Visualizing Variation in Collections of Translations and Adaptations of Cultural Heritage Texts

Tom Cheesman
Swansea University
Dept of Languages
SA2 8PP
00 44 1792 604030
t.cheesman@swansea.ac.uk

Kevin Flanagan
Swansea University
Dept of Languages
SA2 8PP
00 44 1792 205678
kevin@kftrans.co.uk

Robert S Laramee
Swansea University
Dept of Computer Science
SA2 8PP
00 44 1792 205678
r.s.laramee@swansea.ac.uk

ABSTRACT

We describe how we went about designing visualization tools for exploratory access to collections of multiple translations of a literary work. Collections of this kind are relatively small. But big collections are collections of small collections. Visualizations should help us to explore on multiple scales, shifting smoothly between (say) metadata on millions of items to (say) investigating how ten or a few hundred closely similar items differ in detail. For example: a small set of differing translations and adaptations of a work in *one* language, from different times and places; or a set of these sets in *different* languages. Collections of translations and adaptations have great potential value in education, research, and creative practices. We are creating ways to explore such collections, prompting various kinds of "noticing" both in them, and in the translated work. Playful and exploratory approaches need to be combined with linguistics analyses and expert understandings of the texts and their contexts.

Categories and Subject Descriptors

H.5.2 User Interfaces; I.2.7 Natural Language Processing; I.3.8 Computer Graphics: Applications; J.5 Computer Applications: Arts and Humanities

General Terms

Algorithms, Measurement, Design, Experimentation, Human Factors, Languages.

Keywords

Visualization, translations, retranslations, adaptations, variation, collections, corpora, Shakespeare, Othello, German

1. INTRODUCTION

Collections of diverse comparable versions of a work, created by translators in different times and places, are a mine of information about past and present-day (inter/trans)cultural and linguistic changes, about translation as process and product, and about the interpretability of translated works. We aim to create applications which enable users to create and explore this kind of collection in ways which encourage cross-cultural curiosity and understanding.

1.1. A new kind of collection: *re*translations

'Cultural heritage texts' ([22]) are objects of study and curation, attributed to an individual writer (Aristotle, Balzac, Confucius, Darwin [11] … … Shakespeare …), or religious scriptures, attributed to God and/or various writers. The texts are typically unstable: extant in multiple versions in the original language of composition – repeatedly redacted, selected, curated – and even more unstable in other languages. Typically, the works, or selections, have not been translated only once per target (translating) language. In many languages, certain works have been and continue to be translated again and again. Usually, when one translation is made, new translations (*re*translations) are soon called for, which are supposedly more accurate, more appropriate for new generations or groups of users, or both ([7, 12, 27]). The quantity and qualitative diversity of retranslations depends on the historical (inter)cultural interests, opportunities and constraints of the relevant language communities and sub-communities.

Works of scripture are well known examples: multiple parallel aligned versions are accessible online (e.g. [14, 19]). But we have created a small Shakespeare translation collection. Translators of Shakespeare are less formally controlled than translators of scripture; they have more individual freedom, both at structural level (omission, re-ordering, addition: adaptation) and at micro-textual level (interpretation, style) ([8]). Individual (re)translators have their idiosyncracies, and vary in terms of skill and conscientiousness ([13]). Suprapersonal factors – the interests of commissioners, publishers, producers, readers, audiences, and other users, and socio-cultural, ideological constraints – strongly affect decisions about *what* and *how* to translate: format, style, text structure; whether to re-work a previous translation, work *against* it, or start from scratch; etc.

Retranslation collections are sets of documents which are comparable at structural and micro-textual levels. They are mines of cultural data, whether items are old and 'classic' (and have often been re-curated), or old and forgotten, or recent. Retranslation collections have been used to study the history of the translating language ([2]). We are more interested in histories of ideas, discourses, genres: socio-cultural traditions and changes.

1.2. A problematic kind of collection

Creating such a collection digitally is practically diffficult and conceptually problematic. Few have tried (but see [9]). Shakespeare translations and adaptations are very numerous in most European languages and several Asian languages. There is limited bibliographical documentation (but see [3, 10, 27]). Many retranslations are in any case unpublished: scripts for productions,

manuscripts or typescripts. Texts actually performed also vary from one live performance to another. Texts can in principle be transcribed from sound recordings of stage, radio, film, television or video performances. Fragments of Shakespeare's work circulate far more widely than whole texts: quotations and allusions (see [26]). Critical and pedagogical commentaries, reference works, biographies and other treatments, films, other media and artforms are as important as the literary texts are, in the general process of 'intercultural transfer' of Shakespeare's works ([17]). Many aspects of intercultural transfer are *not* captured by studying translation texts only. But translations are key documents and vehicles of transfer, and they are uniquely amenable to machine-assisted comparative study using visualization tools.

1.3 Our collection

We created a digital collection of c.40 German-language versions of Shakespeare's play, *Othello* (c.1604). They date from 1766 to 2010: mostly printed editions, with a dozen unpublished typescripts (or pdfs); structurally conservative translations, except for eight adaptations (with major omissions, additions, re-orderings). Before adopting a digital approach, Cheesman studied versions of one short extract (a string of 14 words) and how these versions convey changing ideologies on the topics of 'race', gender, and political power ([4, 5]). He is also crowd-sourcing a multilingual collection of versions of that speech-string ([6]).

1.4. Initial questions and hypothesis: extracting 'Rich Points' from variation

Shakespeare's 500-year-old language poses many problems of understanding for modern English readers, even experts. The copious annotations in modern editions help readers solve some of those problems. For translators, further problems arise as they must interpret Shakespeare's text and express that interpretation appropriately. For example, the short speech-string Cheesman studied is rich in wordplay and ambiguity. Translators must express a selective interpretation of it, so they reveal their ideological tendencies. Any piece of language to be translated may contain such densely significant, diversely interpretable moments. These moments have been called 'Rich Points' ([1]). The concept of Rich Points "aims at a sophisticated *theory of noticing*" which we still lack ([1], p.687). Our visual interfaces aim to make such "noticing" more enjoyable and productive.

Rich Points for translation may not be the moments in a text which pose problems for interpretation in the source culture. Certain Rich Points may be noticed only by some translators: those working in particular translating languages, at particular periods, or with particular purposes. It is hard to predict which parts of a work are Rich Points for translation. But given a collection of translations, we can use an algorithm to point us towards likely candidates.

We assumed: (a) that translators in a given language-culture ought to notice the same Rich Points, so their work can be queried collectively; (b) that the symptom of a Rich Point, in translations, is that translators translate more variously at Rich Points than they do otherwise, because they express a different interpretation.

On the basis of these intuitions we devised a way of quantifying and visualizing how much each translator has translated differently from (all or selected) others, at each point in the text. We designed a visual interface for exploring this phenomenon (3.3 below). In principle it can be used with any collection of comparable, part-parallel aligned versions of works.

2. DATA AND 'BACK-END' APPLICATION

First, we created an aligned part-parallel corpus of 37 German versions of *Othello* 1.3 (= Act 1, Scene 3: c.10% of the text; with limited time and money, we prepared parallel fragments of many versions, rather than a few full versions.) We had to: digitise the texts; manually clean OCR output; normalise text layout; code speech-prefixes, stage directions, speech text; meanwhile build an application for defining, tagging and (part-automatically) aligning segments (all speeches and selected sub-speech word-strings) between each version and a 'base text' (a curated English text of the play); also build a database to store all the texts together with segmentation and alignment information. No tool for constructing and querying a one-to-many part-parallel text collection existed. This application was developed by Flanagan. It comprises the 'Prism' segmentation and alignment tool, and the 'Ebla' database. These constitute the 'back end' of the toolset, released in 2012 at www.delightedbeauty.org/vvv (fully open installation: guests can alter data) and at www.delightedbeauty.org/vvvclosed (research installation: permission required to alter data).[1]

3. 'FRONT-END' VISUALIZATIONS

The 'front end' of the toolset, designed by Studio Nand, affords access to the texts through experimental visual interfaces. Three prototype exploratory visualizations were built.

3.1 Time-map

A scalable map with time-slider ([24]) affords an overview of the collection's metadata and access to descriptions and texts, including two geographical metadata-points per version (visually linked places of creation and of publication). A future iteration can show places of performance and other historical diffusion, for any collections of versions.

3.2 Alignment Maps and parallel navigation

Alignment Maps visualize texts comparably in terms of their
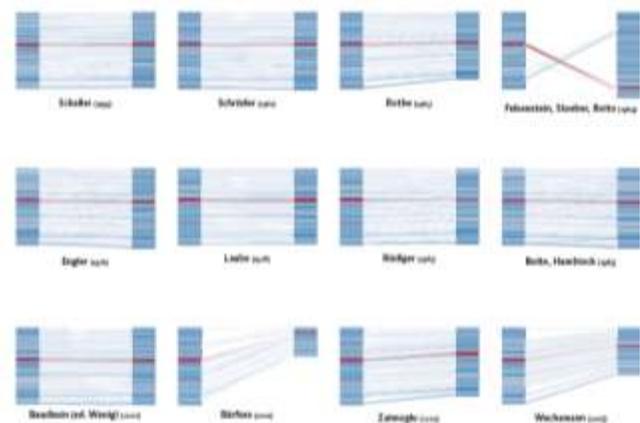
Figure 1: Alignment Maps (12 x *Othello* 1.3 in German)

partly aligned segment structures (Fig.1). This visual approach is generalisable to any collections of comparably structured items. Here the left-hand vertical strip in each pair represents the speech-

structure of our English base text; play texts are structured as series of speeches by speakers (character roles), interspersed with stage directions. Here each right-hand strip represents the speech-structure of a German version. Each horizontal bar in a strip represents a speech; its thickness denotes length in words. Connecting lines denote alignments. We can see at a glance where versions expand, contract, omit, add, or re-order, relative to the base text, and to each another.

Figure 2: Dynamic parallel text navigation

The vertical strip representing speech-structure in Fig.1 becomes a navigational aid for close reading in the dynamic parallel view (Fig.2). Here each bar is coded for the speaker. Users can filter to highlight all speeches by a speaker, ordered by speech length or text flow; click on a bar to dynamically scroll the view to that part of the text; and re-synchronise parallel texts at a click.

We still need a flexible view on *multiple* parallel texts. Studio Nand has produced sketches for such an interface ([23, 25]), which re-scales (zooms) smoothly between close reading views and distant comparative views of structure and selected analytic features, allowing for multiperspectival filtering, sorting, and on-demand feature highlighting. A mid-scale view is shown (Fig.3).
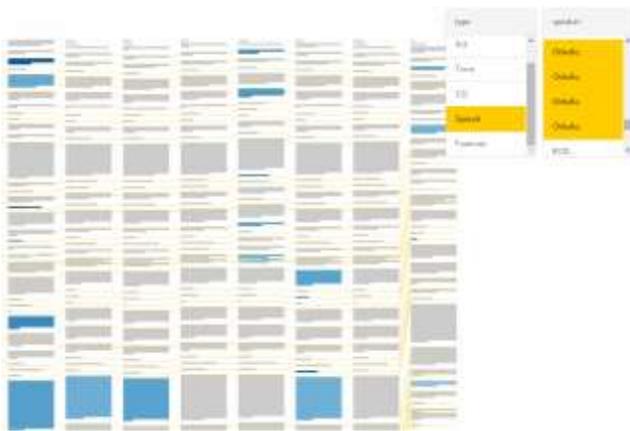
Figure 3: Sketched multiple parallel text view

3.3. "Eddy" and "Viv" – exploring variation, finding "Rich Points"

In our prototype "Eddy and Viv" interface, users scroll the English base text (Fig.4, left column) and click any selected segment (i.e. any speech, or any word-string defined as a segment in Prism) in order to call up all (or a selected subset of) aligned

translations of that segment (Fig.4, right column; machine 'back'-translations are also retrieved). The algorithm "Eddy" sorts all the segment translations in order of their relative predictability, based on the words used in each. "Eddy" builds concordances for each version, calculates a notional average concordance, and finds the deviation of each concordance from the average. Low "Eddy" values indicate segment translations which have many words in common with others, i.e. is closer to 'normal', or more predictable. High "Eddy" values indicate more 'abnormal' or unpredictable translations. (For details see [6].)

The algorithm "Viv" ('variability in variation') finds the average of "Eddy" values for each base text segment (adjusting for the variation associated with differences in segment lengths – see [6]).
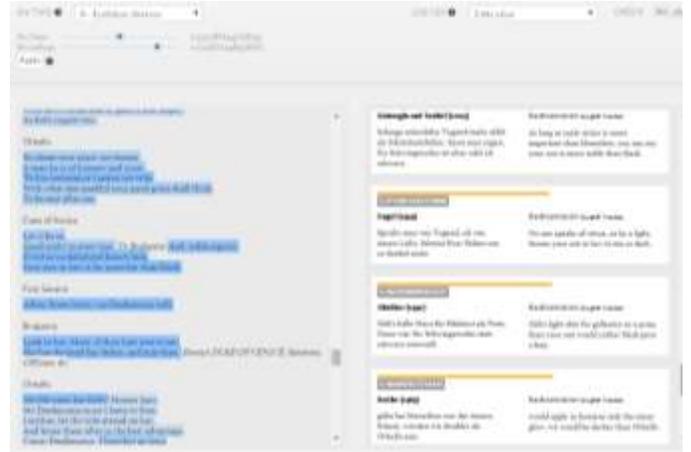
Figure 4: 'Eddy and Viv' interface

"Viv" values are mapped onto base text segments in the display as a varying colour background. Darker colour means a higher "Viv" value, i.e. relatively greater variation (instability, unpredictability) among translations of that segment. Segments with high "Viv" values are algorithmically derived candidates for Rich Points in the base text, as read by multiple translators.

Eddy and Viv also find "Poor Points": segments where most translators use more than averagely *similar* wording. Unexpectedly, these are also interesting: the indicated moments of consensus can be counter-intuitive to domain experts.

The metrics currently used in the prototype produce noisy results. But already "the interface prompts various kinds of 'noticing' and encourages an essentially playful and exploratory approach to the 'data'" ([18], cf. [19]).

4. REFLECTIONS

Designing, building, experimenting with and re-designing such interfaces is a many-staged iterative process. We have many desiderata for future iterations.

More various overviews of a collection, its subdivisions by metadata values, and the intrinsic differences and relations among items are desirable. Visualizations of comparative algorithmic stylometric analyses are good at identifying genre and period clusters, and similarities which may be due to intertextual dependency (borrowing, influence, plagiarism) ([21]). Word co-occurrence cluster networks ([16]), topic modeling ([13]), and other comparative analyses using natural language processing (NLP) tools may discover further interesting differentiations and relations among (part)parallel versions.

Users should be able to experiment with text deformations, e.g. lemmatisation, synonym bundling, stopword culling, and variant spelling normalisation. Such NLP operations can't just be entrusted to machines: users need to work manually on texts and/or word lists. In exploring the collection, users could then learn not just about translation, languages, literatures and cultural histories, but also about different modes of manipulation and exploration of text collections, through practice.

Users should be able to investigate parallel chunk selections, e.g. in *Othello* the different uses of language in different character roles; also vertical selections of keywords, semantic fields, features of discourse, rhetorical tropes, metaphors, idioms, etc. This means more facilities for machine-assisted annotation of aligned texts. User annotation must also include flagging inaccurate or debatable metadata, text data, segment definitions and alignments. Users should be empowered to edit machine translations. Since 'back-translations' are just as provisional as any translations are, this could be one fruitful way of bringing users into dialogue around such collections. Of course, we want to align comparable collections in multiple languages, which raises interesting questions for benchmarking variation in applications like the Eddy and Viv interface.

Users can upload, segment and align corpora (collections), but this should be made easier. We are also contemplating an application for translation practice, where groups of learners' translations of assigned texts constitute collections, and visualizations support review and assessment by teachers, peers, and students. Though very different from collections of pre-existing versions of cultural heritage works, collections of trainee translators' texts would also be interesting for researchers. The two kinds of collections can converge. We envisage a collection of many comparable collections. Eventually it might be a big one.

5. REFERENCES

[1] Agar, M. 1999. How to ask for a study in Qualitatisch. *Qualitative Health Research* 9/5 (1999): 684-697.

[2] Altintas, K. et al. 2007. Language change quantification using time-separated parallel translations. *Literary and Linguistic Computing* 22/4 (2007): 375-393.

[3] Blinn, H. and W.G.Schmidt. 2003. *Shakespeare – deutsch: Bibliographie der Übersetzungen und Bearbeitungen*. Erich Schmidt Verlag.

[4] Cheesman, T. 2012. Thirty times 'more fair than black': *Othello* re-translation as political re-statement. *Angermion* 4 (2012): 1-52. DOI= 10.1515/9783110239201.1

[5] Cheesman, T. et al. 2010--. Delighted beauty: multilingual crowdsourcing. At: www.delightedbeauty.org.

[6] Cheesman, T., K.Flanagan and S.Thiel. 2012-13. Translation Array Prototype 1: project overview. At: www.delightedbeauty.org/vvv.

[7] Delabastita, D. Shakespeare. 2009. In *Routledge Encyclopedia of Translation Studies*, M. Baker and G. Saldanha Eds. Second edition. Routledge, 263-269

[8] Delabastita, D. 2012. Shakespeare and translation. a guide to further reading. In *Shakespeare and the Language of Translation*, T.Hoenselaars Ed. Revised edition. Arden Shakespeare, 289-316.

[9] Drabek, P. et al. 2003-10. *Kapradi: Czech translations of Shakespeare*. At: http://www.phil.muni.cz/kapradi/.

[10] Engler, B., et al. 2003--. *Sh:in:E – Shakespeare in Europe*. At: http://shine.unibas.ch.

[11] Fry, B. 2009. *The preservation of favoured traces*. At: http://benfry.com/traces.

[12] Gürcaglar, S.T. 2009. Retranslation. In *Routledge Encyclopedia of Translation Studies*, M. Baker and G. Saldanha Eds. Second edition. Routledge, 233-236.

[13] Jockers, M. 2013. *Macroanalysis: Digital Methods and Literary History*. U Illinois Press.

[14] Johansson, S. 2011. Between Scylla and Charybdis: on individual variation in translation. *Languages in Contrast* 11/1 (2011): 3-19. DOI= 10.1075/lic.11.1.03joh

[15] LifeChurch.tv. 2013--. YouVersion (online Bible resource). At: www.bible.com.

[16] Meara, P.M. 2014. Texts as vocabulary networks. Paper at Formulaic Language Research Network Conference, Swansea University.

[17] Pym, A. 2010. *Translation and Text Transfer: An Essay on the Principles of Intercultural Communication*. Revised edition. Tarragona: Intercultural Studies Group. At: http://usuaris.tinet.cat/apym/publications/TTT_2010.pdf.

[18] Ramsay, S., 2014. Personal communication to Tom Cheesman.

[19] Ramsay, S., 2011. *Reading Machines. Toward an Algorithmic Criticism*. Second edition. U Illinois Press.

[20] Royal Aal al-Bayt Institute for Islamic Thought. 2002--. Online Quranic resource. At: www.altafsir.com.

[21] Rybicki, J. 2014. Visualizing literature: artistic statistics. In *The Art of Literature, Art in Literature*, B.Bucała et al. Ed. Kraków: WUJ.

[22] Schmidt, D. A. 2010. The inadequacy of embedded markup for cultural heritage texts. *Literary and Linguistic Computing*, 25/3 (2010), 337-356.

[23] Thiel, S. 2014. *Visualizing Translation Variation. Designing Tools for Literary Scholars inTranslation Studies and Linguistics*. Masters thesis. Bauhaus University Weimar.

[24] Thiel, S. and Studio Nand. 2013. *Othello time map*. http://othellomap.nand.io.

[25] Thiel, S. and Studio Nand. 2014. *Visualizing translation variation*. At: www.tinyurl.com/transvis2014.

[26] Trillini, R.H. et al. 2010--. *HyperHamlet*. At http://www.hyperhamlet.unibas.ch/.

[27] UNESCO. 1938--. *Index Translationum*. At: http://portal.unesco.org/culture/en/ev.php-URL_ID=7810&URL_DO=DO_TOPIC&URL_SECTION=201.html.

[28] Venuti, L. 2004. Retranslations: the creation of value. In *Translation and Culture*, K.M.Faull Ed. Bucknell UP, 25-38.