The Internet: Changing the Language
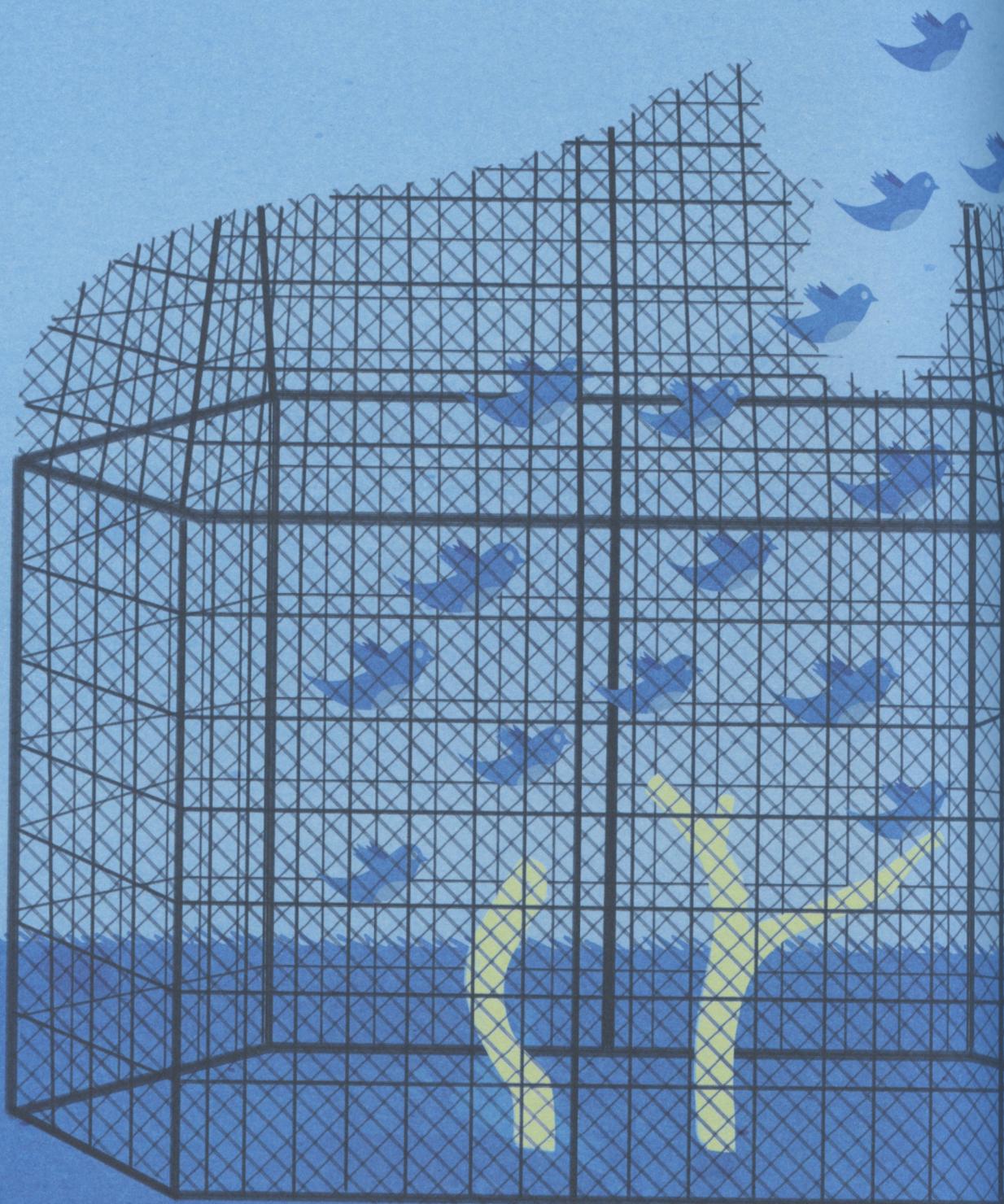
David Crystal

Honorary Professor of Linguistics, Bangor University

David Crystal

davidcrystal.com

David Crystal is Honorary Professor of Linguistics at the
University of Bangor, and works from his home in Holyhead,
North Wales, as a writer, editor, lecturer, and broadcaster.
He read English at University College London, specialized
in English language studies, then joined academic life as
a lecturer in linguistics, first at Bangor, then at Reading,
where he became professor of linguistics. His books related
to his article include *Language and the Internet* (Cambridge
University Press, 2001) and *Internet Linguistics* (Routledge,
2011), and he is the inventor of the semantic targeting
approaches to online advertising, iSense and Sitescreen.
An autobiographical memoir, *Just a Phrase I'm Going Through:
My Life in Language*, was published in 2009 by Routledge.

Sites and services that have changed my life

oed.com

davidcrystal.com

shakespeareswords.com

johnbradburnepoetry.com

The Internet: Changing the Language

It is, of course, too soon to say what permanent effect the Internet will have on languages. Electronically mediated communication (EMC) has been in routine use for only around twenty years, and this is an eyeblink in the history of a language. It takes time—a lot of time—for a change to emerge, for individuals to get used to its novelty, for them to start using it in everyday speech and writing, and for it eventually to become so widely used that it becomes a permanent feature of a language, recorded in dictionaries, grammars, and manuals of style. There are already some telltale signs of what may happen, but everything has to be tentative.

The Difficulty of Generalization

All general statements about EMC are inevitably tentative because of the nature of the medium. Its size, for a start, makes it difficult to manage: there has never been a corpus of language data as large as this one, containing more written language than all the libraries in the world combined. Then there is its diversity, which defies linguistic generalization: the stylistic range of EMC includes the vast outputs found in e-mail, chat rooms, the web, virtual worlds, blogging, instant messaging, text messaging, and Twitter, as well as the increasing amount of linguistic communication encountered in social networking forums such as Facebook, each output presenting different communicative perspectives, properties, strategies, and expectations.

The speed of change makes it difficult to keep pace. How can we generalize about the linguistic style of e-mails, for example? When it first became prevalent, in the mid-1990s, the average age of e-mailers was in the 20s, and it has steadily risen. To take one year at random: the average in the UK rose from 35.7 to 37.9 between October 2006 and October 2007 (Nielsen 2007). This means that many e-mailers, for example, are now senior citizens. The consequence is that the original colloquial and radical style of e-mails (with their deviant spelling, punctuation, and capitalization) has

been supplemented by a more conservative and formal style, as older people introduce norms derived from the standard language. Similarly, the average age of a Facebook user has sharply risen in the past decade, from a predominantly young person's medium to a medium for everyone: in 2012 it was 40.5 years (Pingdom 2013).

But it is not solely a matter of age. The pragmatic purpose of a piece of EMC can alter, sometimes overnight. A good example is Twitter which, when it arrived in 2006, used the prompt "What are you doing?" The result was a range of tweets which were inward-looking, using lots of first-person pronouns and present tenses. Then in November 2009 Twitter changed its prompt to "What's happening?" This made the tweets outward-looking, with lots of third-person pronouns, and a wider range of tense forms. The result was a shift in the aims and linguistic character of Twitter, which took on more of the features of a news service, as well as attracting more advertising content.

EMC as Writing or Speech

EMC, for the moment, is predominantly a written medium. This will not always be so. Voice over Internet (VoI) is rapidly increasing, and already it is possible to engage in many kinds of interactions without the fingers touching the keyboard at all, using speech-to-text software. The technique is a long way from perfection: systems have recurrent problems with regional accents, speed of speech, background noise, and the interpretation of proper names. But these will reduce as time goes by.

Some people say that in 50 years' time keyboards will be redundant, but this is unlikely because speech and writing perform very different and complementary functions. EMC relies on characteristics belonging to both sides of the speech/writing divide.

The graphic character of EMC is best illustrated by the web, which in many of its functions (e.g., databasing, reference publishing, archiving,

advertising) is no different from traditional situations which use writing; indeed, most varieties of written language (legal, religious, and so on) can now be found on the web with little stylistic change other than an adaptation to the electronic medium. In contrast, the situations of e-mail, chat groups, virtual worlds, and instant messaging, though expressed through the medium of writing, display several of the core properties of speech. They are time-governed, expecting or demanding an immediate response; they are transient, in the sense that messages may be immediately deleted (as in e-mails) or be lost to attention as they scroll off the screen (as in chat groups); and their utterances display much of the urgency and energetic force which is characteristic of face-to-face conversation. The situations are not all equally *spoken* in character. We *write* e-mails, not *speak* them. But chat groups are for *chat*, and people certainly *speak* to each other there—as do people involved in virtual worlds and instant messaging.

Another distinctive feature of EMC writing is that, apart from in audio/video interactions (such as Skype or iChat), it lacks the facial expressions, gestures, and conventions of body posture and distance which are so critical in expressing personal opinions and attitudes and in moderating social relationships. The limitation was noted early in the development of the medium, and led to the introduction of *smileys* or *emoticons*. Today there are some sixty or so emoticons offered by message exchange systems. It is plain that they are a potentially helpful way of capturing some of the basic features of facial expression, but their semantic role is limited. They can forestall a gross misperception of a speaker's intent, but an individual emoticon still allows a large number of readings (happiness, joke, sympathy, good mood, delight, amusement, etc.) which can only be disambiguated by referring to the verbal context. Without care, moreover, they can foster misunderstanding: adding a smile to an utterance which is plainly angry can increase rather than decrease the force of the *flame*. So it is not surprising to see the use of emoticons falling, as time goes by. People have realized that they do not solve all communication problems in EMC, and may even add to them.

New Communicative Opportunities in EMC

When we consider EMC as a species of written language, and compare it with traditional modes of writing, certain novel properties are immediately apparent. However, these properties are nothing to do with the standard conception of writing as a combination of vocabulary, grammar, and orthography. EMC has certainly introduced a few thousand new words into English, for example, but these make up only a tiny fraction of the million+ words that exist in that language. There is nothing revolutionary here. Similarly, the grammar of written English, as seen in EMC, displays no novelty in comparison with what was used before—no radically different word orders (syntax) or word endings (morphology). And despite the way people manipulate certain features of the orthography, such as simplifying punctuation marks or using them excessively, or adding the occasional emoticon, the writing system on the whole looks very similar to what existed in pre-EMC days. The novelty of EMC writing lies elsewhere, in the opportunities it presents for fresh kinds of communicative activity, and in the development of new styles of discourse.

There is a contrast, first of all, with the space-bound character of traditional writing—the fact that a piece of text is static and permanent on the page. If something is written down, repeated reference to it will encounter an unchanged text. Putting it like this, we can see immediately that EMC is not by any means like conventional writing. A *page* on the web often varies from encounter to encounter (and all have the option of varying, even if page-owners choose not to take it) for several possible reasons: its factual content might have been updated, its advertising sponsor might have changed, or its graphic designer might have added new features. Nor is the writing that we see necessarily static, given the technical options available which allow text to move around the screen, disappear/reappear, change color, and so on. From a user point of view, there are opportunities to *interfere* with the text in all kinds of ways that are not possible in traditional writing. A page, once downloaded to the user's screen, may have its text cut, added to, revised, annotated, even totally restructured, in ways that nonetheless retain the character of the original. The possibilities are causing not a little anxiety among those concerned about issues of ownership, copyright, and forgery.

Secondly, EMC outputs display differences from traditional writing with respect to their space-bound presence. E-mails are in principle static

and permanent, but routine textual deletion is expected procedure (it is a prominent option in the management system), and it is possible to alter messages electronically with an ease and undetectability which is not possible when people try to alter a traditionally written text. Messages in asynchronic chat groups and blogs tend to be long-term in character; but those in synchronic groups, virtual worlds, and instant messaging are not. In the literature on EMC, reference is often made to the *persistence* of a conversational message—the fact that it stays on the screen for a period of time (before the arrival of other messages replaces it or makes it scroll out of sight).

Thirdly, we see differences between some EMC outputs and traditional writing when we ask how complex, elaborate, or contrived they are. Certain outputs are very similar to what happened before. In particular, the web allows the same range of planning and structural complexity as would be seen in writing and printing offline. But for chat groups, virtual worlds, and instant messaging, where the pressure is strong to communicate rapidly, there is much less complexity and forward planning. Blogs vary greatly in their constructional complexity: some are highly crafted; others are wildly erratic, when compared with the norms of the standard written language. E-mails also vary: some people are happy to send messages with no revision at all, not caring if typing errors, spelling mistakes, and other anomalies are included in their messages; others take as many pains to revise their messages as they would in non-EMC settings.

Fourthly, traditional writing is visually decontextualized: normally we cannot see the writers when we read their writing, and we can give them no immediate visual feedback, as we could when talking to someone in face-to-face conversation. In these respects, EMC is just like traditional writing. But web pages often provide visual aids to support text, in the form of photographs, maps, diagrams, animations, and the like; and many virtual-world settings have a visual component built in. The arrival of web-cams is also altering the communicative dynamic of EMC interactions, especially in instant messaging, and some interesting situations arise. I observed an anomalous one recently, where A and B were attempting to use an audio/video link via iChat, but B's microphone was down. As a result B could hear A but A could not hear B, who thus had to resort to her keyboard. A's spoken stimulus was followed by B's written response.

After a somewhat chaotic start, the conversation settled down into a steady rhythm.

Fifthly, we can compare the factual content of EMC and traditional writing. The majority of the latter is factually communicative, as is evident from the vast amount of reference material in libraries. A focus on fact is also evident on the web, and in many blogs and e-mails; but other EMC situations are less clear. Within the reality parameters established by a virtual world, factual information is certainly routinely transmitted, but there is a strong social element always present which greatly affects the kind of language used. Chat groups vary enormously: the more academic and professional they are, the more likely they are to be factual in aim (though often not in achievement, if reports of the amount of flaming are to be believed). The more social and ludic chat groups, on the other hand, routinely contain sequences which have negligible factual content. Instant message exchanges are also highly variable, sometimes containing a great deal of information, sometimes being wholly devoted to social chitchat.

Sixthly, traditional writing is graphically rich, as we can immediately see from the pages of many a fashion magazine. The web has reflected this richness, but greatly increased it, the technology putting into the hands of the ordinary user a range of typographic and color variation that far exceeds the pen, the typewriter, and the early word processor, and allowing further options not available to conventional publishing, such as animated text, hypertext links, and multimedia support (sound, video, film). On the other hand, as typographers and graphic designers have repeatedly pointed out, just because a new visual language is available to everyone does not mean that everyone can use it well. Despite the provision of a wide range of guides to Internet design and desktop publishing, examples of illegibility, visual confusion, over-ornamentation, and other inadequacies abound. They are compounded by the limitations of the medium, which cause no problem if respected, but which are often ignored, as when we encounter screenfuls of unbroken text, paragraphs which scroll downwards interminably, or text which scrolls awkwardly off the right-hand side of the screen. The problems of *graphic translatability* are only beginning to be appreciated—that it is not possible to take a paper-based text and put it on a screen without rethinking the graphic presentation and even, sometimes, the content of the message.

EMC, then, offers new communicative possibilities in the way people can manipulate written language. And already we can see how these opportunities are creating new kinds of electronic discourse.

New Kinds of Text

Every time a new technology arrives, we see the growth of new kinds of discourse, reflecting the aims and intentions of the users. Printing introduced us to such notions as newspapers, chapter organization, and indexes. Broadcasting brought sports commentary, news reading, and weather forecasting. EMC is no different. The content displayed on a screen presents a variety of textual spaces whose purpose varies. There is a scale of online adaptability. At one extreme, we find texts where no adaptation to EMC has been made—a PDF of an article on screen, for example, with no search or other facilities—in which case, any linguistic analysis would be identical with that of the corresponding offline text. At the other extreme, we find written texts which have no counterpart in the offline world. Here are four examples.

Texts whose aim is to defeat spam filters

We only have to look in our e-mail junk folder to discover a world of novel texts whose linguistic properties sometimes defy analysis:

supr vi-agra online now znwygghsxp
VI @ GRA 75% off regular xxp wybzz lusfg
fully stocked online pharmac^y
Great deals, prescription d[rugs

It is possible to see a linguistic rationale in the graphological variations in the word *Viagra*, for example, introduced to ensure that it avoids the word-matching function in a filter. We may find the letters spaced (*V i a g r a*), transposed (*Viarga*), duplicated (*Viaggra*), or separated by arbitrary symbols (*Vi*agra*). There are only so many options, and these can

to a large extent be predicted. There have been huge advances here since the early days when the stupid software, having been told to ban anything containing the string S-E-X, disallowed messages about Sussex, Essex, and many other innocent terms. There is also an anti-linguistic rationale, as one might put it, in which random strings are generated (*wybzz*). These too can be handled, if one's spam filter is sophisticated, by telling it to remove any message which does not respect the graphotactic norms of a language (i.e., the rules governing syllable structure, vowel sequence, and consonant clusters).

Texts whose aim is to guarantee higher rankings in web searches

How is one to ensure that one's page appears in the first few hits in a web search? There are several techniques, some nonlinguistic, some linguistic. An example of a nonlinguistic technique is the frequency of hypertext links: the more pages link to my site, the more likely my page will move up the rankings. An example of a linguistic technique is the listing of key words or phrases which identify the semantic content of a page in the page's metadata: these will be picked up by the search engine and given priority in a search. Neither of these techniques actually alters the linguistic character of the text on a page. Rather different is a third technique, where the text is manipulated to include keywords, especially in the heading and first paragraph, to ensure that a salient term is prioritized. The semantic difference can be seen in the following pair of texts (invented, but based on exactly what happens). Text A is an original paragraph; text B is the paragraph rewritten with ranking in mind, to ensure that the product name gets noticed:

The Crystal Knitting-Machine is the latest and most exciting product from Crystal Industries. It has an aluminum frame, comes in five exciting colors, and a wide range of accessories.

The Crystal Knitting-Machine is the latest and most exciting product from Crystal Industries.

- The Crystal Knitting-Machine has an aluminum frame.
- The Crystal Knitting-Machine comes in five exciting colors.
- The Crystal Knitting-Machine has a wide range of accessories.

Some search engines have got wise to this technique, and try to block it, but it is difficult, in view of the various paraphrases which can be introduced (e.g., *Knitting-Machine from Crystal*, *Crystal Machines for Knitting*).

Texts whose aim is to save time, energy, or money

Text messaging (a different sense of the term *text*, note) is a good example of a genre whose linguistic characteristics have evolved partly as a response to technological limitations. The limitation to 160 characters (for Roman alphabets) has motivated an increased use of nonstandard words (of the *c u l8r* type), using logograms, initialisms, shortenings, and other abbreviatory conventions. The important word is *partly*. Most of these abbreviations were being used in EMC long before mobile phones became a routine part of our lives. And the motivation to use them goes well beyond the ergonomic, as their playful character provides entertainment value as an end in itself as well as increasing rapport between participants. I have developed this point in my *Txtng: the Gr8 Db8*.

Another example of a new type of text arising out of considerations of convenience is the e-mail which uses *framing*. We receive a message which contains, say, three different points in a single paragraph. We can, if we want, reply to each of these points by taking the paragraph, splitting it up into three parts, and then responding to each part separately, so that the message we send back then looks a bit like a play dialogue. Then, our sender can do the same thing to our responses, and when we get the message back, we see his replies to our replies. We can then send the lot on to someone else for further comments, and when it comes back, there are now three voices framed on the screen. And so it can go on—replies within replies within replies—and all unified within the same screen typography. People find this method of response extremely convenient—to an extent, for there comes a point where the nested messages make the text too complex to be easily followed.

Related to framing is intercalated response. Someone sends me a set of questions, or makes a set of critical points about something I have written. I respond to these by intercalating my responses between the points made by the sender. For clarity, I might put my responses in a different

color, or include them in angle brackets or some such convention. A further response from the sender might lead to the use of an additional color; and if other people are copied in to the exchange, some graphical means of this kind, to distinguish the various participants, is essential.

Texts whose aim is to maintain a standard

Although the Internet is supposedly a medium where freedom of speech is axiomatic, controls and constraints are commonplace to avoid abuses. These range from the excising of obscene and aggressive language to the editing of pages or posts to ensure that they stay focused on a particular topic. Moderators (facilitators, managers, wizards... the terminology is various) have to deal with organizational, social, and content-related issues. From a textual point of view, what we end up with is a sanitized text, in which certain parts of language (chiefly vocabulary) are excluded. It is not clear how far such controls will evolve, as the notion of textual responsibility relating to the libel laws is still in the process of being tested.

A good example of content moderation is in the online advertising industry, where there is a great deal of current concern to ensure that ads on a particular web page are both relevant and sensitive to the content of that page. Irrelevance or insensitivity leads to lost commercial opportunities and can generate extremely bad PR. Irrelevance can be illustrated by a CNN report of a street stabbing in Chicago, where the ads down the side of the screen said such things as "Buy your knives here"—the software being unaware that the weapons sense of *knife* in the news report did not match the cutlery sense of *knife* in the ad inventory. Insensitivity can be illustrated by a German page which was describing heritage visits to Auschwitz; the same silly software, having found "gas" mentioned several times on the page, linked this with a power company's ads for "cheap gas," much to the embarrassment of all concerned. One solution, known as *semantic targeting* (and now available in Ad Pepper Media's iSense and Sitescreen products) carries out a complete lexical analysis of web pages and ad inventories so that subject matter is matched and ad misplacements avoided. In extreme cases, such as a firm which does not want its ad to appear on a particular page (e.g., a child clothing manufacturer on an adult porn site), ads can be blocked from appearing. As a result, from a content point of view, the text that appears on a page appears more

semantically coherent and pragmatically acceptable than would otherwise be the case.

Texts Sans Frontières

All the texts mentioned so far have one thing in common: they are easily identifiable and determinate. They have definable physical boundaries, which can be spatial (e.g., letters and books) or temporal (e.g., broadcasts and interviews). They are created at a specific point in time; and once created, they are static and permanent. Each text has a single authorial or presenting voice (even in cases of multiple authorship of books and papers), and that authorship is either known or can easily be established (except in some historical contexts). It is a stable, familiar, comfortable world. And what the Internet has done is remove the stability, familiarity, and comfort.

Written texts are defined by their physical boundaries: the edges of the page, the covers of the book, the border of the road sign... Spoken texts are defined by their temporal boundaries: the arrival and departure of participants in a conversation, the beginning and end of a broadcast, the opening and closing of a lecture... Internet texts are more problematic. Sometimes, as with a text message or an instant-message exchange, we can clearly identify the start and the finish. But with most Internet outputs there are decisions to be made, as the following examples show.

- Does a single e-mail message constitute a text, or is the text everything available on a screen at a particular point in time, including previously exchanged messages that have not been deleted and any framed or intercalated responses sent by the recipient? And does one include unchanging biodata, such as the sender's address, web links, and taglines?

- *A fortiori*, does an entire website constitute a text, or are the texts the individual elements of the menu (Home, About, Contact, Help...), or the individual pages, or the functional elements seen on these pages (main text, advertisements, comments...)? The distinction has commercial importance in online advertising, where an ad server is likely to serve a different

range of ads to the top page of a site compared to its constituent pages. Sky TV, for example, at one point had a banking ad at the top of its home page, and a video games ad at the top of its sport page. And should we include translations? Many websites now are multilingual, with a list of language choices on the home page. Are these part of the same text, or are they different texts?

- If an e-mail, tweet, instant message, blog, or other output includes an obligatory hypertext link, is that link to be considered as part of the text? By *obligatory* I mean a link that forms part of the structure of a sentence or which provides information that is critical to the understanding of the page, such as "Please go to www... for details," or the links used in tweets.

- If security is an obligatory element (e.g., asking for user names, pass-words, or other authentication), is this to be considered as part of the text? Are the glosses or images which appear when a mouse hovers over a string to be considered as part of the text? And do we include the key-words which identify the page, and which may not appear on the screen, but are only visible when one looks at the underlying code, as here?

```
<HEAD>
<TITLE>Stamp Collecting World</TITLE>
<META name="description" content="Everything you wanted to know
about stamps, from prices to history.">
<META name="keywords" content="stamps, stamp collecting, stamp
history, prices, stamps for sale">
</HEAD>
```

- How are we to define a text in an Internet output which is continuously growing, as in a social networking site, a chat room, a blog forum, or a bulletin board, which might last indefinitely? In these cases there is a dynamic archive, which in some cases goes back many years. Are asso-ciated comments to be considered part of the text? As they are elicited by the main text, and are semantically (and sometimes grammatically) dependent on it, they cannot be taken as independent texts in their own right. There is an asymmetrical relationship: the main text has autonomy: it does not need comments to survive; but comments could not exist

without a main text. And there is no theoretical limit to the number of comments a post might elicit.

- Similarly, how are we to define a text in an Internet output which is continually changing—where there is permanently scrolling data, regularly updated, such as stock-market reports and news headlines? Here there may be no archive: old information is deleted as it is replaced. The content comes from an inventory which is fixed at any one point in time, but frequently refreshed. Some sequences that appear on-screen are cyclical (such as the recurring headlines we see in a news-ticker service or a retail store); others are randomly generated (such as the pop-up ads or banner ads taken from a large inventory, which may change in front of your eyes every few seconds.

- What do we do with a message sequence (as in e-mails or a bulletin board) where the subject line identifies a semantic thread? Is the text the set of messages that relate to that thread (as in items 4 and 9 below)? They may be separated by other messages, as in this example from a Shakespeare forum:

  4 Arden3 The Merchant of Venice
  5 Thoughts on Double Falsehood
  6 Arden3 Sir Thomas More
  7 2011 Blackfriars Conference Announcement
  8 From New York to Santa Fe
  9 Arden3 The Merchant of Venice

- Do we follow the header? If so, what do we do with cases where (a) the discussion continues but someone changes the header in the subject line, or (b) the header in the subject line remains the same, but the discussion veers off-topic? Which takes priority?

- Are we to include in the text elements automatically inserted by cookies, such as site preferences, shopping cart contents, and visitor tracking, or the features which are available to users, such as helplines and analytics reports?

- How do we view texts rendered incomplete by the technology, as when a tweet exceeds the 140-character limit and is truncated by the software? This is shown by ellipsis dots on screen.

The traditional notion of *text* is inadequate to handle these cases. A broader, more inclusive notion is going to be needed. Clearly, what we see in all these examples are aggregates of functional elements, which interact in various ways in different Internet outputs. We need terms for both the elements and the aggregates. Dürscheid and Jucker (2011), for example, call the elements "communicative acts," and the aggregates "communicative act sequences." Doubtless other proposals will be forthcoming, as linguists explore these phenomena in more detail. In the meantime, here are some general observations.

Panchronicity

The above examples are not a complete list of the boundary decisions which have to be made when we are trying to identify Internet texts, but they are representative of what is *out there*. And they raise quite fundamental questions. In particular, Ferdinand de Saussure's classical distinction between synchronic and diachronic does not adapt well to these kinds of communication, where everything is diachronic, time-stampable to a micro-level. Texts are classically treated as synchronic entities, by which we mean we disregard the changes that were made during the process of composition and treat the finished product as if time did not exist. But with many electronically mediated texts there is no finished product. And in many cases, time ceases to be chronological.

For example, I can in 2011 post a message to a forum discussion about a page which was created in 2004. From a linguistic point of view, we cannot say that we now have a new synchronic iteration of that page, because the language has changed in the interim. I might use in my message vocabulary that has entered the language since 2004, or show the influence of an ongoing grammatical change. Content is inevitably affected. I might refer to Twitter—something which would not have been possible in 2004, for that network did not appear until 2006. I might even—as is possible with Wiki pages—insert information into the main text of a page which could not have been available at the time of the page's creation. In the case of my blog, I might go back to a post I wrote in 2004 and edit it to include material from 2013.

We need a new term for this curious conflation of language from different time periods. We are very familiar with texts which include language from earlier periods (archaisms). We need a way of describing features of texts which include language from later periods. The traditional term for a chronological mismatch is anachronism—when something from a particular point in time is introduced into an earlier period (before it existed) or a later period (after it ceased to exist). Anachronisms can be isolated instances—as when Shakespeare introduces striking clocks into ancient Rome (in *Julius Caesar*)—or a whole text can be anachronistic, as when a modern author writes a play about the seventeenth century and has everyone speak in a twenty-first-century way. But these cases don't quite capture the EMC situation, where a chronological anomaly has been introduced into an original text. This is a new take on the grammatical notion of *future in the past*—or, perhaps better, *back to the future*. And I think we need a new term to capture what is happening. A text which contains such futurisms cannot be described as synchronic for it cannot be seen as a single *état de langue* (Saussure's term for a state of the language at a particular point in time): it is a conflation of language from two or more *états de langue*. Nor can it be described as diachronic, for the aim is not to show language change between these different états. Such texts, whose identity is dependent on features from different time frames, I call panchronic.

Wiki pages, such as those seen on Wikipedia, are typically panchronic. They are the result of an indefinite number of interventions by an indefinite number of individuals over an indefinite number of periods of time (which become increasingly present as time goes by). We are only 20 or so years into the web, so the effect so far is limited; but think ahead 50 or 100 years, and it is obvious that panchronicity will become a dominant element of Internet presence. From a linguistic point of view, the result is pages that are temporally and stylistically heterogeneous. Already we find huge differences, such as standard and nonstandard language coexisting on the same page, often because some of the contributors are communicating in a second language in which they are not fluent. Tenses go all over the place, as this example illustrates (reproduced exactly as it appeared in Wikipedia):

> *Following his resignation, Mubarak did not make any media appearances. With the exception of family and a close circle of aides, he reportedly refused to talk to anyone, even his supporters. His health*

was speculated to be rapidly deteriorating with some reports even alleging him to be in a coma. Most sources claim that he is not longer interested in performing any duties and wants to "die in Sharm El-Sheikh."

On 28 February 2011, the General Prosecutor of Egypt issued an order prohibiting Mubarak and his family from leaving Egypt. It was reported that the former president was in contact with his lawyer in case of possible criminal charges against him. As a result, Mubarak and his family had been under house arrest at a presidential palace in the Red Sea resort of Sharm el-Sheikh. On Wednesday 13 April 2011 Egyptian prosecutors said they had detained former president Hosni Mubarak for 15 days, facing questioning about corruption and abuse of power, few hours after he was hospitalized in the resort of Sharm el Sheik.

Note the way for example, we move from past tense to present tense in paragraph 1, and from *was* to *had* in paragraph 2. Note also the way *former president Hosni Mubarak* is introduced in the last sentence, as if this were a new topic in the discourse. Note the three different spellings of the Red Sea resort. And how are we to interpret such nonstandard usages as *was speculated*, *in case of*, and *few hours*?

In pages like this, traditional notions of stylistic coherence, with respect to level of formality, technicality, and individuality, no longer apply, though a certain amount of accommodation is apparent, either because contributors sense the properties of each other's style, or a piece of software alters contributions (e.g., removing obscenities), or a moderator introduces a degree of leveling. The pages are also semantically and pragmatically heterogeneous, as the intentions behind the various contributions vary greatly. Wiki articles on sensitive topics illustrate this most clearly, with judicious observations competing with contributions that range from mild through moderate to severe in the subjectivity of their opinions. And one never knows whether a change introduced in a wiki context is factual or fictitious, innocent or malicious.

The problem exists even when the person introducing the various changes is the same. The author of the original text may change it—refreshing a web page, or revising a blog posting. How are we to view the relationship

between the various versions? This is not the first time we have encountered this problem. It is a familiar problem for medievalists faced with varying versions of a text. It is a routine question in the case of, say, Shakespeare: Did he (or someone else) go back and revise an earlier manuscript? It is something we see all the time in the notion of a *second edition*, where the two layers of text may be separated by many years. But what is happening on the Internet is hugely different from the traditional process of revision, because it is something that authors can do with unprecedented frequency and in unprecedented ways. A website page can be refreshed, either automatically or manually.

The issue is particularly relevant now that print-on-demand texts are becoming common. It is possible for me to publish a book very quickly and cheaply, printing only a handful of copies. Having produced my first print run, I then decide to print another, but make a few changes to the file before I send it to the POD company. In theory (and increasingly common in practice), I can print just one copy, make some changes, then print another copy, make some more changes, and so on. The situation is beginning to resemble medieval scribal practice, where no two manuscripts were identical, or the typesetting variations between copies of Shakespeare's First Folio. The traditional terminology of *first edition*, *second edition*, *first edition with corrections*, ISBN numbering, and so on, seems totally inadequate to account for the variability we now encounter. The same problem is also present in archiving. The British Library, for example, launched its Web Archiving Consortium a few years ago. My website is included. But how do we define the relationship between the various time-stamped iterations of this site, as they accumulate in the archive?

Anonymity

I mentioned five criteria above: texts have definable physical boundaries; they are created at a specific point in time; they are static and permanent; they have a single authorial or presenting voice; and—apart from in some historical contexts—authorship is either known or can easily be established. None of these criteria are necessarily present on the Internet. And in the case of the last of these, its absence presents linguists with a

particularly difficult situation. When we classify texts into types we rely greatly on extralinguistic information. This is something we have learned from sociolinguistics and stylistics: the notion of a language variety (or register, or genre, or whatever) arises from a correlation of linguistic features with extralinguistic features of the situation in which it occurs, such as its formality or occupational identity. In principle we know the speaker or writer—whether male or female, old or young, upper class or lower class, scientist or journalist, and so on. And when we do research we try to take these variables into account in order to make our study comparable to others or distinguishable from others in controlled ways. In short, we know who we are dealing with.

But on the Internet, a lot of the time, we don't. The writer is anonymous. In a wide range of Internet situations, people hide their identity, especially in chat groups, blogging, spam e-mails, avatar-based interactions (such as virtual reality games and Second Life), and social networking. These situations routinely contain individuals who are talking to each other under nicknames (nicks), which may be an assumed first name, a fantasy description (topdude, sexstar), or a mythical character or role (rockman, elfslayer). Operating behind a false persona seems to make people less inhibited: they may feel emboldened to talk more and in different ways from their real-world linguistic repertoire. They must also expect to receive messages from others who are likewise less inhibited, and be prepared for negative outcomes. There are obviously inherent risks in talking to someone we do not know, and instances of harassment, insulting or aggressive language, and subterfuge are legion. Terminology has evolved to identify them, such as flaming, spoofing, trolling, and lurking. New conventions have evolved, such as the use of CAPITALS to express *shouting*.

While all of these phenomena have a history in traditional mediums, the Internet makes them present in the public domain to an extent that was not encountered before. But we do not yet have detailed linguistic accounts of the consequences of anonymity. All that is clear is that traditional theories don't account for it. Try using Gricean maxims of conversation to the Internet (Grice 1975): our speech acts, he says, should be truthful (the maxim of quality), brief (the maxim of quantity), relevant (the maxim of relation), and clear (the maxim of manner). Take quality: do not say what you believe to be false; do not say anything for which you lack evidence.

Which world was Grice living in? A pre-Internet world, evidently. Pragmatics people traditionally assume that human beings are nice. The Internet has shown that they are not. Is a pedophile going to be truthful, brief, relevant, and clear? Are the people sending us tempting offers from Nigeria—beautifully pilloried in Neil Forsyth's recent book, *Delete This at Your Peril* (2010)? Are extreme-views sites (such as hate racist sites) going to follow Geoffrey Leech's (1983) maxims of politeness (tact, generosity, approbation, modesty, agreement, sympathy)? And if brevity was the soul of the Internet, we would not have such coinages as blogorrhea and twitterrhea.

Electronically mediated communication is not the first medium to allow interaction between individuals who wish to remain anonymous, of course, as we know from the history of telephone and amateur radio; but it is certainly unprecedented in the scale and range of situations in which people can hide their identity, and exploit their anonymity in ways that would be difficult to replicate offline. And the linguist is faced with a growing corpus of data which is uninterpretable in sociolinguistic or stylistic terms. A different orientation needs to be devised, in which intention and effect become primary, and identity becomes secondary.

The Future

The biggest question marks to do with change on the Internet relate to the way EMC is developing—always difficult to predict as technology rapidly changes. Most of my observations about written language are based on what I have seen on the large screen of my computer. But it is a fact that Internet access is becoming increasingly mobile. Indeed, in some parts of the world, where a wired electricity supply is unreliable or absent (such as a great deal of Africa), the only way of reaching the Internet is via mobile phones. So what happens, in terms of legibility, when a page containing a large amount of visually encoded information is presented on a small screen? How is the information reorganized? What is lost and what is gained? If, as the mobile phone industry is predicting, the majority of Internet access will soon be through handheld devices, then how relevant will be all the generalizations about EMC character that have hitherto been based only on an analysis of large-screen displays?

Finally, this paper has largely focused on written language. The main issue for the future will be how to deal with the increased presence of spoken outputs, as a result of growth in Voice over Internet and mobile communication. There are several new kinds of speech situation here, such as the modifications which are introduced into conversation to compensate for the inevitable lag between participants, automatic speech-to-text translation (as when voicemail is turned into text messages), text-to-speech translation (as when a web page is read aloud), voice recognition interaction (as when we tell the washing machine what to do), and voice synthesis (as when we listen to GPS driving instructions). Each of these domains is going to introduce us to new kinds of output over the next twenty years. Evidently, we ain't seen nothin' yet.

References

Ad Pepper Media.
http://www.adpepper.com/advertiser/overview (accessed March 18, 2013).

British Library.
UK Web Archive. 2008. http://www.webarchive.org.uk/ukwa/

Crystal, David.
Txtng: The Gr8 Db8. Oxford: Oxford University Press, 2008.

Dürscheid, Christa, and Andreas H. Jucker.
"Text As Utterance: Communication in the Electronic Media." Paper presented at the conference "Language As a Social and Cultural Practice: Advances in Linguistics," University of Basel, 2011.

Forsyth, Neil [also known as Bob Servant].
Delete This at Your Peril. Edinburgh: Birlinn, 2010.

Grice, H. Paul.
"Logic and Conversation." In Peter Cole and Jerry L. Morgan, eds. *Syntax and Semantics 3: Speech Acts*. New York: Academic Press, 1975. 41–58.

Leech, Geoffrey.
Principles of Pragmatics. London: Longman, 1983.

Nielsen Company.
"The Ageing UK Internet Population." Nielsen Online News Release, December 18, 2007. http://www.netratings.com/pr/pr_071218_UK.pdf

Pingdom.
"Internet 2012 in Numbers." *Pingdom*, January 16, 2013. http://royal.pingdom.com/2013/01/16/internet-2012-in-numbers