Language, Translation and the Web

Mark Shuttleworth

1 INTRODUCTION AND DEFINITIONS

The area of 'translation and the Web' represents a novel alignment of concepts and subsumes how translation is implemented, enabled, supported, promoted, discussed and made available over the Web. While part of translation studies, it also interrelates with the interlocking lines of enquiry referred to as 'Web studies', 'Web science', 'internet studies' or 'internet science' (see Consalvo & Ess 2011; O'Hara & Hall 2013).

Two terms need to be defined before we launch into the discussion:

Web 2.0 (or the '*read/write Web*') Since the early 2000s this term has been used to reflect the collaborativity built into many websites and on-line applications that permits users to interact within virtual communities, creating their own content and contributing to that of other users.

UGC (User-Generated Content) UGC is 'any form of content such as blogs, wikis, discussion forums, posts, chats, tweets, podcasting, digital images, video, audio files, advertisements and other forms of media that was created by users of an online system or service' (Moens, Li & Chua 2014: 7).

2 HISTORICAL PERSPECTIVES

Folaron (2010) approaches translation and the Web from the perspective of Web studies (2010: 449). The present chapter follows Gaspari's (2015) more recent

discussion of what he terms 'online translation', placing it in the context of language on the Web and adding further topics to his discussion's already extensive repertoire.

3 CORE ISSUES AND TOPICS

3.1 The language of the English Web

This section describes the English Web; the situation in other languages is believed to be parallel to what is discussed here.

Today's Web contains a proliferation of digital 'cybergenres'. Some – including advertisements, dictionaries and scholarly papers – are 'extant' genres (Shepherd & Watters 1998: 98). 'Novel' cybergenres (ibid., 99) include, for example, institutional websites, personal homepages, blogs and FAQs.

A number of factors characterise cybergenres. Some pages are obviously the work of several writers; on-going updating means that the text associated with a URL may change between visits while dynamic pages are assembled on the basis of the precise information requested by the user (e.g. on-line store or library catalogue pages). Some linguistic heterogeneity and prefabricatedness can therefore be observed.

Another factor that characterises many web pages is their nature as 'interrupted linear' (Crystal 2006: 204), where the unidimensional text flow is broken not only by white space and paragraph ends but also by screen ends and, in certain contexts, ellipses to indicate omitted text (e.g. a Google search results list).

Alternatively, web pages can be made up of 'non-linear' text fragments (ibid.) that can be read in any order, depending on the reader's interest.

In addition, Web documents tend to be hypertext so that moving between different documents is facilitated.

According to Crystal, language usage in cybergenres also tends to be distinctive in vocabulary, orthography, grammar, pragmatics and style (2011: 57-77). Lexical innovations include words such as *blogger*, *FAQs*, *netizenship* and *twitterati*. Characteristic orthographic usage includes the use of InterCaps and non-standard spellings. Grammatical innovation is often restricted to certain groups of users, an example being the replacement of plural ending *-s* by *-z* to refer to pirated software (e.g. *tunez*, *gamez*, etc.: see Crystal 2011: 67). As an instance of a pragmatically motivated Web practice Crystal cites including keywords in a page's title and first paragraph to help ensure a high search engine ranking (2011: 72). Finally, new Web styles include graphic richness, innovative structural elements (e.g. posts, comments, etc.) and the limitation of tweets to 140 characters (Crystal 2011: 77).

Guides on writing for the Web give similar advice. Usability.gov recommends short sentences and paragraphs, the front-loading of important information, and bulleted and numbered lists (2016). Barr and the Senior Editors of Yahoo! advise shaping text for on-line reading, making it easy to scan, and writing for the world (2011, Chapter 1). Many cybergenres also display an acceptance of non-native English.

3.2 The languages of the Web

With the implementation of successive versions of the Unicode standard
(http://www.unicode.org/versions/Unicode8.0.0) on the encoding of characters
in different scripts, and following the combined work of the ITU (International
Telecommunications Union: http://www.itu.int) and UNESCO
(http://en.unesco.org/themes/building-knowledge-societies) on information and
communications infrastructure and on access to knowledge, and that of ICANN
(Internet Corporation for Assigned Names and Numbers: https://www.icann.org)
on agreeing a set of multilingual and multiscript conventions for the naming of
internet entities (Crystal 2011: 84), the infrastructure is in place for the internet
to take on a fully multi-language identity. According to W3Techs, 53.7% 'of all
the websites whose content language [they] know' use English, Russian
accounts for 6.3%, German for 5.7%, Japanese for 5.0%, Spanish for 4.9%,
French for 4.1%, Portuguese for 2.6%, Italian for 2.1% and Chinese for 1.9%
(2016, figures as of March 2015). In terms of numbers of internet users, English
speakers number 872.9m, Chinese 704.5m, Spanish 256.8m, Arabic 168.1m,
Portuguese 131.9m, Japanese 114.9m, Russian 103.1m, Malay 98.9m, French
97.2m and German 83.7m (Internet World Stats 2015a, figures as of November
2015).

Crystal reports identifying 1000 languages that have 'a modicum of presence on
the Web' (2006: 233). Google searches can be undertaken in 348 languages
(Thomas 2015), while versions of Wikipedia exist in 292 languages as of March
2016 ('List of Wikipedias' 2016).

3.3 Language learning on the Web

An early mention of Web-Assisted Language Learning (WALL) can be found in Gitsaki & Taylor (1999: 143). WALL can take several forms. The Web can be exploited for authentic materials (Timmis 2015: 133-45), which can be edited and turned into learning activities using software such as Hot Potatoes (https://hotpot.uvic.ca). Language-learning sites, often commercial, offer course materials that can be downloaded or accessed on-line (http://www.rosettastone.co.uk; https://www.linguaphone.co.uk), and software that performs specific language-learning functions – such as testing vocabulary – is also available to download. Sites designed to explain grammar have been in existence for a long time (e.g. http://www.verbix.com).

All of these approaches still play important roles. However, Web 2.0 has enabled a previously unthinkable level of interactivity. Thomas (2009) discusses Skype, virtual learning environments, blogging, podcasts, social networking, video, corpora and courseware management systems. To these may be added wikis (Alm 2006; Shuttleworth 2014) and subtitling (Lertola 2012), for example. Each of these exploits the potential of Web 2.0 for the collaborative development of resources and for making use of UGC. Newer kinds of language-learning application and website have also appeared, including Duolingo (http://www.duolingo.com), Wordreference (http://www.wordreference.com) and Anki (http://ankisrs.net) (Shuttleworth 2014).

3.4 On-line machine translation as a utility

Free, on-line translation has been available since the launch of Altavista's Babelfish in 1997, which was based on the rule-based Systran machine

translation (MT) system. The original Babelfish service was initially available in ten language pairs, although that number was gradually increased. Since 1997 websites offering MT have grown in number and diversity: Hutchins (2009) lists 56 free and charging on-line MT services, for example.

In the autumn of 2007, Google Translate (https://translate.google.com), which had been launched the previous year, transferred all its language pairs from SYSTRAN to its own statistical MT system (Chitu 2007), making the addition of further language pairs to the system significantly more straightforward, and free on-line MT more widely available.

The Google Translator Toolkit (https://translate.google.com/toolkit) was launched in 2009 (Garcia & Stevenson 2009: 16). This permits collaboration, interactive use of MT, an enhanced post-editing interface and a limited integration of translation memory and glossaries. In addition, while the direct use of Google Translate remains free, since 2011 (Google Cloud Platform 2015) Google have introduced charges for the use of its API via third-party programs because of 'the substantial economic burden caused by extensive abuse' (ibid.). Nevertheless, a free Google Translate Website Translator plug-in can be added to any website, and a free mobile app is available.

The second most widely used on-line MT system is Microsoft Translator, launched as Bing Translator (http://www.bing.com/translator) in 2009. This is currently available for 52 languages. An extension of the Translator, the Microsoft Translator Hub (http://www.microsoft.com/en-us/translator/hub.aspx) enables users to build their own MT systems. Other systems include Worldlingo

(http://www.worldlingo.co.uk), iTranslate4 (http://itranslate4.eu/en/) and Systranet (http://www.systranet.com/translate). Baidu (http://fanyi.baidu.com) is the main Chinese system, offering translation between 27 languages within a Chinese-language interface. The original Babelfish service no longer operates, although the similarly-named Babel Fish Corporation (https://www.babelfish.com) still maintains a presence.

The output from such systems almost always needs to be post-edited before it can be used for any serious purpose. While much research is being directed towards post-editing (Allen 2003; O'Brien, Simard & Specia 2013) and translation quality (Aiken & Balan 2011; Hampshire & Salvia 2010), the linguistic features of the outputs from different types of MT system have attracted little attention (although see Shen 2010). Likewise, little work exists on the differing characteristics of post-edited MT output and human-translated text.

In the early 2000s the principal uses (all non-specialised) of free on-line MT were to discover information, translate the user's own web page, communicate in another language, provide entertainment, and learn a language (Yang & Lange 2003: 201-2). Since then, indiscriminate, uninformed use of on-line MT has led to a number of high profile gaffs, one of the best known being the decision taken by the Italian government in 2001 to use an on-line MT engine to produce English-language versions of ministers' biographies, which were uploaded to the official government website, apparently without checks (Messina 2001). More recently unedited output from on-line MT seems to be used to translate Wikipedia content from one of its language versions to another, despite Wikipedia's own deprecation of such a practice ('Wikipedia:Translation' 2015).

On-line MT may be in the process of bringing about a fundamental change in the way that interlingual communication is being framed and implemented by significant sections of the translation industry. However, views of professional translators on its use vary, with some embracing the new technology while appreciating its limitations and others feeling frustrated when asked to perform post-editing (Stejskal 2010; Kelly 2014a). For reasons of privacy and confidentiality, those who offer services incorporating statistical MT in their workflow are ethically obliged to inform their clients of this as texts that have been translated are generally automatically incorporated in the system's database.

3.5 The Web as Corpus

The most immediate way to exploit the Web as a corpus is via search engines, an approach described by Bernardini et al. as the Web as a 'corpus surrogate' (2006: 10-11). This approach can be considerably enhanced through the use of Web concordancers (e.g. http://www.webcorp.org.uk or http://www.kwicfinder.com: see Fletcher 2007), which add some of the features of more standard corpus search tools, such as the KWIC format.

A second approach, designated Web as 'corpus shop' by Bernardini et al. (2006: 11-12), involves making a purpose-built corpus by downloading material from the Web. This can involve the use of BootCaT, a set of tools designed to 'bootstrap corpora and terms from the Web' (BootCaT 2015) through a series of automated search engine queries based on 'seeds', or 'terms that are expected to be typical of the domain of interest' (BootCaT 2015). The idea of 'DIY corpora' – defined by Gatto as 'corpora created from the web for a specific purpose' such

as translation or terminological research (2014: 138) – is related but tends to refer to single-use corpora that are *ad hoc* in nature (see Zanettin 2002 and Gatto 2014 among others).

The other two approaches proposed by Bernardini et al. are 'Web as corpus proper' (2006: 13) and 'mega-corpus/mini-Web' (2006: 13-14). The former refers to using the Web as a corpus that can provide representative information on the language of the Web. The latter involves compiling a huge corpus that combines features that are web-like (e.g. size, type of interface, contemporaneity) with those that are corpus-like (e.g. annotation, sophisticated querying language, stability) (see Gatto 2014: 37, 167-71).

Hundt et al. point out that 'we still know very little about the size of this "corpus", the text types it contains, the quality of the material included or the amount of repetitive "junk" that it "samples"' (2007: 2-3). This, along with the question of representativeness (Gatto 2014: 43-5) and the ephemeral nature of much Web content, which renders replication problematic, has caused some scepticism as to the validity of the 'Web as Corpus' concept. Nevertheless, even writers who voice such concerns favour exploiting the Web as a resource for corpus building, and their scepticism only extends to some of the senses described by Bernardini et al. (2006). Gatto argues that while an approach based on the use of commercial search engines does not usually find favour in traditional corpus linguistics, it is 'gaining prominence in language teaching, and indeed it is possibly one of the most widespread – albeit unacknowledged – uses of the Web as Corpus even beyond the corpus linguistics community' (2014: 3).

3.6 On-line translation resources and software

The Web is a repository for terminological and text-based resources, aligned data and specialist software (delivered in the form of SaaS, or 'Software as a Service'), all of which can be used to support translators' work.

Enríquez Raído reports that the internet has been the most used source of information for translators for several years, even though the dynamic manner in which sources exist on the Web gives rise to issues of evaluation, selection and use of the information located (2013: 2). There are a wide range of approaches to the task of locating the terminological, textual and information resources that translators need in order to solve problems on an on-going basis. Ready-made solutions include accessing major on-line databases (e.g. http://untermportal.un.org, http://iate.europa.eu); consulting sites providing links to glossaries and dictionaries (http://www.lexicool.com/dictionary-search.asp, http://lai.com/thc/glmain.html); searching one of the large translator directories (http://www.proz.com, http://translatorscafe.com; see Section 3.9). More open-ended search strategies include consulting a search engine with a gradually evolving search expression to try to pin down an exact TL terminological usage. A small number of apps either already exist (Terminology by Agile Tortoise: http://agiletortoise.com) or are under development (TermSeeker by UCL: https://termseeker.wordpress.com).

Popular sources for textual assets include Wikipedia and Google Books. Textual resources collected from the internet can be invaluable sources of linguistic and technical information for translators, and, if gathered together and accessed via the appropriate software, can function as DIY corpora. Some corpus linguistics

systems boast versions of BootCaT that allow efficient assembly of corpora (http://www.sketchengine.co.uk).

According to Isabelle et al., 'existing translations contain more solutions to more translation problems than any other available resource' (Isabelle et al. 1993: 1137), and parallel text (also known as aligned or bilingual text, or bitext) has been described as 'the fuel that drives modern machine translation systems', the Web being an excellent source for such data (Smith et al. 2013: 1).

However, while large-scale natural language processing projects have exploited resources such as the Internet Archive (http://www.archive.org; Resnik & Smith 2003) and the Common Crawl (https://commoncrawl.org; Smith et al. 2013), the options for individual users lacking the necessary expertise are more limited, although resources such as OPUS (http://opus.lingfil.uu.se; Tiedemann 2012) and those of the European Parliament (http://www.statmt.org/europarl) and the United Nations (http://www.uncorpora.org) are available.

Translation memories, a special type of parallel data, can be downloaded or consulted on-line. One of the first publicly available TMs was the VLTM, or 'Very Large Translation Memory' (https://www.wordfast.net/?whichpage=jobs), which started to be offered free for use with the Wordfast translation memory tool in about 2007. Other services include MyMemory (https://mymemory.translated.net) and WeBiText (http://www.webitext.com/bin/webitext.cgi). TAUS (https://www.taus.net/data/taus-data-cloud) offers a large set of aligned data, most of it is only available for payment. In all cases, the sharing of translation

data has to take place with all due attention to legal considerations (van der Meer & Joscelyne 2013).

3.7 Collaborative and crowdsourced translation

Crowdsourcing and crowdsourced translation (Howe 2006; see further "Translation, Interpreting and New Technologies") has been facilitated by the appearance of Web 2.0. Other terms and concepts such as collaborative, community, volunteer and amateur translation also exist (see O'Hagan 2011a: 13-14).

Fernández Costales argues that the difference between collaborative and crowdsourced translation is one of hierarchy (2013: 96): while both use volunteers, the former involves direct networking between equals whereas in the latter the collaborative effort is 'managed, directed or sponsored' by an organisation (ibid.). The former scenario can be observed in the TED Open Translation Project (DePalma 2009; Olohan 2014; TED Conferences n.d. a), for example, in which over 16,000 translators have collaborated with transcribers to produce some 85,000 translations in 110 languages (ibid.) using the Amara on-line subtitling tool (http://www.amara.org). In contrast, before inviting its bilingual users to help translate its interface into the 140 languages in which it is currently available, Facebook put a system in place that ensures that one model of segment translation and quality approval is strictly followed (Ellis 2009, O'Hagan 2009: 112, Facebook n.d.).

Désilets and van der Meer (2011: 30-34) discuss five 'common issues' in collaborative translation:

1. Business goals:

Advantages of the business model used by Facebook include cost saving, community involvement, enhancing brand loyalty, reducing turnaround time, improving coverage of low-density languages and ephemeral UGC and producing translations that reproduce the linguistic idiosyncrasies of the users of the content (Désilets & van der Meer 2011: 31; see also Ellis 2009). According to Kelly, further attractive features include improvement in quality caused by the involvement of the end-users (2009).

2. Quality control:

In crowdsourcing, this often involves the end-user translators voting for what they consider to be the best translation of each text string, making the translation process self-regulating. Désilets and van der Meer comment that the quality sometimes exceeds that of more traditional processes (2011: 32). By contrast, in their 'controlled crowdsourcing' (Kelly 2012), the non-profit association Translators without Borders vets volunteers before they start translating because the urgency of many projects allows no time for vetting later (ibid.).

3. Crowd motivation

Among volunteer translators motivation can be very high because of their emotional investment in the content they are translating (Désilets & van der Meer 2011: 32). In such contexts companies frequently offer intangible benefits, such as Facebook's rewards and leaderboards (Facebook n.d.) and TED crediting (TED Conferences n.d. b).

4. The role of professionals

Even though companies such as Facebook have made crowdsourcing part of their translation strategy, they employ professionals to check quality (Ellis 2009: 236, 238) and to translate legal documentation (Sargent 2008).

5. Parallelism and de-contextualisation

Much collaborative translation involves breaking down content for distribution among translators. For example, Asia Online's project of translating the English Wikipedia into Thai (DePalma 2011) involved rapid machine translation of 3.5 million Wikipedia articles followed by large-scale post-editing of the articles in segmented form (Morera-Mesa et al.2013: 11).

Morera-Mesa et al. (2013) study workflow patterns used in crowdsourced translation, focusing in particular on questions such as whether alternative translations should be visible to all users and how to assess translations. These and other aspects of crowdsourced work patterns can now be managed by software systems such as Smartling Translation Management System (http://www.smartling.com/platform), Synble Get Localization (http://about.synble.com/?page_id=28), Transifex (http://www.transifex.com) and CrowdFlower (http://www.crowdflower.com) (see also Orr Priebe 2009).

Ethical issues raised by crowdsourced and collaborative translation include the remuneration of participants, the impact on the public perception of translation, the potential positive effect on the visibility of minority languages (McDonough Dolmaya 2011a), and the applicability of professional codes to collaborative

translation, and the nature of new non-professional codes and the new shared ethos (Drugan 2011).

3.8 Wikipedia translation

In early 2016, Wikipedia existed in 292 language versions ('List of Wikipedias' 2016; see also Ayers, Matthews and Yates 2008: 407-18 and Petzold 2012). Articles can be edited by anybody, so that each article becomes a kind of 'evolving continuity' ('Meta:Translate Extension' 2014).

Wikipedia's size and scope have led to it being utilised as a corpus in a number of lexical research packages (e.g. http://www.sketchengine.co.uk and http://corp.hum.sdu.dk). In addition, researchers tap into its potential in a range of computational linguistics applications, including entity linking (or 'wikification': Milne & Witten 2008), word sense disambiguation (Mihalcea 2007) and measuring semantic relatedness (Explicit Semantic Analysis: Gabrilovich & Markovitch 2006).

Much has been written on the language, style and structure of Wikipedia. For the English version, Ayers et al. (2008) stress the importance of readability (2008: 175) and characterise Wikipedia's tone as 'direct, crisp, and contemporary' (2008: 176). They observe that American English and Commonwealth English coexist (2008: 177), a situation that is interestingly mirrored by Simplified and Traditional Chinese on the Chinese version (and sometimes even within the space of a single article). Style recommendations are made in many of the different language versions, with the English articles 'Wikipedia:Manual of Style' (2016) and 'Wikipedia:Writing better articles' (2016) having 98 and 32 so-called interwiki links respectively. Shuttleworth characterises Wikipedia writing as

hybrid in nature as a result of having been authored by multiple writers, and comments on a blurring of the distinction between translation and original writing (2015a). Geolinguists Liao and Petzold provide a network graph based on interwiki links and illustrating the encyclopaedia's major language nodes and the routes along which information is disseminated from one Wikipedia to another – which frequently involve the English version (2010:11).

While it is generally recognised that the encyclopaedia contains material that has been translated from other language versions, it is not certain how much there is or where its greatest concentrations are. The patchwork nature of Wikipedia discourse means that a short translated passage may be present alongside material of a different provenance. Translated material – or material needing to be translated – can, however, be located by referring to the relevant list pages (e.g. 'Category:Translated pages' 2016; 'Category:Articles needing translation from foreign-language Wikipedias' 2016; 'Wikipedia:Pages needing translation into English' 2016). Studying an article's Talk and History tabs can also lead to the identification of specific sections of translated discourse (Shuttleworth 2015b). Human translation appears to be only one of a number of procedures for the cross-language expansion of Wikipedia, others including paraphrase, non-native writing and un-post-edited MT (Shuttleworth 2015a).

3.9 Translation blogs and translator networking sites

The practice of blogging has been taken up rapidly, the period of 2006-2014, for example, seeing the number of blogs in existence grow from 35.77m to 260.47m (Meinel et al. 2015: 8). As of February 2016, the American Translators Association lists 92 translation-related blogs (2016), although there may be far

more. Translation likewise has a strong presence on the microblogging service Twitter.

Myers has written an extended study of the discourse of blogs and wikis (2010). McDonough Dolmaya discusses the 'working conditions, emerging technologies, ethical challenges and other aspects of the profession' (2011b: 77) and lists the most popular functions within a sample of fifty translation blogs focusing on English, French and Spanish as: offering advice and opinions, sharing news, resources and personal experiences, asking for feedback and sharing material of a personal nature that is not directly linked to translation (2011b: 86). In this way McDonough Dolmaya identifies translation blogs as an important means of studying sociological aspects of translation (2011b: 91).

There may be up to a couple of dozen English-language translator networking sites. These fulfil the combined functions of directory, portal, discussion forum and marketplace. Possibly the best known is ProZ.com. Founded in 1999 and intended principally for freelance translators, it permits its members to post and/or bid for job offers, provide feedback on clients, collaborate on terminology (via KudoZ, the ProZ.com terminology service) and participate in on-line training (ProZ.com 2016). TranslatorsCafe.com, another highly prominent translator directory, is discussed by McDonough (2007: 804-11; see also Pym et al. 2013: 136-8).

The professionally-oriented social networking service LinkedIn is also popular among translators. LinkedIn enables users to form networks and increase their professional visibility, to market themselves and conduct more effective job

searches, or to advertise for and recruit staff. LinkedIn Groups bring together people with shared purposes or interests, such as networking with clients, sharing advice about translation agencies or discussing specialist areas of translation.

3.10 Fansubbing and related practices

Fansubbing is based on a notion of fandom that builds on the concept of the 'prosumer' (or person who is simultaneously a 'producer' and a 'consumer').

In the early 1990s fansubbers worked with VHS copies of Japanese anime series, imprinting the subtitles in the video and distributing tapes by post (Lee 2011: 1138). Since 2000 the process has gone fully digital, with the 'raw' (i.e. unsubtitled) source usually being obtained from DVDs, television broadcasts, peer-to-peer networks or contacts in Japan or elsewhere (Hatcher 2005; Díaz Cintas & Muñoz Sánchez 2006); finished products are likewise typically distributed using the peer-to-peer file-sharing protocol (Lee 2011: 1137; e.g. via sites such as http://myanimelist.net) or else viewed directly over the Web (e.g. at http://www.crunchyroll.com). The subtitling tool of choice for most fansubbers is Aegisub (http://www.aegisub.org).

Fansubbers are unpaid volunteers, motivated by the desire to share with fellow fans new material from the genres that interest them. Typically they are not trained in subtitling but compensate for this through detailed domain knowledge.

Fansubs can be of a very good quality (Hatcher 2005), although they have been observed to differ from the output of professional subtitlers: for example, they tend to be longer and more oriented towards ST norms, and they contain more

features specific to spoken language and also more language errors (Wilcock 2013: 103-6). Furthermore, much modern fansubbing output pursues a relatively foreignising translation approach, with over 90% retaining significant numbers of honorifics (e.g. 'san', 'sensei') and Japanese words and even displaying explanatory notes on the screen simultaneously with the subtitles (OtaKing77077 2008). However, it has exceeded official translation in terms of the range and variety of material that is subtitled, which has traditionally included titles that would not otherwise have been available, because they are either not well known or intended for a niche audience (Lee 2011: 1138; Hatcher 2005).

While technically illegal, fansubbing has been tolerated by the licensed distributors, for a number of reasons: a fansubbed version can help draw a tentative audience towards a new anime series, while the work of fansubbers can help distributors to decide what to license as well as bringing new material – and indeed new translation and subtitling talent – to their attention (Hatcher 2005).

The related phenomenon of scanlation is the process of scanning Japanese manga, Chinese manhua, Korean manhwa and other genres of comic book, translating them and making them available via the Web (Lee 2011: 1132; see also Lee 2009). Like fansubbing, scanlation is an amateur activity that is almost always carried out without the copyright holder's permission.

O'Hagan argues that the kind of UGT produced by fansubbers and scanlators is likely to have far-reaching consequences for the translation profession, with

individual users and consumers becoming empowered as they take their place within a rapidly evolving community of practice (2009: 115). Fans cease to be seen as 'helpless victims of mass culture', taking on instead more positive attributes through their engagement in different forms of resistance (Chu 2013: 260).

3.11 Web localisation

One definition of localisation – albeit one that favours the business aspects of the activity – sees it as 'the process of modifying products or services to account for differences in distinct markets' (LISA 2007: 11).

The business imperative behind the drive to localise websites and other digital products (such as software, video games and apps) is the geographical distribution of internet users, the preference on their part for an internet experience in their own language and the often excellent return on investment that localisation can yield (Internet World Stats 2015b; Globalization and Localization Association 2016). However, other types of material – such as personal homepages or not-for-profit websites – may also require localisation.

Localisation is one of the linked 'GILT' processes of globalisation, internationalisation, localisation and translation, which play a role in adapting e-content to different linguistic, cultural and business settings. These processes are frequently referred to by abbreviations – G11N, I18N, L10N and T9N – that are based on each term's letter count.

Globalisation involves the organisational preparations that a company needs to put in place prior to going global (Jiménez-Crespo 2013: 24).

Internationalisation, on the other hand, entails 'enabling a product at a technical level' (LISA 2007: 17) in order to prevent the need for substantial redesign during the localisation process; this can include allowing space for text expansion and ensuring that the language is as culture-neutral as possible. Once these stages have been completed the website can be localised for use in the target country. This involves replacement of all 'user-visible natural-language strings' (Pym 2010: 134) with corresponding TL text strings (translation) and adaptation of date and number formats (part of a website's 'locale'), colour scheme, appearance, layout, cultural specificity and conventions for interacting with the interface, making use all the appropriate technology in the process (localisation). While cultural adaptation often forms a major part of localisation, the internationalisation process can be used to ensure that the presence of features requiring such adjustment is kept to a minimum.

The task of converting the text tends to be outsourced to freelance translator-localisers. Besides being bilingual and bicultural, Jiménez-Crespo describes localisers as translators with 'an expandable degree of technological and management competence' (2013: 165), including, for example, an understanding of mark-up languages such as HTML and XML, Cascading Style Sheets and scripting languages such as JavaScript and PHP, as well as an ability to conduct QA and other procedures.

Representatives of the localisation industry tend to characterise translation as a mere stage in the complex localisation process (see for example Nichols 2015), and localisation and related processes of adaptation are increasingly distanced from 'traditional translation' through the use of terms and concepts such as

'glocalisation' (Mazur 2007) and 'transcreation' (Bernal Merino 2006: 32; Kelly 2014b; Nichols 2015). In contrast, the translation scholar Gouadec states that '[t]o all intents and purposes, localisers are a particular category of translators who "translate" material that is embedded in media other than paper or print or audiovisual media' (2007: 114).

4 MAIN RESEARCH METHODS

The subject of language, translation and the Web is extremely varied, taking in a large number of individual topics. Large amounts of data useful for researching the Web are available on various sites (e.g. http://www.internetlivestats.com, http://www.internetworldstats.com, http://www.alexa.com and https://archive.org) or, for example, for Wikipedia within the encyclopaedia itself (e.g. 'Wikipedia:Translation' 2015, 'Category:Translated pages' 2016, 'List of Wikipedias' 2016 and 'List of Lists of Lists' 2015) so that the principle of using the Web to study the Web, or using Wikipedia to study Wikipedia, seems to apply. Theoretical approaches to the study of language on the Web are drawn from many different branches of linguistics, while when it comes to translation all three of Holmes' original research focuses – i.e. product, process and function (2004: 72-3) – are brought into play.

5 CURRENT DEBATES AND FUTURE DIRECTIONS

The Web is expected to continue to evolve rapidly and the direction of future research will depend on new developments within its object of study. That said, research paradigms need to be elaborated in greater detail in nearly every area that has been discussed, our current knowledge of most of these being limited.

6 IMPLICATIONS FOR PRACTICE

The above discussion highlights at least two potential challenges to traditional translation practices, namely MT and crowdsourced translation, both of which have had a significant impact on the translation industry. Furthermore, for certain sectors of the industry, translating 'the wiki way' is becoming an increasingly attractive and popular option (Translation the Wiki Way n.d.). Besides this, there is likely to be an increasing use of corpora among translators although the reliability of data accessed via the Web is likely to remain an issue.

7 FURTHER READING

Ayers, Phoebe, Charles Matthews & Ben Yates (2008) *How Wikipedia works and how you can be a part of it*, San Francisco: No Starch Press, Inc., available at http://en.wikibooks.org/wiki/How_Wikipedia_Works (accessed 29 February 2016).

An excellent all-round introduction to Wikipedia.

Crystal, David (2011) *Internet linguistics: A student guide*, Abingdon & New York: Routledge.

An introductory overview of the study of internet linguistics.

Gatto, Maristella (2014) *Web as corpus: Theory and practice*, London: Bloomsbury.

A comprehensive and detailed study of how Web content can be exploited and studied using corpus linguistics methodologies.

Jiménez-Crespo, Miguel A. (2013) *Translation and web localization*, Abingdon:

Routledge.

A detailed overview of the theory and practice of web localisation.

O'Hagan, Minako (ed.) (2011) *Translation as a social activity*, *Linguistica*

Antverpiensia New Series: Themes in translation studies 10/2011, available at

https://lans-tts.uantwerpen.be/index.php/LANS-TTS/issue/view/14 (accessed

1 March 2016).

A very useful collection of articles on various aspects of collaborative translation.

8 RELATED TOPICS

Computation, language and translation, Corpus linguistics, translation and

interpreting, Genre Analysis and translation, Language and translation in film,

Translation, Interpreting and New Technologies.

9 REFERENCES

Aiken, Milam & Shilpa Balan (2011) 'An Analysis of Google Translate Accuracy',

in *Translation Journal* 16:2, available at

http://translationjournal.net/journal/56google.htm (accessed 5 October 2015).

Allen, Jeffrey (2003) 'Post-editing', Chapter 16 in Harold Somers (ed.)

Computers and Translation: A Translator's Guide, Amsterdam and Philadelphia:

John Benjamins, 297-318.

Alm, Antonie (2006) 'CALL for autonomy, competence and relatedness:

Motivating language learning environments in Web 2.0', in *The JALT CALL*

Journal 2:3, 29-38, available at http://journal.jaltcall.org/articles/2_3_Alm.pdf

(accessed 24 February 2016).

American Translators Association (2016) 'American Translators Association (ATA) Blog Trekker', available at http://www.atanet.org/resources/blog_trekker.php (accessed 11 February 2016).

Ayers, Phoebe, Charles Matthews & Ben Yates (2008) *How Wikipedia works and how you can be a part of it*, San Francisco: No Starch Press, Inc., available at http://en.wikibooks.org/wiki/How_Wikipedia_Works (accessed 29 February 2016).

Barr, Chris & the Senior Editors of Yahoo! (2011) *The Yahoo! Style Guide: The ultimate sourcebook for writing, editing and creating content for the Web*, London: Macmillan (Kindle Edition).

Bernal Merino, Miguel (2006) 'On the translation of video games', in *The Journal of Specialised Translation*, No. 6, 22-36, available at http://www.jostrans.org/issue06/art_bernal.php (accessed 10 December 2015).

Bernardini, Silvia, Marco Baroni & Stefan Evert (2006) 'A WaCky introduction', in Marco Baroni & Silvia Bernardini (eds) *WaCky! Working papers on the Web as Corpus*, Bologna: Gedit Edizioni, available at http://wackybook.sslmit.unibo.it (accessed 26 February 2016), 9-40.

BootCaT (2015) 'Simple utilities to Bootstrap corpora and terms from the Web', available at http://bootcat.sslmit.unibo.it (accessed 27 February 2016).

Category:Articles needing translation from foreign-language Wikipedias (2016) Available at https://en.wikipedia.org/wiki/Category:Articles_needing_translation_from_foreign-language_Wikipedias (accessed 28 January 2016).

Category:Translated pages (2016) Available at
https://en.wikipedia.org/wiki/Category:Translated_pages (accessed 28
January 2016).

Chitu, Alex (2007) 'Google switches to its own translation system', post dated 22
October 2007, in *Google operating system: Unofficial news and tips about
Google*, available at http://googlesystem.blogspot.co.uk/2007/10/google-
translate-switches-to-googles.html (accessed 23 October 2015).

Chu, Donna S. C. (2013) 'Fanatical labour and serious leisure: A case of
fansubbing in China', in Lee, Francis L. F., Louis Leung, Jack L. Chu & Donna S.
C. Chu (eds) *Frontiers in New Media Research*, New York & Abingdon:
Routledge, 259-77.

Consalvo, Mia & Charles Ess (eds) (2011) *The Handbook of Internet Studies*,
Hoboken: Wiley-Blackwell.

Crystal, David (2006) *Language and the Internet*, 2nd Edition, Cambridge:
Cambridge University Press.

Crystal, David (2011) *Internet linguistics: A student guide*, Abingdon & New
York: Routledge.

DePalma, Donald A. (2009) 'dotSUB Extends TED's "Ideas Worth Spreading" to
Hundreds of Languages', post dated 14 May 2009, in *Global Watchtower*,
available at
http://www.commonsenseadvisory.com/Default.aspx?Contenttype=ArticleDetA
D&tabID=63&Aid=553&moduleId=391 (accessed 9 January 2016).

DePalma, Donald A. (2011) 'The largest translation project…so far', post dated
13 January 2011, in *Global Watchtower*, available at
http://www.commonsenseadvisory.com/Default.aspx?Contenttype=ArticleDetA
D&tabID=63&Aid=1180&moduleId=391 (accessed 24 January 2016).

Désilets, Alain & Jaap van der Meer (2011) 'Co-creating a repository of best-practices for collaborative translators', in Minako O'Hagan (ed.), 27-45.

Díaz Cintas, Jorge & Pablo Muñoz Sánchez (2006) 'Fansubs: Audiovisual Translation in an Amateur Environment', in *The Journal of Specialised Translation* 6, 37-52, available at http://www.jostrans.org/issue06/art_diaz_munoz.pdf (accessed 12 April 2015).

Drugan, Joanna (2011) 'Translation ethics wikified: How far do professional codes of ethics and practice apply to non-professionally produced translation?', in Minako O'Hagan (ed.), 111-127.

Ellis, David (2009) 'A case study in community-driven translation of a fast-changing website', in Nuray Aykin (ed.) *Internationalization, Design and Global Development*, Berlin: Springer Verlag, 236-44, available at http://link.springer.com/chapter/10.1007%2F978-3-642-02767-3_26 (accessed 10 December 2014).

Enríquez Raído, Vanessa (2013) 'Introduction', Chapter 1 in *Translation and Web searching*, New York & Abingdon: Routledge, 1-7.

Facebook (n.d.) 'Translation App Guide', available at https://www.facebook.com/translations/guide/?cmsid=217893761755049 (accessed 9 January 2016).

Fernández Costales, Alberto (2013) 'Crowdsourcing and Collaborative Translation: Mass Phenomena or Silent Threat to Translation Studies?' in *Hermēneus. Revista de Traducción e Interpretación*, No. 15, 85-110, available at http://recyt.fecyt.es/index.php/HS/article/viewFile/30295/15892 (accessed 29 December 2015).

Fletcher, William H. (2007) 'Concordancing the Web: Promise and problems, tools and techniques', in Marianne Hundt, Nadja Nesselhauf & Carolin Biewer (eds), *Corpus linguistics and the Web*, Amsterdam & New York: Rodopi, 25-45.

Folaron, Deborah (2010) 'Web and Translation', in Yves Gambier & Luc van Doorslaer (eds) *Handbook of translation studies, Volume 1*, Amsterdam & Philadelphia: John Benjamins, 446-450.

Gabrilovich, Evgeniy & Shaul Markovitch (2006) 'Overcoming the brittleness bottleneck using Wikipedia: Enhancing text categorization with encyclopedic knowledge', in *Proceedings of the 21st National Conference on Artificial Intelligence (AAAI)*, 1301–1306, available at http://www.aaai.org/Papers/AAAI/2006/AAAI06-204.pdf (accessed 27 January 2016).

Garcia, Ignacio & Vivian Stevenson (2009) 'Google Translator Toolkit: Free web-based translation memory for the masses', in *MultiLingual* September 2009, 16-18.

Gaspari, Federico (2015) 'Online translation', in Sin-wai Chan (ed.) *Routledge encyclopedia of translation technology*, Abingdon & New York: Routledge, 578-593.

Gatto, Maristella (2014) *Web as corpus: Theory and practice*, London: Bloomsbury.

Gitsaki, Christina & Richard Paul Taylor (1999) 'Bringing the WWW into the ESL classroom', in Keith Cameron (ed.) *CALL & the learning community*, Exeter: Elm Bank Publications, 143-160.

Globalization and Localization Association (2016) 'Why Localize?', available at https://www.gala-global.org/industry/language-industry-facts-and-data/why-localize (accessed 18 February 2016).

Google Cloud Platform (2015) 'Translate API FAQ', available at

   https://cloud.google.com/translate/v2/faq (access 26 October 2015).

Gouadec, Daniel (2007) *Translation as a profession*, Amsterdam & Philadelphia:

   John Benjamins.

Hampshire, Stephen & Carmen Porta Salvia (2010) 'Translation and the

   Internet: Evaluating the Quality of Free Online Machine Translators', in

   Quaderns: Revista de traducció 17, 197-209, available at

   http://www.raco.cat/index.php/QuadernsTraduccio/issue/view/14558/showToc

   (accessed 10 December 2014).

Hatcher, Jordan S. (2005) 'Of otaku and fansubs: A critical look at anime online

   in light of current issues in copyright law', available at

   http://www2.law.ed.ac.uk/ahrc/script-ed/vol2-4/hatcher.asp (accessed 26

   November 2015).

Holmes, James S. (2004) 'The Name and Nature of Translation Studies', in

   Lawrence Venuti (ed.) *The translation studies reader*, 2nd edition, London:

   Routledge, 180-92. (First published 1972.)

Howe, Jeff (2006) 'Crowdsourcing: A Definition', in *Crowdsourcing Blog*,

   available at

   http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing_a.html

   (accessed 11 December 2014).

Hundt, Marianne, Nadja Nesselhauf & Carolin Biewer (2007) 'Corpus linguistics

   and the web', in Marianne Hundt, Nadja Nesselhauf & Carolin Biewer (eds),

   Corpus linguistics and the Web, Amsterdam & New York: Rodopi, 1-5.

Hutchins, John (2009) *Compendium of translation software: Directory of

   commercial machine translation systems and computer-aided translation

   support tools*, 15th edition, available at

http://www.hutchinsweb.me.uk/Compendium-15.pdf (accessed 14 October 2015).

Internet World Stats (2015a) 'Top ten internet languages – world internet statistics', available at http://www.internetworldstats.com/stats7.htm (accessed 22 February 2016).

Internet World Stats (2015b) 'World internet users statistics and 2015 world population stats', available at http://www.internetworldstats.com/top20.htm (accessed 18 February 2016).

Isabelle, Pierre, Marc Dymetman, George Foster, Jean-Marc Jutras, Elliott Macklovitch, François Perrault, Xiaobo Ren & Michel Simard (1993) 'Translation Analysis and Translation Automation,' in *TMI-93: The Fifth International Conference on Theoretical and Methodological Issues in Machine Translation*, Kyoto, Japan, 1133-1147, available at http://mt-archive.info/IBM-1993-Isabelle.pdf (accessed 1 December 2015).

Jiménez-Crespo, Miguel A. (2013) *Translation and web localization*, Abingdon: Routledge.

Kelly, Nataly (2009) 'Freelance Translators Clash with LinkedIn over Crowdsourced Translation' post dated on 19 June 2009, available at http://www.commonsenseadvisory.com/Default.aspx?Contenttype=ArticleDetAD&tabID=63&Aid=591&moduleId=391 (accessed 29 October 2015).

Kelly, Nataly (2012) 'Translators without Borders Prepares to Bridge the Last Language Mile', *The Huffington Post*, posted 12 March 2011, updated 2 February 2012, available at http://www.huffingtonpost.com/nataly-kelly/translators-without-borde_b_1122452.html (accessed 18 November 2015).

Kelly, Nataly (2014a) 'Why So Many Translators Hate Translation Technology', in *Huffington Post* 19 June 2014, available at http://www.huffingtonpost.com/nataly-kelly/why-so-many-translators-h_b_5506533.html (accessed 9 November 2015).

Kelly, Nataly (2014b) 'Where Transcreation Ends and Localization Begins', 23 July 2014, available at https://www.smartling.com/blog/transcreation-ends-localization-begins (accessed 10 December 2015).

Lee Hye-Kyung (2009) 'Between fan culture and copyright infringement: Manga scanlation', in *Media, Culture & Society* 31:6, 1011–1022, available at http://mcs.sagepub.com/content/31/6/1011.full.pdf+html (accessed 16 February 2016).

Lee, Hye-Kyung (2011) 'Participatory media fandom: A case study of anime fansubbing', in *Media, Culture & Society* 33:8, 1131-1147, available at http://mcs.sagepub.com/content/33/8/1131.full.pdf+html (accessed 12 April 2015).

Lertola, J. (2012) 'The effect of the subtitling task on vocabulary learning' in A. Pym and D. Orrego-Carmona (eds) *Translation research project 4*, Tarragona: Intercultural Studies Group, Universitat Rovira i Virgili, 61-70, available at http://isg.urv.es/publicity/isg/publications/trp_4_2012/5-Lertola.pdf (accessed 24 February 2016).

Liao, Han-teng & Thomas Petzold (2010) 'Analysing Geo-linguistic Dynamics of the World Wide Web: The Use of Cartograms and Network Analysis to Understand Linguistic Development in Wikipedia', in *Journal of Cultural Science*, 3:2, 1-18, available at http://cultural-science.org/journal/index.php/culturalscience/article/view/44/128 (accessed 27 January 2016).

LISA (2007) *LISA Globalization Industry Primer*, Romainmôtier, Switzerland: Localization Industry Standards Association.

List of Lists of Lists (2015) Available at https://en.wikipedia.org/wiki/List_of_lists_of_lists, date updated 17th July 2015 (accessed 21st July 2015).

List of Wikipedias (2016) Available at http://en.wikipedia.org/wiki/List_of_Wikipedias, date updated 22 February 2016 (accessed 23 February 2016).

Mazur, I. (2007) 'The metalanguage of localization: Theory and practice', in *Target* 19:2, 337-57.

McDonough, Julie (2007) 'How do Language Professionals Organize Themselves? An Overview of Translation Networks', in *Meta: Journal des Traducteurs / Meta: Translators' Journal*, 52:4, 793-815.

McDonough Dolmaya, Julie (2011a) 'The ethics of crowdsourcing', in Minako O'Hagan (ed.), 97-110.

McDonough Dolmaya, Julie (2011b) 'A Window into the Profession: What Translation Blogs Have to Offer Translation Studies', in *The Translator* 17:1, 77-104.

Meinel, Christoph, Justus Broß, Philipp Berger & Patrick Hennig (2015) *Blogosphere and its exploration*, Berlin Heidelberg: Springer.

Messina, Sebastiano (2001) 'E Palazzo Chigi creò l'inglese', *La Repubblica*, 8 December 2001, available at http://ricerca.repubblica.it/repubblica/archivio/repubblica/2001/12/08/palazzo-chigi-creo-inglese.html?ref=search (accessed 22 October 2015).

Meta:Translate Extension (2014) Available at

http://meta.wikimedia.org/wiki/Meta:Translate_extension, date updated 3rd

April 2014 (accessed 27 August 2014).

Mihalcea, Rada (2007) 'Using Wikipedia for AutomaticWord Sense

Disambiguation', in *Proceedings of HLT-NAACL 2007, Rochester, NY, April

2007*, Rochester, NY: Association for Computational Linguistics, 196–203,

available at http://www.aclweb.org/anthology/N07-1025 (accessed 27 January

2016).

Milne, David & Ian H. Witten (2008) 'Learning to link with Wikipedia', in

*Proceedings of the 17th ACM conference on information and knowledge

management, Napa Valley, California, USA, October 26-20, 2008*, New York,

NY, USA: ACM, 509-518, available at

http://www.cs.waikato.ac.nz/~ihw/papers/08-DNM-IHW-

LearningToLinkWithWikipedia.pdf (accessed 27 January 2016).

Moens, Marie-Francine, Li Juanzi & Tat-Seng Chua (2014) *Mining user generated

content*, Boca Raton: Chapman & Hall/CRC.

Morera-Mesa, Aram, J. J. Collins, & David Filip (2013) Selected crowdsourced

translation practices', delivered at *Translating and the Computer* 35, 28-29

November 2013, Paddington, London, available at http://www.mt-

archive.info/10/Aslib-2013-Morera-Mesa.pdf (accessed 24 January 2016).

Myers, Greg (2010) The discourse of blogs and wikis, London: Continuum.

Nichols, Brian (2015) 'The Difference between Translation and Localization for

Multilingual Website Projects [Definitions]', 28 July 2015, available at

http://content.lionbridge.com/the-difference-between-translation-and-

localization-for-multilingual-website-projects-definitions (accessed 10

December 2015).

O'Brien, Sharon, Michel Simard & Lucia Specia (2013) *Proceedings of the 2nd Workshop on Post-editing Technology and Practice (WPTP-2)*, MT Summit XIV Workshop, Nice, France, September 2, 2013, available at http://www.mt-archive.info/10/MTS-2013-W2-TOC.htm (accessed 6 November 2015).

O'Hagan, Minako (2009) 'Evolution of user-generated translation: Fansubs, translation hacking and crowdsourcing', in *The Journal of Internationalization and Localizatio*n, 1:1, 94-121, available at https://www.academia.edu/4462788/Evolution_of_User-generated_Translation_Fansubs_Translation_Hacking_and_Crowdsourcing (accessed 12 April 2015).

O'Hagan, Minako (ed.) (2011) *Translation as a social activity*, *Linguistica Antverpiensia New Series: Themes in translation studies* 10/2011, available at https://lans-tts.uantwerpen.be/index.php/LANS-TTS/issue/view/14 (accessed 1 March 2016).

O'Hagan, Minako (2011a) 'Community translation: Translation as a social activity and its possible consequences in the advent of Web 2.0 and beyond' in Minako O'Hagan (ed.), 11-23.

O'Hara, Kieron & Wendy Hall (2013) 'Web Science', in William H. Dutton (ed.) *The Oxford Handbook of Internet Studies*, Oxford: OUP, 48-68.

Olohan, Maeve (2014) 'Why do you translate? Motivation to volunteer and TED translation', in *Translation Studies* 7:1, 17-33.

Orr Priebe, Shelly (2009) 'Tom Sawyer – A crowdsourcing pioneer?', in *tcworld* November 2009, available at http://www.tcworld.info/e-magazine/outsourcing/article/tom-sawyer-a-crowdsourcing-pioneer (accessed 20 January 2016).

OtaKing77077 (2008) 'Anime Fansub Documentary PART 2', available at

https://www.youtube.com/watch?v=yoJ_BWQ9Kow (accessed 1 March 2016).

Petzold, Thomas (2012) '36 Million language pairs', in *Journal of Cultural Science*

5:2, 106-119, available at http://cultural-

science.org/journal/index.php/culturalscience/article/view/55/137 (accessed 2

November 2014).

Pimienta, Daniel, Daniel Prado & Álvaro Blanco (2009) 'Twelve years of

measuring linguistic diversity in the Internet: Balance and perspectives',

available at http://unesdoc.unesco.org/images/0018/001870/187016e.pdf

(accessed 22 February 2016).

ProZ.com (2016) 'ProZ.com – company overview', available at

http://www.proz.com/about (accessed 13 February 2016).

Pym, Anthony (2010) *Exploring translation theories*, Abingdon & New York:

Routledge.

Pym, Anthony, François Grin, Claudio Sfreddo & Andy L. J. Chan (2013) *The*

status of the translation profession in the European Union, London & New York:

Anthem Press.

Resnik, P. & N. A. Smith (2003) 'The web as a parallel corpus', in *Computational*

Linguistics, 29:3, 349–380.

Sargent, Benjamin B. (2008) 'Community Translation Lifts Facebook to Top of

Social Networking World', post dated 14 August 2008, in *Global Watchtower*,

available at

http://www.commonsenseadvisory.com/Default.aspx?Contenttype=ArticleDetA

D&tabID=63&Aid=525&moduleId=391 (accessed 10 January 2016).

Shen, Ethan (2010) 'Comparison of online machine translation tools', available

at http://www.tcworld.info/e-magazine/translation-and-

localization/article/comparison-of-online-machine-translation-tools (accessed 14 October 2015).

Shepherd, M. & C. Watters (1998) 'The evolution of cybergenres', in R. Sprague (ed.) *Proceedings of the XXXI Hawaii International Conference on System Sciences*, Los Alamitos CA: IEEE-Computer Society, 97-109.

Shuttleworth, Mark (2014) 'Approaches to language learning: Blending tradition with innovation', in *Proceedings of SILK conference, 14-16 February 2014, Hydro Hotel, Penang, Malaysia*, CD, available at https://ucl.academia.edu/MarkShuttleworth (accessed 24 February 2016).

Shuttleworth, Mark (2015a) '5th IATIS Conference Wikipedia presentation', *IATIS 5th International Conference*, Belo Horizonte, Brazil, 7-10 July 2015, slides available at https://www.academia.edu/13808267/5th_IATIS_Conference_Wikipedia_presentation (accessed 29 February 2016).

Shuttleworth, Mark (2015b) 'Wikipedia translation: Collaborativity, translation and the web', *ARTIS@UCL 2015 Multidimensional methodologies: Collaboration and networking in translation research*, UCL, UK, 15-16 June 2015, slides available at https://www.academia.edu/12982446/Wikipedia_translation_Collaborativity_translation_and_the_web (accessed 1 July 2015).

Smith, Jason, Herve Saint-Amand, Magdalena Plamada, Philipp Koehn, Chris Callison-Burch & Adam Lopez (2013) 'Dirt cheap Web-scale parallel text from the Common Crawl', in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, ACL 2013, 4-9 August 2013, Sofia, Bulgaria, Volume 1: Long Papers, 1374-1383.

Stejskal, Jiri (2010) 'Freelance Translators Talk about Machine Translation',

available at http://www.cetra.com/blog/freelance-translators-talk-about-

machine-translation (accessed 9 November 2015).

TED Conferences (n.d. a) 'TED Open Translation Project', Available at

https://www.ted.com/about/programs-initiatives/ted-open-translation-project

(accessed 9 January 2016).

TED Conferences (n.d. b) 'Get started', Available at

https://www.ted.com/participate/translate/get-started (accessed 21 January

2016).

Thomas, Jason (2015) 'The internet is killing the world's languages. Can Google

help?', available at http://bigthink.com/connected/language-extinction-google

(accessed 23 February 2016).

Thomas, Michael (ed.) (2009) *Handbook of research on Web 2.0 and second

language learning*, Hershey: Information Science Reference.

Tiedermann, Jörg (2012) 'Parallel Data, Tools and Interfaces in OPUS', in

*Proceedings of the 8th International Conference on Language Resources and

Evaluation (LREC 2012)*, 21-27 May 2012, Istanbul, Turkey, 2214-2218.

Timmis, Ivor (2015) *Corpus linguistics for ELT: Research and practice*, Abingdon

& New York: Routledge.

Translation the Wiki Way (n.d.) 'How will massive online collaboration impact the

world of translation', available at http://www.wiki-

translation.com/How+will+massive+online+collaboration+

impact+the+world+of+translation (accessed 1 March 2016).

Unbabel (2015) 'Top Languages of the internet, today and tomorrow', available

at https://blog.unbabel.com/2015/06/10/top-languages-of-the-internet

(accessed 22 February 2016).

Usability.gov (2015) 'Writing for the Web', available at

http://www.usability.gov/how-to-and-tools/methods/writing-for-the-web.html

(accessed 10 December 2015).

van der Meer, Jaap & Andrew Joscelyne (2013) 'Clarifying Copyright on

Translation Data', available at https://www.taus.net/think-

tank/articles/translate-articles/clarifying-copyright-on-translation-data

(accessed 12 December 2015).

W3Techs (2016) 'Usage statistics of content languages for websites, February

2016', available at

http://w3techs.com/technologies/overview/content_language/all (accessed 2

March 2016).

Wikipedia:Manual of Style (2016) Available at

https://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style, date updated 28

February 2016 (accessed 29 February 2016).

Wikipedia:Pages needing translation into English (2016) Available at

https://en.wikipedia.org/wiki/Wikipedia:Pages_needing_translation_into_Englis

h, date updated 28 January 2016 (accessed 28 January 2016).

Wikipedia:Translation (2015) Available at http://en.wikipedia.org/wiki/

Wikipedia:Translation, date updated 3 November 2015 (accessed 9 November

2015).

Wikipedia:Writing better articles (2016) Available at

https://en.wikipedia.org/wiki/Wikipedia:Writing_better_articles, date updated

16 February 2016 (accessed 29 February 2016).

Wilcock, Simone (2013) *A comparative analysis of fansubbing and professional

DVD subtitling*, unpublished MA Thesis, University of Johannesburg, available

at https://ujdigispace.uj.ac.za/handle/10210/8638 (accessed 14 February 2016).

Yang, Jin & Elke D. Lange (2003) 'Going live on the Internet', in Harold L. Somers (ed.) *Computers and TRANSLATION: A translator's guide*, Amsterdam & Philadelphia: John Benjamins, 191-210.

Zanettin, Federico (2002) 'DIY Corpora: The WWW and the translator', In Belinda Maia, Johann Haller & Margherita Ulrich (eds) *Training the language services provider for the new millennium*, Porto: Faculdade de Letras da Universidade do Porto, 239-48.

BIOGRAPHICAL NOTE

Mark Shuttleworth is interested in translation studies terminology, metaphor in translation, translation technology, translator training, medical translation and, more recently, translation and the Web. He has launched and run Masters programmes at both Leeds University and Imperial College London. He is currently based at University College London.