The sparsity suggests that these may indeed be atypical instances of "confusion" (to use Seidlhofer's term: 2004, p. 220), and also accords with the findings of Motschenbacher, whose ELF corpus of speech at a press conference had *who-which* exchange in only 2.1% of relative clauses, a "clearly marginal" proportion (2013, p. 167), leading him to conclude that it may be misguided to regard it as a typical feature. (Where it did occur, there were no comprehension problems.)

Another pairing of relative pronouns is *that* and *which*, whose usage, however, is not predicated on personhood of the antecedent, but on whether the relative clause is restrictive: thus, *the bottle that was empty* (i.e. not the full one) but *the bottle, which was empty* (new information). Texts in the corpus disregard this distinction when it comes to *which*:

- *a scanner/digitizer which can scan photos* (German)
- your screens are nice ... but your crack which you do, it's total horror (Czech)
- effects like multicolor which are depended on the cpu speed (Polish)

A single German example of *what* replacing *which* (*"to je jedno"*, *what has the same meaning in czech as "egal" in german*) can be disregarded as a 'false friend', since *was* is the German for both *what* and relative *which*.

In a frequency list for the BNC, which mainly comprises traditional writing, *that*, *which* and *who* rank 13th, 33rd, and 54th respectively (Leech, Rayson, & Wilson, 2001). In the demo corpus, they rank 14th, 105th, and 76th, i.e. *which* is appreciably less common, even allowing for these words not always being relative pronouns. This reflects the speech-like nature of CMC:

[T]he proportion of *thats* to *whichs* is far higher in speech than in writing [because] the kind of clause properly begun with *which* is rare in speech with its short detached sentences, but very common in the more complex & continuous structure of writing. (Fowler, 1927/2002, p. 635)

#### 4.2.2.2 Tag question not matching main verb

Formulating tag questions can pose a challenge to NNES because English tag questions vary according to the preceding verb (e.g. *you do, don't you?* but *you will, won't you?*), whereas numerous European languages have invariant tag questions: German *nicht war? oder?*; Russian да?; French *n'est-ce pas?*, and so on. The invariant *innit*, found in certain varieties of English (Martínez, 2015), is generally viewed as a low-prestige or non-standard form and would therefore not be encountered in the classroom.

ELF speakers frequently use tag questions that do not obey the rule, either because the verb is an explicit mismatch (e.g. *you know him, isn't it?*) or because it is bypassed entirely, as with *or*? or *no*? (Seidlhofer, 2004).

74 tag questions were found in the corpus, but virtually all showed agreement of auxiliary with main verb, though one Ukrainian text used *isn't it* invariantly: *not 3d-multicolour, but look pretty! isn't it?*. There were no occurrences of *or?*, and all 35 instances of *no?* were not tag questions (e.g. *do you remember my joke??? no??*): ten of them were in

British texts. However, tag questions are highly typical of interactive conversation, and often serve as prompts to an interlocutor (Svartvik & Leech, 2006, p. 204), which may explain their rarity in the non-interactive written form.

#### 4.2.2.3 Redundant or innovative preposition use

Preposition use is idiomatic (*in a meeting*, but *at a party*; *with help from*, but *with the aid of*) and can therefore be a difficult area for NNES. 14 redundant and 52 innovatively-used prepositions were found in the corpus, out of 1535 potential matches. Classification, however, is somewhat arbitrary, since for example *fighting with abortion* (Polish) could be either innovative (for *fighting against*) or redundant.

The redundancies show no obvious pattern: e.g. *the town was re-named in chemnitz* (German); *if I forget with hello for somebody* (Czech); *a new members to join with us* (Lithuanian). Dewey (2007, p. 103) proposes that the apparently redundant *discuss about* may be an extension of *discussion about*; this is the case with *I enjoy of amiga's music* (Belarusian), where *enjoyment of music* would be unexceptional.

In the innovatively-used group, *at* is underutilised, often turning into *in* or *on*, e.g. *effect in a top of screen* (Ukrainian); *started to work in computer company* (Czech); *screens which were digitized on party* (Polish). While there is (limited) potential for confusion here – *in, on, at the house* are three different situations in NES English – none of the matched cases are ambiguous.

On occasionally indicates software use, where NES would tend to use *in*: hence, compiled on hisoft basic (Austrian); made on sample-tracker program (Polish); compose tunes on soundtracker (Polish). This could be seen as regularisation with on used for hardware, as in stereograms were designed on pc and vicok learned to code on amiga.

Significantly more instances of the feature might have been found had the search pattern permitted intervening determiners and adjectives, as in *bfox is happy 2 see u on this nice demoparty* (Russian); however, inaccuracies in automated tagging made this impractical. Omitted prepositions are also a known ELF feature (Dewey, 2007), but an automated search was infeasible since many common verbs can be both transitive and intransitive.

#### 4.2.2.4 Replacing infinitive-constructions with 'that' clauses

Only two instances were found: *we planned that current compilation will be presents* [i.e. presented], from a Russian text, and *all politicians want that berlin gets the olympic games* (German). Both L1s do in fact construct such expressions in this way: *I want you to stay* = *ich will, dass du bleibst* = я хочу что бы ты остался [*I want that SUBJUNCTIVE-PARTICLE you stayed*]; so these isolated cases can be plausibly regarded as L1 transfer.

Though it may sound alien to NES, there is no reason why this feature should cause communication failure, and it serves to regularise the kind of inconsistency in English whereby *I hope, believe, expect that you know* are acceptable but \**I want, need, intend that you know* are not. Such cases represent large sets of historical rules that must be learned but add no particular value to the language.

A possible additional benefit for NNES is that *that*-clauses are easy to construct, involving the subordination of a simple sentence, whereas the infinitive cannot stand alone and is therefore less frequently encountered.

#### 4.2.2.5 Innovative word order with adverbials

45 instances were found, from at least eight countries. 11 were of the clause-initial kind, e.g. *simply we can swop these programs* (Belarusian); *probably you know* (Czech); and the rest were clause-medial, e.g. *we can mention also a few freaks*; *i've seen lately a very good demo* (both Polish). Potential communication failures can be mitigated by context:

- *a representative... will receive the prizes if the work will win a prize. in other case the author of the work will receive only the diploma by mail.* (Russian: i.e. a certificate by mail and nothing else)
- you don't know what you don't know, only you know what you know (Polish)

(It is worth noting here that scrolltexts are not purely 'read-only' but do often serve as part of an epistolary communication flow, with sceners exchanging greetings, morsels of news, and plans for collaboration in alternate texts. The researcher found virtually no evidence of problems with understanding: practically the only obvious communication failure was a cultural one, regarding a joky remark taken the wrong way.) Some of the adverbial innovations are only arguably 'innovative': if *now you're listening to f-16 fighting simulator music* (Polish) is somehow unidiomatic (one would expect *you're now listening*), then so is *already he is staying up late at nights*, yet the latter occurs in a British text.

45 out of 33,896 adverbs is 0.01%, making the feature rare. However, the search did not comprehend clause-final adverbials. Also, under a transfer hypothesis, L1 is a factor: Dewey (2007, p. 123) gives *I like very much playing golf* as a pattern typical of Latinate languages, which are virtually absent from the largely Eastern-European demo corpus.

# 4.2.3 Semantic phenomena

#### 4.2.3.1 Frequent use of verbs of high semantic generality

194 instances of this feature were found, from an approximate total of 73,500 verbs. They spanned at least 11 countries, though Ukraine was particular prominent, and were mainly *make* (94) and *do* (43), which overlap in many of the corpus L1s (e.g. Czech *udělat*, Russian делать) and even in NES usage have more flexible roles than often imagined (Dewey, 2007, pp. 150-151). *Make* was mainly used in a vague sense of creation, e.g. making code or a magazine, but also more broadly: *make a little party* (Polish); *please made hello to our group* (Ukrainian); *i made great races through the whole towm* [sic] (German).

The transparent or delexicalised nature of the semantically general verbs allows one to replace another almost arbitrarily: *i had* [i.e. made] *some attempts at writing* (German); *where you take* [get] *so nice sample??* (Czech); *took* [made] *the heavy effort* (Polish). General verbs are also often used with a noun that could itself have served as the verb: *doing nostalgic tries to load old games* (Czech), *do a research about the atari st computer* (Greek); such clauses are based around important nouns with the verb serving as connective 'glue', and it is conceivable that this supports NNES vocabularies developed informally by acquiring the 'names for things' (e.g. at demoparties) rather than through formal instruction.

#### 4.2.3.2 Overdoing explicitness

Overdoing explicitness can help to reinforce clarity (Vettorel, 2014, p. 149). It is not tied to any particular grammatical form and may span clauses, as in the Japanese emergency warning "When directions are given by staff members during evacuation, please follow the directions" (Backhaus, 2015, p. 202). The adjective-noun search strategy outlined in Table 2 is therefore undesirably narrow; but it would be impossible to identify arbitrary cases of semantic redundancy using only grammatical tagging.

12 matches were found, and the added explicitness can usually be regarded as resolving a potential ambiguity. Five cases name a language, e.g. *English language* where *English* would suffice; but all of the languages are also nationalities, and *all people know English* might otherwise be misinterpreted in ELF as 'everybody knows some English people'. *Acoustic piano* distinguishes the instrument from a synthesiser. *Compress packers* (data compression programs) are usually simply *packers*, but perhaps the writer felt this bore clarifying as relatively rare technical jargon (9 occurrences in corpus).

*Enough number* and *most favourite*, while not native-sounding, are semantically no more problematic than the idiomatic *sufficient number* and *most perfect*, so their presence in this category is arguable. Similarly, *great piece of good work* could be viewed as an intensifier structure. *Pink colour* parallels Seidlhofer's (2004) example of *black colour*. Possibly the redundant noun in such cases serves as a syntactic 'anchor' for the adjective, making it easier to form a typical sentence with subject and object.

*Your 18-years birthday* represents an overexplicitness rather characteristic of the Eastern European portion of the corpus: that of number and year. Other examples observed, but not matched by the algorithm (since numbers were not tagged as adjectives), include *founded in year 1992, beginning of 1999 year, intro of 1995 year*, and the demo title *Happy New 1997 Year*.

#### 4.2.3.3 Transitive verbs used as if intransitive

Dewey (2007, p. 91) found numerous instances of this feature in his NNES conversation data, but in the demo corpus only six instances were found, all in situations that would usually have a pronominal object. While there might be some reason why this feature is more common in speech (e.g. inability to go back and edit), more probably the clause-finality search condition was too restrictive: the researcher observed such missed examples as we'd like to tell about ourselves (Belarusian) and i wanna say about effect (Ukrainian).

Some potential examples of the opposite case, i.e. intransitive verbs used transitively, were also observed, e.g. *if u don't wanna listen it* (Polish); *look next part; wait new demo* (both Russian). Such cases, however, appear to be simplifications of phrasal verbs, which are notoriously difficult to learn (Mart, 2012), and so might be looked upon as a radical form of regularisation: that is, dropping various 'default' prepositions (*listen to, look at, wait for*) that are so strongly correlated with those verbs as hardly to obscure the meaning by their absence. (Less common scenarios, such as listening *for* or looking *under*, would be far less amenable to preposition-dropping.) This hypothesis also explains usages such as *i say you "good-bye"* (Belarusian) and *we present u first part* (Ukrainian) that would otherwise have to be regarded as ditransitive.

From a cross-linguistic perspective, the transitivity of a clause is properly a matter of degree, i.e. gradable rather than binary (Næss, 2007, pp. 22-24). For this reason, and particularly where the object is a pronoun denoting a contextually recent referent, strictly following the transitivity conventions for verbs is not always necessary for successful communication.

#### 4.2.3.4 Lexical innovation

The 50 most frequent innovative lexical items from the corpus are listed in Appendix D. Other than the names of specific hardware and software, they comprise nine abbreviations, nine sensational spellings, six speech sounds, and six items of general demoscene jargon.

Far from being restricted to a tiny subset of the language as helpless 'learners', NNES do not hesitate to experiment creatively with English (Poppi, 2013, p. 58), and this is evident from the less common innovations in the corpus. Neologisms may fill a lexical gap (*panny*: a panning effect; *outro*: a closing part; *muzaker*: an inferior musician) or lend an air of informality or humour (*collie*: collection; *messy*: message; and *drinkware* for alcoholic drinks, listed alongside the software and hardware used in producing a demo). Words may be formed by derivation (*decrunch*: to decompress data), conversion (*disking games*, i.e. modifying cassette versions to run from disk), or borrowing (*tacts/tackts*: processor cycles, from Russian такт, or Czech and Polish *takt*).

22

This is not to say that all atypical usages are innovation; some are evidently learner errors or false friends, as with 'demone'... stands shortly for demo one (Austrian; i.e. abbreviatedly), or every point has proper colour (Czech; vlastní means both 'own' and 'proper'). Nevertheless, a widespread anomaly may gradually become a norm within the community of practice: graphician (graphic artist), found in ten texts from five countries, is more likely to have spread between sceners than to have been coined independently in every case; and the adjective unrealizable, highly uncommon in English, was used by German, Polish, and Slovak sceners before making its appearance in five British texts.

# 4.3 Netspeak features

# 4.3.1 Abbreviations

14 of Crystal's 103 netspeak abbreviations were found in the corpus. Some were uncommon: *asap*, *b4*, *lol*, *ruok*, and *ttfn* occurred once each. As Crystal notes (2006, p. 92), not all abbreviations occur in every online context; however, one or two chat abbreviations did occur in scrolltexts, as in *beer is calling*, *brb* [be right back].

*Thx* occurred 30 times (from at least three countries) versus *tnx*'s one, suggesting that *thx* won out as the notation for *thanks* in the community. The most common abbreviation was *btw*, found 37 times from at least seven countries. Some abbreviations were characteristic of particular writers, as with the one British scener who used *cu l8r m8* in multiple texts. *Imho*, though common on the English-speaking Internet, occurred four times and only in texts from the former USSR, which might suggest local spreading.

Some further abbreviations, such as fx for *effects* and *FLP* for *flexible line position* (a kind of animated text effect), can be found among the innovative lexical items listed in Appendix D.

# 4.3.2 Capitalisation

Capitalisation on the Internet is often absent or arbitrary, with a "lower-case default mentality" (Crystal, 2012, pp. 90-92). The demo corpus does not match the Internet in the latter respect. 276 texts had mainly lower-case tokens while 450 had mainly upper-case tokens; furthermore, the lower-case texts generally used standard capitalisation (first letter of sentence, etc.) rather than omitting it altogether, and only seven texts were in unmixed lower case.

One reason for the prevalence of upper case might simply be the technological era: early software often displayed messages in capitals, and some computers had no lower case at all (Mace, 1983). Also, scrolltexts are usually presented in large and attractive fonts, which (due to memory constraints) are sometimes only designed in upper case, seen as more important in the era before texting and the Web. Custom fonts are also used to create special characters, such as Cyrillic letters or graphical icons, and the texts sometimes refer to these:

- the logos are by me, and the font is the same one used in the intro (British)
- *KAZ CODED: [], [], [], AND MAT CODED: [], []* (Polish; the boxes display animated miniatures of earlier visual effects)
- PLEASE SORRY ME FOR BAD ENGLISH, BUT ... I DON'T HAVE MEMORY FOR RUSSIAN FONT (Russian)

In a few cases, two authors write a text collaboratively, one using lower case and one using upper case (mapped to a second font on the display), so that the reader can follow turn-taking in the discourse: *orson started out going to draw some crappy little egg type gay guy!!! I DID NOT!!! liar!!!* 

Finally, a chaotic alternation of cases is occasionally seen, particularly in Russian texts (possibly because the Latin alphabet is less of a fixed entity to Cyrillic eyes, and more amenable to visual play?): *Make DemO NoT vIruS!!!!!*; *We aLSo reTUrnEd AnD AliVe*; *LaSt pAgE iS cOmMiNg...* This deliberately unconventional style is the typographical counterpart of the sensational spellings discussed next.

## 4.3.3 Non-standard spelling

Deliberately non-standard, or sensational, spelling is acceptable in conversational netspeak (Crystal, 2006, p. 93) and is also a feature of underground countercultures such as hip-hop and graffiti, which bear parallels to the demoscene (Czerski, 2014): in both cases, oppositional digression from orthographic norms establishes a distinct group identity (Jaffe, 2000) and enables creative play with language.

Crystal's example of final *z* for *s* is widely evidenced in the corpus, from at least 11 countries, especially the Russian-speaking Belarus and Russia, and sometimes with atypically long or formal words (*camcorderz, insectz, organizerz*). Particularly common are *greetz/greetingz* (and variations: 129 occurrences) and *guyz* (38). *z* may also replace *s* in situations other than verb and noun endings, and for non-/z/ pronunciations (*thiz, dizlike, tremendouz*), suggesting that it has lost its phonological basis and become a visual flourish. *Thiz* even occurs in one British text. Analogous consonant changes include those in *skool, freax, thanx, greetinx, disx (disks)*, and *muzzax (music)*, while vocalic eye-dialect is evident in *yer (your)* and *kewl (cool)*.

Non-standard spelling sometimes reflects the speaker's L1 phonology, particularly with the /i:/ vowel and the relatively rare fricatives / $\theta$ / and / $\delta$ /: hence *gritinks* (*greetings*; Austrian), *birsday* and *ozers* (*birthday* and *others*; Russian), *reliz* (*release*), *steel* (*still*), and *pipl* (*people*). On occasion, this is done with humorous intent, as in the Polish text that caricatures a drinking session: *after sekond botyl hi waz qlajt drunk and lejing ander ze tejbl*.... [After second bottle he was quite drunk and laying under the table...]

# 4.3.4 'Time-stretching' of syllables with repeated letters

Repeating letters in a word is one of the forms of exaggerated spelling and punctuation that mimic prosody and paralanguage in online conversations (Crystal, 2006). This feature is apparently universal in the demo corpus, occurring 1,572 times in scrolltexts from at least 16 countries. Table 3 shows markedly higher feature densities for Britain and Ukraine, but these are the result of two or three particularly prolific individuals who favour the style.

Describing the feature as time-stretching of syllables proved to be inaccurate, since the lengthened portions often fail to correspond to the (typically vocalic) sites of lengthening in emphatic speech: *errorrrrrrr* (Austrian), *rulezzz* (Czech), *wwwwaaaaakkkkkeeeee up*  (German). In this respect it differs from Werry's conception of Internet chat, where reduplicated letters are said to be directly indicative of "drawn-out or expressive intonation ... to be read with the simultaneous involvement of the ear and eye" (1996, pp. 57-59); rather, in the demoscene, this typing affectation serves as an anarchic way to convey 'manic' states of enthusiasm and rage:

- now we is stoooooaaaaannnneeeed as a dog (British)
- *i would have to fill up 1000000 bytes with text. aaaaarrrrrrgghhhhhh....* (German)
- *uffffffffff!!! stop this english!* (Slovak)
- such a brilliant piece of music! yeaaaaaaaaaa! (Ukrainian)
- *aaaamiiigaaa is the beeeessstttt!!!!* (Polish)

Nevertheless, Werry's observations regarding the tendency toward verbal play are of at least some relevance: the newness of the medium and consequent freedom from traditional contexts apply to demos as much as to online chat, though anonymity in the demoscene tends to be partial (because of the postal and electronic distribution network by which demos are shared) and texts are only transitory insofar as they scroll on and off the screen: software programs, unlike chat-room conversations, are recorded works that can be replayed at will.

## 4.4 Other observations

## 4.4.1 Regularisation of noun and verb forms

Jenkins (2013, p. 33) observes that "ELF speakers are speeding up the regularization of English that is already in progress and always has been". Some points involving grammatical regularisation in the demo corpus have already been discussed (4.2.1.1, 4.2.2.3-4, 4.2.3.3), but there are further areas where regularisation is particularly apparent, involving verb and noun inflections.

Regularised past forms of verbs in the corpus include *beated*, *catched*, *choosed*, *cutted*, *drawed*, *drinked*, *heared*, *leaded*, *maked*, *rised*, and *spended*. Nouns are regularised in various ways: sometimes by back-formation from a plural-sounding singular or mass noun: *the first big thank* (Czech); *the last demo of this serie* (German); ...*to tell you some news*... *one of 'em is bad* (Belarusian); and sometimes by adapting an irregular plural to take regular - *s: hello mens* (Polish); *a few mices* (Ukrainian).

However, the most noticeable regularisation of nouns is 'countification', resulting in forms such as *advices*, *homeworks*, and *informations*. This process has been remarked upon by Jenkins (2012, p. 489), not as an ELF-specific feature but as one that also occurs in World Englishes and even native varieties; and the demo corpus does indeed include a small number of British texts using unlikely plurals of mass nouns:

- e-3 gave us some codes [i.e. program code], and i've sort of, "customised" them a bit
- *i can always attempt musics if the need arises*
- we hope you enjoy the music's!

# 4.4.2 Code-switching and code-mixing

Cheshire (2002, p. 32) observes that young Europeans frequently adapt English into their own languages: this is an additive process rather than a replacement, and is associated with the expression of personal identity. Unsurprisingly, then, given the international context and group affiliations of the demoscene, 170 texts (23.4%) contained code-switching or mixing. Where predicted L1s were available, these were mostly Polish (28), Czech (16), and Russian (16), which approximately reflects the total L1 distribution.

English in demos is often used for conventional openings and/or closings. For example, one Russian text begins *hello everybody!!!* but switches to Russian after eight short sentences; another text is entirely in Czech until it finishes with *this is all. the end. bye-bye*. (A very few demos offer parallel bilingual texts, but this is a rare luxury due to memory constraints and translation effort.) An impending switch is sometimes signalled in the text:

- so, let's speak on russian? yes! da!
- now take your dictionary because now comes czek language!

The ubiquitous greetings list is another area that may employ English. Typically, people from the author's home country are greeted in the common L1, and those from other countries in English. (Sebba [2012] makes a similar observation regarding multilingual speech: changes of addressee may result in a change of code.) One Russian text briefly

switches into English only to say *hello to all spectrum users*????, while other NNES explicitly insist on English as the language of communication, at least in certain contexts:

- english contacts only in writing! (Czech)
- *if you are not russian write me in english*
- why are most of the demos, that come from other countries not written in english? i think it is much easier to understand the messages, when you write your texts in english. (German)
- in the first part i used english anywhere i was sure that greeted person will understand me, otherwise i used czech. in the second part i used czech if the greeted was czech or slovak, if he was member of other nation, used english.

It appears, therefore, that English serves as a de-facto international language for communicating with the demoscene in general – but without precluding the concomitant use of other languages. While this special status of English might be disparaged as linguistic imperialism (Phillipson, 1992), the ability to mix and switch makes it decidedly less of a barrier than in wider sociopolitics. For example, the United Nations has almost 200 member states but only six 'official' languages, while, in the European Union, certain member languages lack the status accorded to others (and, consequently, their users must work harder to make their voices heard):

[S]ome regional languages, such as Catalan and Welsh, have gained a status as *co-official* languages of the European Union. The official use of such languages *can be authorised on the basis of an administrative arrangement* concluded between the Council and the requesting Member State. (European Commission, 2006; my italics)

Apart from switching, code-mixing (i.e. within individual phrases) also occurs, and – since European NNES are more likely to be multilingual than the British (Piller, 2012) – this actually puts the NES at a communicative disadvantage:

- *co z* [what about] *vectors graphics for my next demo?* (Polish)
- this is prawie [almost] the best like dire straits...and sandra is nawet [even] better...
  (Polish)
- жела́ю тебе [I hope you] to pass your final exams successfully! (Russian)

# 4.4.3 Attitudes to English

A comprehensive analysis of NNES sceners' attitudes to English is beyond the scope of this study: that would best be served by an ethnographic or interview-based approach, and at the very least would require the retention and translation of those texts that were discarded as containing little or no English. However, a simple search for the word *English* uncovers a range of attitudes that, in the context of the debate over ELF fluency versus learner error, it would be imprudent to overlook.

The English language is a recurring subject in scrolltexts, whether the writer is discussing issues of acceptability, criticising the perceived need for English, or simply explaining why English is – or is not – being used. These are only a few examples:

- *i also would like to know if you want me to write in slang in my intro texts*
- don't tell me about my odd english. i just can't image how englishmen can speak it?!
- profsoft has told me i must write in english because this demo will be offered to your sinclair or sinclair user [British magazines]
- sorry! i don't now english and i like better write polish.
- my english isn't as perfect as yours. it was the first reason [for not using English] and the second is in history it was russian language and now it should be english? no!!!!
- *i'm* sorry that there's no english text in 3d gallery part but i found it difficult to write the tutor text in english (my knowledge of english isn't good enough to write such a difficult text).
- *don't expect also, that all people* [at a forthcoming demoparty] *know english language. we don't want you would feel disappointed.*

There are, furthermore, around 40 texts in which the writer apologises for bad English and/or characterises himself as a deficient speaker, and many more that plead learner status as the reason for any mistakes.

All of this leads back to the thorny issue identified by Maley (2009): how – and why – are we to offer ELF as a "pedagogical alternative" (Jenkins, 2007, p. xii) when NNES tend to self-identify as learners and prefer to learn a standard variety (Groom, 2012), and while the ELF debate remains an academic abstraction of which most teachers are "sublimely unaware" (Maley, 2009, pp. 194-196)? Maley's suggestion of concentrating on existing World Englishes is unlikely to make sense in Europe, since – as with ELF – "legitimation,

codification and standardization [of Euro-English] is proving to be a challenging endeavour" (Modiano, 2006, p. 229).

# **CHAPTER 5: CONCLUSION**

# **5.1 Summary and implications**

The demoscene corpus has provided a cross-section of a thriving, collaborative international community that uses non-standard English with considerable fluency and success, even for complex topics such as technical restrictions on contributing program code. While sceners are aware of the divergent features of their varieties, and often regard them as errors, the 'errors' prove to be little impediment in practice, and sometimes (as with the netspeak features) are actively maintained and propagated as part of co-constructing a group identity – to the extent that some non-standard items have been partially adopted by NES.

The analysis demonstrated the presence of both ELF and netspeak features, though not all of those features, nor only those features. Some, such as time-stretching of syllables and the use of verbs of high semantic generality, are more or less universal, while others, such as insertion and omission of articles, are correlated with the writer's L1; these latter cases, however, are mostly explainable as L1 transfer. A significant part of the non-standard usage represents regularisation of one form or another, albeit a hyper-accelerated and idiolectal style of regularisation rather than the ongoing, but glacially slow, regularisation of native Englishes.

Perhaps the key factor of both ELF and netspeak that surfaces in the demo corpus is *flexibility*: in ELF terms, a pragmatic willingness to adapt the language to serve the interests of clarity and simplicity, and in netspeak terms, an urge to do so out of a spirit of adventure and creativity. Failure (or refusal!) to obey lexicogrammatical norms is not a reasonable criterion for error in demoscene interactions, and indeed obeying them unswervingly might be perceived as an error of register, like using slang in a formal public speech. The only true standard of error in such interactions is implied by Dewey's final research question: "Which items in 'non-standard' L1 English lexicogrammar *lead to miscommunication*, and what *might thus constitute an error* in lingua franca usage?" (2007, p. 54; my italics). As for netspeak, even unintentional slips cause "little or no disruption" (Crystal, 2006, p. 116), and the "moral panic" around the supposed linguistic deterioration resulting from text-messaging has proven unjustified (Crystal, 2011, pp. 3-4).

31

Flexibility by its nature can hardly be codified – which makes questionable the extent to which ELF can ever be taught and examined, especially when individual ELF speakers are inconsistent in the structures they use (Dewey, 2007; Poppi, 2013) – but the rich success of the demoscene in communicating and collaborating, and its analogue in the modern 'global village' where so many Englishes are spoken, suggests that it may be of value to be less prescriptive and less proscriptive in teaching English to non-native speakers, and – while we will never eliminate lists of exceptions and Gradgrindian points of grammar – to place more emphasis on communicative strategies and outcomes.

# **5.2 Limitations**

This study involved less ELF data than originally envisioned. It was not possible to determine the originating country for almost 40% of texts, due to incomplete information on the national origins of many demogroups; in addition, more than 20% of the country-identified texts originated in Britain, and these were on average more than twice as long as the others.

The feature density metric did not work well for very small subcorpora. For example, the densities in Table 3 for Argentina, Latvia, and Turkey appear artificially high, simply because there are very few texts from those countries. This might have been ameliorated by the use of inverse document frequency, a.k.a. term specificity (Spärck Jones, 1972); similar techniques could have brought to the fore any bias resulting from particular individuals producing atypically large amounts of text.

Automated part-of-speech tagging achieved low accuracy in some cases, identified as a risk in section 3.3; this was mitigated, however, by the manual filtering of results to remove false positives, though the latter were numerous and filtering was therefore time-consuming.

While nationality could be discussed to some extent, it was not possible to compare other demographics, since such details are rarely evident from scrolltexts and in any case there is likely to be little variation: Czerski (2014) suggests an almost exclusively male demoscene aged 14 to 24, which largely coincides with the observations of Montfort et al. (2012, p. 212) on the demographics of early programming hobbyists.

[9,700 words]

# LIST OF REFERENCES

- Anthony, L. (2014). AntConc (Version 3.4.3) and TagAnt (Version 1.2.0). Tokyo, Japan: Waseda University. Retrieved from http://www.laurenceanthony.net/
- Backhaus, P. (2015). Attention, please! A linguistic soundscape/landscape analysis of ELF information provision in public transport in Tokyo. In K. Murata (Ed.), *Exploring ELF in Japanese academic and business contexts: Conceptualisation, research and pedagogic implications* (pp. 194-209). Abingdon: Routledge.
- Baker, P. (2010). Sociolinguistics and corpus linguistics. Edinburgh: Edinburgh University Press.
- Bell, J. (2010). Doing your research project: A guide for first-time researchers in education, health and social science. Maidenhead: Open University Press.
- Breiteneder, A. (2009). English as a lingua franca in Europe: An empirical perspective. *World Englishes*, 28(2), 235-269.
- British Council. (2016, January 24). *Verbs followed by to + infinitive*. Retrieved from British Council: https://learnenglish.britishcouncil.org/en/english-grammar/verbs/verbsfollowed-infinitive
- Bruckman, A. (2002, April 4). *Ethical guidelines for research online*. Retrieved from http://www.cc.gatech.edu/~asb/ethics/
- Butterick, M. (2014). *All caps: Fine for one line or less*. Retrieved from Practical Typography: http://practicaltypography.com/all-caps.html
- Cheshire, J. (2002). Who we are and where we're going: Language and identities in the New Europe. In P. Gubbins, & M. Holt (Eds.), *Beyond boundaries: Language and identity in contemporary Europe* (pp. 19-34). Buffalo, NY, Clevedon: Multilingual Matters.
- Cogo, A., & Dewey, M. (2012). *Analysing English as a lingua franca: A corpus-driven investigation*. London: Continuum.
- *comp.sys.sinclair FAQ: File formats*. (2005). Retrieved from http://www.worldofspectrum.org/faq/reference/formats.htm

- Crystal, D. (2006). *Language and the Internet* (2nd ed.). Cambridge: Cambridge University Press.
- Crystal, D. (2011). Internet linguistics: A student guide. New York: Routledge.
- Crystal, D. (2012). English as a global language. Cambridge: Cambridge University Press.
- Czerski, P. (2014). Maszyny, które mogą wszystko. Ha!art(47), pp. 10-24.
- Demozoo. (n.d.). Retrieved from http://demozoo.org/
- Dewey, M. (2007). English as a lingua franca: An empirical study of innovation in lexis and grammar (Unpublished PhD thesis). King's College, London.
- European Commission. (2016, February 26). *Official languages of the EU*. Retrieved from European Commission: http://ec.europa.eu/
- Fowler, H. W. (1927/2002). A dictionary of modern English usage (1st ed.). Oxford: Oxford University Press.
- Gilquin, G. (2010). *Corpus, cognition and causative constructions*. Amsterdam: John Benjamins.
- Groom, C. (2012). Non-native attitudes towards teaching English as a lingua franca in Europe. *English Today*, 28(1), 50-57.
- Gubbins, P., & Holt, M. (Eds.). (2002). *Beyond boundaries: Language and identity in contemporary Europe*. Buffalo, NY, Clevedon: Multilingual Matters.
- HarperCollins. (2016, January 25). Collins Official Scrabble Words. Retrieved from HarperCollins UK: http://www.harpercollins.co.uk/9780007589166/collins-officialscrabble-words
- Herring, S. C. (Ed.). (1996). *Computer-mediated communication: Linguistic, social and cross-cultural perspectives*. Amsterdam: John Benjamins.
- House, J. (2012). English as a lingua franca and linguistic diversity. *Journal of English as a Lingua Franca, 1*(1), 173-175.
- Hunston, S. (2002). Corpora in applied linguistics. Cambridge: Cambridge University Press.
- ISO. (2016, January 23). *Country codes ISO 3166*. Retrieved from ISO: http://www.iso.org/iso/country\_codes

- Jaffe, A. (2000). Introduction: Non-standard orthography and non-standard speech. *Journal* of Sociolinguistics, 4(4), 497-513.
- Jenkins, J. (2000). *The phonology of English as an international language*. Oxford: Oxford University Press.
- Jenkins, J. (2007). *English as a lingua franca: Attitude and identity*. Oxford: Oxford University Press.
- Jenkins, J. (2012). English as a lingua franca from the classroom to the classroom. *ELT Journal*, 66(4), 486-494.
- Jenkins, J. (2013). English as a lingua franca in the international university: The politics of academic English language policy. London & New York: Routledge.
- Kachru, B. B., Kachru, Y., & Nelson, C. L. (Eds.). (2006). *The handbook of World Englishes*. Oxford: Blackwell.
- Kelion, L. (2014, December 5). Syntax era: Sir Clive Sinclair's ZX Spectrum revolution. Retrieved from BBC News: http://www.bbc.co.uk/news/technology-30333671
- Kollock, P., & Smith, M. (1996). Managing the virtual commons: Cooperation and conflict in computer communities. In S. C. Herring (Ed.), *Computer-mediated communication: linguistic, social and cross-cultural perspectives* (pp. 109-128). Amsterdam: John Benjamins.
- Korenman, J., & Wyatt, N. (1996). Group dynamics in an e-mail forum. In S. C. Herring (Ed.), Computer-mediated communication: linguistic, social and cross-cultural perspectives (pp. 225-242). Amsterdam: John Benjamins.
- Lean, T. (2013). Mediating the microcomputer: The educational character of the 1980s British popular computing boom. *Public Understanding of Science*, 22(5).
- Leech, G., Rayson, P., & Wilson, A. (2001). Word frequencies in written and spoken English: based on the British National Corpus. Retrieved from http://ucrel.lancs.ac.uk/bncfreq/flists.html
- Linguasorb. (2016, January 24). *100 most common English verbs list*. Retrieved from Linguasorb: http://www.linguasorb.com/english/verbs/most-common-verbs/
- Mace, S. (1983). British Dragon pursues home computer market. InfoWorld, 5(48), p. 24.

Mair, C. (Ed.). (2003). The politics of English as a world language. Amsterdam: Rodopi.

- Maley, A. (2009). ELF: A teacher's perspective. *Language and Intercultural Communication*, 9(3), 187-200.
- Marecki, P. (2015). *Textual demoscene*. Retrieved from http://nickm.com/trope\_tank/TROPE-15-01.pdf
- Mart, C. T. (2012). How to teach phrasal verbs. *English Language Teaching*, 5(6), 114-118.
- Martínez, I. P. (2015). Variation, development and pragmatic uses of innit in the language of British adults and teenagers. *English Language and Linguistics*, *19*(3), 383-405.
- Modiano, M. (2006). Euro-Englishes. In B. B. Kachru, Y. Kachru, & C. L. Nelson (Eds.), *The handbook of World Englishes* (pp. 223-239). Oxford: Blackwell.
- Modiano, M. (2009). Inclusive/exclusive? English as a lingua franca in the European Union. *World Englishes*, 28(2), 208-223.
- Montfort, N. (2013). *No code: Null programs*. Retrieved from http://nickm.com/trope\_tank/TROPE-13-03.pdf
- Montfort, N., Baudoin, P., Bell, J., Bogost, I., Douglass, J., Marino, M. C., . . . Vawter, N. (2012, November). 10 PRINT CHR\$(205.5+RND(1)); : GOTO 10. Retrieved from http://nickm.com/trope\_tank/10\_PRINT\_121114.pdf
- Motschenbacher, H. (2013). *New perspectives on English as a European lingua franca*. Amsterdam: John Benjamins.
- Murata, K. (Ed.). (2015). *Exploring ELF in Japanese academic and business contexts: Conceptualisation, research and pedagogic implications.* Abingdon: Routledge.
- Næss, Å. (2007). Prototypical transitivity. Amsterdam: John Benjamins.
- Needle, J. (2015). *Spectaculator, ZX Spectrum emulator*. Retrieved from http://www.spectaculator.com/
- Orlowski, A. (2010, June 21). *Twitter on a ZX Spectrum*. Retrieved from The Register: http://www.theregister.co.uk/2010/06/21/vintage\_computer\_fair/
- Pemberton, L., & Shurville, S. (Eds.). (2000). Words on the Web: Computer mediated communication. Exeter: Intellect.

- Pennycook, A. (2003). Beyond homogeny and heterogeny: English as a global and worldly language. In C. Mair (Ed.), *The politics of English as a world language* (pp. 3-18). Amsterdam: Rodopi.
- Phillipson, R. (1992). Linguistic imperialism. Oxford: Oxford University Press.
- Pillar, J. (1992, October). Shock (PD). Your Sinclair, p. 40.
- Piller, I. (2012, July 18). *Multilingual Europe*. Retrieved from Language on the Move: http://www.languageonthemove.com/multilingual-europe/
- Poppi, F. (2013). *Global interactions in English as a lingua franca: how written communication is changing*. Oxford: Peter Lang.
- Schmid, H. (1994). Probabilistic part-of-speech tagging using decision trees. *Proceedings of the International Conference on New Methods in Language Processing*, (pp. 44-49).
   Manchester, UK.
- Sebba, M. (2012). Researching and theorising multilingual texts. In M. Sebba, S. Mahootian,
  & C. Jonsson (Eds.), *Language mixing and code-switching in writing: Approaches to mixed-language written discourse* (pp. 1-26). London: Routledge.
- Sebba, M., Mahootian, S., & Jonsson, C. (Eds.). (2012). Language mixing and codeswitching in writing: Approaches to mixed-language written discourse. London: Routledge.
- Seidlhofer, B. (2004). Research perspectives on teaching English as a lingua franca. *Annual Review of Applied Linguistics*, 24, 209-239.
- Sondheim, A. (2001). Codework. American Book Review, 22(6), 3-4.
- Spärck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, 28, 11-21.
- Stachniak, Z. (2015). Red clones: The Soviet computer hobby movement of the 1980s. *IEEE* Annals of the History of Computing, 37(1), 12-23.
- Svartvik, J., & Leech, G. (2006). *English: One tongue, many voices*. Houndmills: Palgrave Macmillan.
- Vettorel, P. (2014). *English as a lingua franca in wider networking: Blogging practices*. Berlin/Boston: De Gruyter.

- Werry, C. (1996). Linguistic and interactional features of Internet Relay Chat. In S. C. Herring (Ed.), *Computer-mediated communication: linguistic, social and crosscultural perspectives* (pp. 47-63). Amsterdam: John Benjamins.
- White, M. (2006). *The body and the screen: Theories of Internet spectatorship*. Cambridge, MA: MIT Press.

# **APPENDICES**

## **Appendix A: Some representative demo texts**

These opening portions of Spectrum demo scrolltexts come from various countries and serve to give a flavour of the style and content of such texts. They have been converted to lower case for readability (Butterick, 2014).

heeeeeey!!! we are here again. are you wonder what is up above the scroll ? this gratful effect was called "belt" it recolect "plasma war" from c-64 a bit , doesn't it ? we think so. we want to tell you something about our group beginining... at the very begining was chaos... next ziutek organized a party and rainbow dreams appeared (from chaos of course). to this secret organisation belonged mat (coder),kaz (coder),ziutek (graphic,musician). then the second party was and e.s.i. was founded (this time rackne (an commodore user) joined the group). in not a long time team got a new member - muad'dib...

#### [Poland: Ethanol Software Inc, The Lyra II]

yeeep! u succesfully start thiz demo which called me as ascent, 'n already c previous preintro !.. how ya like it ?..ha ?.. by the way, in preintro u may find smth word, recruit all letterz of which in smth part, u go out 2 hidden ...(think about thiz)... well, in ascent only one (!) part: ...classik 3d-reality, computed in realtime wiz a few control keyz !.. otherz are preintro, intro, notice, strictly part, epilogue, post script, digital end ....that'z all.... in thiz intro u may touch keyz: 8, 9, 7, 6 'n of coz space - exit 2 next notice ... all kodez, grafix in thiz demo waz done by s.system of enigma group 'n all muzak waz written by mirage of <eg> too !!

#### [Ukraine: S.System / Enigma, Ascent]

hello friends... terminator presents you:....7up 1.... sampling by rec in 26th march of 1991 in buenos aires - argentina. national greetings to:turbo soft(live) - piluso(nothing demo) - rajsoft - dack - hacker chris - mac - the alien gilbert- mhc(perete) - diego - niki(cara con palanca). if you want swop soft for spectrum write to: [name and address redacted] buenos aires-argentina. other sampling for my colection. end of scroll!!!

[Argentina: Terminator, 7 UP 1]

hallo and welcome in the demo-lition by tiger's claw. this demo is called so because here is a demolition effect in part 2. (part one has greetz, ifos and music. part two is the meteor part. in part 3 you can read something about my new game.), but firstly i want to say something very important. this demo is writen in basic and compiled with hisoft's compiler, only the scroll routine is in assembler (a simple ldir) the garfield pic is ripped from sam, because i'm working currently on a spectrum game, which needs my full time to create some graphics.

## [Austria: Tiger's Claw, Demo-lition]

do you see the ball in upper third? textured ball! it more coolest than ballscroller you see ... part info : ba ... stop !!! ball can jump on flp !!! o-o-ops! continue >> ball size 56 pix, scroll eat 7000 tackts, ball mapping - approx. 57000 tackts, music player - 4500 tackts and 1000 tackts free time. sorry, but this effect available only on pentagon 128. we wanna say - our demo reached to end. we would be proud to present it to enlight'96. we work on this demo not so long time - 1.5 month of non-intensive coding.

[Russia: X-Trade, *Illusion*]

# **Appendix B: Custom software tools for text extraction**

A number of software tools were developed and used in the text extraction process summarised in the methodology. Source code for these programs, written in the C# language, is available on the accompanying CD.

**SplitFilesIntoFoldersByPrefix** was used to split the unsorted files from the Demozoo archive into a number of folders, one for each demo. **DeleteFoldersOfSmallIntros** then discarded any folders representing a demo of 4 kilobytes or smaller in size (i.e. where the metadata specified its type as "64b Intro", "128b Intro", "256b Intro", "512b Intro", "1K Intro", or "4K Intro"). Such productions, written as a competitive exercise in minimalism (Montfort, 2013), are too small to contain significant text.

Individual snapshots were then generated as described in the methodology. The emulator was configured as a Russian 'Pentagon 128' clone, since this provides the best overall compatibility for demos (including many that would not run on an original Spectrum

model). Snapshots were saved in the uncompressed SNA format (comp.sys.sinclair FAQ, 2005), since data compression might interfere with the texts.

**SnapshotTextHunter** (Figure 1) facilitated the manual recovery of text from the demo snapshots. Its graphical interface presents the entire binary contents of a snapshot in text format, with edit and save facilities to allow those portions not actually representing text to be quickly removed. This tool also handled the conversion of the Spectrum character set to modern ASCII (American Standard Code for Information Interchange), which differs in a few particulars. Any non-convertible characters (including binary code and data) were replaced with ¶ paragraph marks for easy identification, and long blocks of such characters were automatically deleted.



Figure 1. Locating text visually in the SnapshotTextHunter tool.

A demo or snapshot was discarded if (i) it contained little or no English text; (ii) the demo could not be successfully launched for some reason (e.g. lack of Pentagon compatibility, or shortcomings in the emulator); (iii) the text that was visible on the screen could not be found in the snapshot, as a result of unconventional storage techniques or encryption by the programmer; or (iv) the snapshot represented a 'cracktro' prepended to a production by another group acting as redistributor: this would otherwise distort the language variable, since the cracking group might be from a different country.

Finally, **RenameSnapshotTexts** carried out automatic renaming of the resulting text files according to the convention *id-creator--title--country*. For example, the file 9794--*Hooy-Program--SoundTracker 20th Anniversary--GB,PL.txt* contains text from the demo *SoundTracker 20th Anniversary* by Hooy-Program, a group whose members hail from Britain and Poland.

## **Appendix C: Custom software tools for data analysis**

Some further software tools were developed to assist in analysing the texts in the corpus. They are described here sequentially, in the order of their use. Again, source code is available on the accompanying CD.

Source filename	Text portion
12635Jano-K & Neuron	to turn off the movie . today we will >>see again angel , in
Angel 2.txt	her favourite situation
12658Dream Makers	no , don " t look out , >>simply we can swop these
AssortyBY.txt	programs . ok
13307PentagramLSD	? ? ? hello to all of you ! >>especially we know and greet :
PL.txt	mac,
PL.txt 13325The Mad Guys	mac, the game with ' emand i 've >>got still stuff to write,
PL.txt 13325The Mad Guys Mad Guys Intro 5 48K, The-	<pre>mac , the game with ' emand i 've &gt;&gt;got still stuff to write , some ratings</pre>
PL.txt 13325The Mad Guys Mad Guys Intro 5 48K, The- -DE.txt	mac , the game with ' emand i 've >>got still stuff to write , some ratings
PL.txt 13325The Mad Guys Mad Guys Intro 5 48K, The- -DE.txt 13349K3LMarwin's	<pre>mac , the game with ' emand i 've &gt;&gt;got still stuff to write , some ratings to draw so nice pics . we hope you &gt;&gt;like also this attr</pre>

Table B1: Fragment of the feature concordance file for 'innovative word order with adverbials'

**FeatureFinder** searched for occurrences of features in the tagged text files making up the corpus, according to the search strategies and heuristics set out in the methodology. For each feature, a *feature concordance file* was generated in CSV (tab-separated) format. Each line in the file contained the source text's filename and the brief portion of text thought to exemplify the feature, with the central word marked with the symbol >> and surrounded by a few tokens of cotext on each side. Each file therefore had the appearance of a concordance, though not having the same node (root word) in every line, but rather an instance of the same

feature. By way of illustration, Table B1 shows five lines from the feature concordance file for the feature 'innovative word order with adverbials'.

The next tool, **AddCountryColumnToCsvs**, processed each feature concordance file and added an extra column on the left representing the originating country (determined from the filename in each line). This column was left blank for demos of unknown national origin and those marked as coming from more than one country (since there is no reliable means of identifying which of those countries the writer of the text came from). Meanwhile, **CountTokensPerCountry** computed the total number of tokens in the corpus for each country, for use in calculating feature densities.

At this point, the researcher manually marked up each of the feature concordance files for false positives. This involved reading through each file in a spreadsheet program and placing the letter 'n' in a marker column for each line that did not in fact exemplify the desired feature.

**CalcFeatureDensities** then computed the feature densities per country by counting the legitimate instances of each feature (i.e. those not manually marked with an 'n') and normalising them according to the previously computed total number of tokens per country.

The capitalisation feature could not be conveniently measured by FeatureFinder because of the fact that texts had been converted to lower case as part of tagging. Therefore a separate program, **CheckCorpusLcaseUcase**, was developed to count the numbers of texts that were primarily lower-case or primarily upper-case.

Finally, **FindLexis** searched the corpus, disregarding punctuation, and generated a list of unfamiliar words, which might represent demoscene jargon, new coinages, or sensational spellings. A word was judged to be unfamiliar if it did not occur among the 267,751 English words in a dictionary file obtained by the author some years ago from a now-defunct Web site and thus now unfortunately unsourceable – though it appears to be an earlier version of Collins Official Scrabble Words (HarperCollins, 2016). The output was a file listing all unfamiliar words in the corpus, ranked by descending frequency. This list was then manually filtered to remove foreign words (primarily found in code-switched passages), everyday proper nouns and abbreviations (*June, English, etc, OK, TV*), names and pseudonyms, and a few obviously non-sensational misspellings (*adress, realy, writting*).

43

# **Appendix D: Innovative lexical items**

This table is a frequency list of the top 50 innovative lexical items in the demo corpus (i.e. those not present in an English dictionary, as detailed in Appendix C).

Rank	Frequency	Word	Function	Notes
1	360	megademo	scene jargon	large multi-part demo
2	345	scrolly	scene jargon	scrolling message
3	259	gfx	abbreviation	graphics
4	257	ty	abbreviation	thank you
5	239	ZX	hardware	
6	234	scroller	scene jargon	scrolling message
7	218	128k	hardware	
8	216	thanx	sensational spelling	thanks
9	207	PC	hardware	
10	156	Atari	hardware	
11	146	erm	speech sound	
12	142	PD	abbreviation	public-domain software
13	121	48k	hardware	
14	105	Sinclair	hardware	
15	101	greetinx	sensational spelling	greetings
16	98	grafix	sensational spelling	graphics
17	95	iz.	sensational spelling	is
18	94	Soundtracker	software	
19	85	greetz	sensational spelling	greets
20	79	IBM	hardware	
21	78	fx	abbreviation	effects
22	63	hmmm	speech sound	
23	62	Z80	hardware	
24	55	C64	hardware	
25	55	NMI	abbreviation	non-maskable interrupt
26	53	graphix	sensational spelling	graphics
27	51	ohhh	speech sound	

Table C1: Innovative lexical items

Rank	Frequency	Word	Function	Notes
28	51	bla	speech sound	
29	50	yer	sensational spelling	your
30	50	scrollies	scene jargon	scrolling messages
31	49	scrollers	scene jargon	scrolling messages
32	48	STS	software	
33	48	fuckin	sensational spelling	fucking
34	43	FLP	abbreviation	flexible line position
35	43	NeOS	software	
36	42	ZX-Spectrum	hardware	
37	42	KB	abbreviation	kilobytes
38	42	Scrunk	software	
39	38	YS	abbreviation	Your Sinclair magazine
40	38	Didaktik	hardware	
41	37	48/128	hardware	
42	37	ohh	speech sound	
43	36	MSX	hardware	
44	36	lamers	scene jargon	incompetent users
45	36	Multiface	hardware	
46	35	guyz	sensational spelling	guys
47	35	Timex	hardware	
48	35	Tasword	software	
49	35	FDD	abbreviation	floppy disk drive
50	35	hur	speech sound	