Originalveröffentlichung in: Gippert, Jost/ Gehrke, Ralf (Hrsg.):
Historical Corpora. Challenges and Perspectives.–
Tübingen: Narr, 2015. S. 77-89. (Korpuslinguistik und interdisziplinäre Perspektiven auf Sprache 5)

TIMOTHY BLAINE PRICE

Multi-faceted alignment

Toward automatic detection of textual similarity in Gospel-derived texts

Abstract

Ancient Germanic Bible-derived texts stand in as test material for producing computational means for automatically determining where textual contamination and linguistic interference have influenced the translation process. This paper reports on the results of research efforts that produced a text corpus; a method for decomposing the texts involved into smaller, more directly comparable thematically-related chunks; a database of relationships between these chunks; and a user-interface allowing for searches based on various referential criteria. Finally, the state of the product at the end of the project is discussed, namely as it was handed over to another researcher who has extended it to automatically find semantic and syntactic similarities within comparable chunks.

1. Introduction

The research and output discussed in this paper took place at the Goethe Universität Frankfurt during a period of over two years, from June 2011 to September 2013, as part of the LOEWE research cluster “Digital Humanities”. Formally named ‘Historische Wechselbeziehungen altgermanischer Sprachen’, the primary goal of the research project was to investigate methods for aligning and comparing old Germanic texts that historically related yet vary due to translation and, in particular, to reconfiguration through artistic retelling, as is the case with the Old Saxon *Heliand*.

Aligning like parts of texts is a primary stumbling block to any textual comparison. For texts that differ minimally, basic comparison based on structural differences can be performed by various popular software options, e.g. Microsoft Word or Adobe Acrobat, which can easily show where sections of text between two files match and/or have been altered. However, the scope of the project discussed here was to expand such capabilities to allow for comparison between texts that are less superficially related (different time periods, distinctive styles, and variant languages).

1.1 The plagiarism model

Plagiarism detection software is able to return the probability that one (chunk of) text has been copied from another. Once a certain threshold of similarity in the two texts is passed, the overall resemblance is considered "not due to chance," suggesting material was copied. By analysing word combinations, i.e. *n-grams*, the computer compares arbitrarily sized chunks of text between the two documents. One way to extend these capabilities is to add syntactic parsing capabilities and lexica, allowing for analysis based on synonymy.

Plagiarism software is not absolutely perfect. Beyond a threshold of complexity, the problem becomes mathematically taxing. Increasing the number of texts to be compared also does this. Other complications can easily develop simply by removing certain assumptions, e.g. that a text was written by a single author.

1.2 Direct borrowing vs. generic influence

Multi-authored texts occur readily: citation of external works technically creates a multi-authored work. In the best-case scenario, an author gives credit to the original author. Yet academics and recent politics have shown us that "borrowing" is common and not always cited responsibly.

Even when an author is honest, it may still be possible that his writing was inspired by others in ways of which he is not fully aware. Unless he directly says so, it is difficult to know whether an author has come into contact with the works of another. This is important for anyone interested in the history of ideas.

Copying is a multifaceted problem: there exists a spectrum ranging from *copying-and-pasting*, through *paraphrasing*, toward *idea theft*. Each step rightward along this spectrum involves linguistic alternations that become increasingly more difficult to track. Between paraphrasing and idea theft is the realm of *translation*, from which yet another axis branches off.

1.3 Translation as "plagiarism"

Determining similarity between two texts in the same language is one thing; doing it between different languages is quite another. Techniques that aid such comparisons include employing sophisticated lexica and pre-determined divisional similarities, the latter being the scope of this paper.

Note, two texts that differ only in language are not normally considered plagiarism *per se*, rather translation. Nevertheless, the two practices are similar. A plagiarist rewords another's writing to fit a situation making the ideas look like his own. A translator also rewords another person's writing into a new situation – a new cultural realm. The difference tends to be a moral one. Still, the new expression of translator and plagiarist contains both copied and novel material.

Normally, a translator strives not to add to or remove from the information stream of the original. In reality, languages differ not only in the word forms used, but also in grammatical structure and in cultural context. Thus, translations will differ from their original to a variable degree depending on these linguistic and cultural differences. It is difficult to develop computational means to handle such flexibility in translations while still finding similarity between texts. In other words, one might consider software that recognizes a translation and its original to be one step beyond plagiarism software. It is no longer looking for common structures based only on *n-grams*; it is also simultaneously measuring these possibilities against the options allowed in the target language. One now needs to be able to handle another axis of potentially infinite complexity.

1.4 Highly disruptive alterations

There are likely many linguistic axes that might break the plagiarism detection model. One is style, i.e., authorial creativity. Through 'translation' we understand the transfer of information between languages with minimal alteration – semantic integrity trumps syntagmatic freedom. Yet not all translated texts fit this description: the assumption of meaning-over-form is likely too restrictive to describe all of what translation entails. Consider musical translation, where lyrics are more variable than a set melody, e.g. syncopation. In this case, the musical structure outweighs information provided by text. Creativity is allowed in semantic translation – and also more variance from the original intentions. A similar phenomenon occurs in translation of non-prose, where a translator may be forced to sacrifice certain poetic features of the original in order to focus on just one, e.g. end-rhyme over semantics.

Automatic text comparison faces this conundrum: texts often display disruptive alterations among otherwise recognisable similarity. In texts where multiple changes to an original have been undertaken simultaneously, *n-gram* permutations are no longer the only concern. Grammatical limitations in the target language combine with poetic patterns affecting both word order and word choice. It may be necessary to deal with each of these challenges in piecemeal fashion, e.g. by limiting the amount of material being compared. Evaluating one extensive text with another is probably too expensive. Thus, for our project, where text sizes reach tens of thousands of lines, I have cut each into more manageable chunks. Where these cuts are to be made depends greatly upon each text.

2. Textual resources

The translations and derivative Biblical texts employed here are not without controversy. On the positive side, this reflects an interest within academia. Determining textual contamination within these texts is important toward answering questions regarding the texts' provenance.

2.1 Old Saxon *Heliand*

The Old Saxon (OS) *Heliand* was presumably written by an anonymous 9th-century author. Yet even this date is speculation assuming other resources must have been used by the author. Furthermore, material evidence from manuscripts gives only vague ideas about the circumstances of the *Heliand*'s creation, generally pointing to a date coinciding with attempts by neighboring tribes to Christianise the text's audience.

The *Heliand* is a re-working of the story of Jesus as presented traditionally. It is the only remaining full text in the now-lost language of an ancient European population, namely Old Saxon. It is not a translation, rather a re-composition of the original: 1) by translation, presumably from a Latin or Old High German (OHG) resource; 2) by change in style from prose to verse with grand elaborations; and 3) by reframing the story culturally through apparently deliberate omission of certain (likely culturally unacceptable) elements. Furthermore, there are some paleographic indications that *Heliand* was once sung. In other words, *Heliand* represents a text that contains multiple confounding alterations. Each of these alone challenges correct alignment with the original.

2.2 Tatian's *Diatessaron*

The *Diatessaron* is a harmony of the Gospels originally penned by the 2nd-century Syrian Tatian. A later translation into Latin (5th century) or its 9th-century translation into OHG is the proposed source behind the *Heliand*. Both Latin and OHG versions existed at the Benedictine scriptorium at Fulda in the early 9th century. Various theories (often circular in their logic) thus also place the *Heliand* there. Whatever the location of origin, the *Diatessaron* itself certainly represents the first steps of conversion from the original Gospels into OS, due to its similar re-organisation of the textual structure, wherein overlapping information from the four was reduced to a single text.

The stage of alteration from the Gospels to the *Diatessaron* is one that, though extensive, can quite easily be resolved using *n-gram* analysis, i.e. for determining which target material stems from which source material

The conversion of the *Diatessaron* to the *Heliand* is significantly more complicated, involving permutation, translation, and innovation. Thanks to his page-by-page linking of the *Heliand* with the OHG *Diatessaron*, Burkhard Taeger's (Behaghel/Taeger 1984) *Heliand* edition provides the reference notation that allows for smaller sections of both texts to be compared with one another relatively easily. This thus provides the bridge between the *Heliand* as target text and the Gospels as its source.

2.3 Otfrid's *Evangelienbuch*

Another old Germanic re-telling of the Gospels is the *Evangelienbuch*, penned in OHG by Otfrid of Weissenburg ca. 830. His presence at Fulda Monastery at that time is documented, which lends some credence to the theories noted above regarding the provenance of the OS *Heliand*.

Similar to the *Heliand*, Otfrid's work is a Gospel harmony retold in poetic verse. These facts have prompted theories that the *Heliand* and the *Evangelienbuch* were derived in succession from the *Diatessaron* at Fulda. However, the two texts differ; consequently, it is inconclusive as to whether the two are products of the same author.

These differences are evident primarily in style: the *Evangelienbuch* shows Latinate (or at least non-Continental Germanic) influence (end-rhyme, syllable counting), whereas the *Heliand* shows the traditional Germanic style (alliteration, loose metrics). Furthermore, the two linguistic varieties, OS and OHG, though genetically related, were already at the time distinct languages, also suggesting that the texts are not simply the product of the same author.

Whatever relationships do exist between the two texts are best seen when compared via the *Diatessaron*. The resulting network of texts – the OS *Heliand*, the OHG *Evangelienbuch*, and the Latin *Diatessaron*, presents a challenge in alignment (due to more than just the differences in language might suggest), because Paul Piper's 1884 reprint of Otfrid gives less-than-systematic chunks of the *Evangelienbuch* and their links with *Diatessaron*, meaning the full bridge between the interlinguistic texts is full of sizeable potholes. To close this gap, yet another version of the *Evangelienbuch*, this time in its Latin version, also needs to be taken into account.

2.4 Traditional Gospels

Eduard Sievers produced (1872) a side-by-side publication of the Latin and OHG *Diatessaron*. In it, he references where the Latin version corresponds to the Vulgate. With this, the complete bridge of relationships is built from the Gospels via the Latin and OHG *Diatessaron* and the OHG *Evangelienbuch* to the OS *Heliand*. It is significant to remember that, despite the linguistic varieties, only four texts are used.

Barring any significant differences resulting from the various languages of the texts, the resulting alignments via Sievers and Behaghel/Taeger create chunks of text that are manageable for comparison purposes: a handful of Gospel verses correspond to more or less individual *Diatessaron* sentences, as well as to ca. 30 lines of the *Heliand*. Section lengths from Otfrid vary widely, at most ca. 200 lines. Considering the full number of lines/verses of all texts involved reaches ca. 20,000 units, comparison groups averaging ca. 60 lines (from all texts) is an advancement on the necessity of comparing full texts with each other.

The Gospel translations relevant to the given project involve six languages (see "Digital Resources" below): Latin (Vulgate), Greek (*Novum Testamentum Graece*), Wulfila's Gothic (5th century), Old English (OE: Wessex Gospels, ca. 990), and two Early New High German versions from Luther (*Septembertestament* [1521] and *Letzter Hand* [1545]). The Gothic and OE versions are mostly complete and stand in place of non-existent OS and OHG full Gospels. As genetically related Germanic languages, Gothic and OE are therefore the closest linguistic comparison between the derived texts (i.e., the *Heliand*, Tatian, Otfrid) and the full Gospels. The Vulgate provides the bridge between all the texts, since the Latin *Diatessaron* seems to correspond nearly word-for-word to it. The Gothic Gospels are generally thought to have been made from the Greek, which is therefore included. Any difference between the full Gospels in Latin, Greek, and Gothic reveals relationships that might have been passed down to texts derived from them. Gothic and OE provide a similar role for comparing the language of the derived Germanic texts.

The Luther translations are included because of rumors that Luther may have been influenced by the *Heliand*. Thus, Luther's disputed translations might owe some of their existence to this purported external influence. The variations in Luther thus provide another wrinkle of difficulty to test our measures, viz. by comparing his language a) to the Latin and Greek, b) to the *Heliand* (and potentially Otfrid, Tatian, Gothic, and OE), and c) to itself. This last point explains the inclusion of two Luther translations – his first and his last, which will aid in determining when his linguistic innovations came into being.

3. Alignment

A key element to the theory that the *Heliand*, the *Evangelienbuch*, and the OHG *Diatessaron* are related is a purported preference in all three for St. Matthew. A known resource at Fulda is the *Matthäuskommentar*, created by Hrabanus Maurus, alleged translator of the *Diatessaron* into OHG. Thus, one further goal is to test this theory: do the *Heliand* and the *Evangelienbuch* truly follow Matthew more commonly? Simply stated, the arguments made to support both of the texts having been penned in Fulda after the OHG *Diatessaron* was completed are cyclical in nature. By either proving or disproving the supposed preference for Matthew in these two derived texts, one will give substantial proof either towards the argument of their link to Fulda, or at least that this argument is unfounded (as such, other theories about the provenance of the *Heliand* and the *Evangelienbuch* might look more promising).

In order to test which Gospel any given section of the *Heliand* and/or the *Evangelienbuch* follows, one needs a summary of where the four Gospels overlap in theme. A table consisting of such relationships was produced first in the late 3rd century by Eusebius of Caesarea. His *Eusebian Tables* account for all parallel sections where any of the four Gospels correspond with all or any of the other Gospels, as well as where each Gospel contains unique storyline. Eusebius made use of divisions created by contemporary 3rd-century Christian philosopher Ammonius of Alexandria, who developed the first means of subdividing the Gospels into verse-like chunks.[1]

For the use of our project, the ten Eusebian Canon Tables and their subordinate Ammonian sections have been borrowed wholesale with one minor addition: an eleventh Canon to account for the Long Ending of Matthew, which developed into Christian canon after Eusebius' time.[2] The "chunk" divisions made to the *Heliand*, the *Evangelienbuch*, and the *Diatessaron* have all been matched via their aforementioned relationships to the full Gospels in the Vulgate (i.e., through the Latin *Diatessaron* analysis by Sievers). To account for material where the *Diatessaron* suggests Matthew as the source, one needs to also consider where these verses overlap with similar thematic material in the other Gospels. This is done via the Eusebian Canon Tables. I have extended the search for overlap to include all four Gospels as possible sources. Thus, a matrix of relationships develops whereby the *Heliand* and the *Evangelienbuch* are compared to thematically similar sections of the *Diatessaron* in both OHG and Latin, which is then compared to the Vulgate Latin equivalent as postulated by Sievers. Since the Gospel chapter-and-verse divisions hold true regardless of language, the other five versions of the Gospels can be referenced easily. From here, a failsafe to prove or disprove Sievers' reading of the Latin *Diatessaron* is provided by taking the Ammonian section to which his suggested Gospel reference belongs, and comparing this Ammonian section with those from the other Gospels (and, apparently in some cases, within the same Gospel) as provided by the Eusebian Canon Tables.

All this produces thematically reduced chunks of each text, presented side-by-side with all the possible source material from the Gospels to which one needs to turn to find the actual source material. Note, this result only produces the correct aligned materials. A future stage of the project will seek to develop the computational algorithms to analyse the linguistic material in all its variant languages to determine relationships and/or similarity.

[1] The now-traditional chapter and verse divisions of the New Testament are Renaissance and Reformation-era inventions by Stephen Langton (chapters, 13th century; Fenlon 1910, "Hebrew Bible") and Robert Estienne (verses, 16th century; Miller/Huber 2004: 173).

[2] As this material was not considered canon during the 2nd century, it was not included by Eusebius in his tables. Nevertheless, as its inclusion into the Canon was accomplished anciently, this material in St. Matthew has had historical impact on other Gospel translations and derivatives. It is thus necessary to include it somehow in the system for research purposes. I have done so by simply assigning it as the eleventh Eusebian Canon Table, in continuation of his systematic.

Despite not yet providing this analysis of similarity, the present stage of the research does provide a necessary tool. Though quite simple in appearance, the relationships produced by this process have successfully linked disparate texts from different time periods and in distinctive styles and variant languages by discovering where these texts parallel one another thematically. The results are 4,560 inter-text groups made up of the related chunks between all the originals and languages involved. Each of these 4,560 groups easily fits on a single webpage (sometimes with minimal scrolling) that any researcher will be able to access via the Internet. Simple naked-eye comparisons of the sub-texts can be performed now, where in the past simply finding where one line from the *Heliand* relates to which verse of, e.g., St. Mark, would have required many hours of tedium.

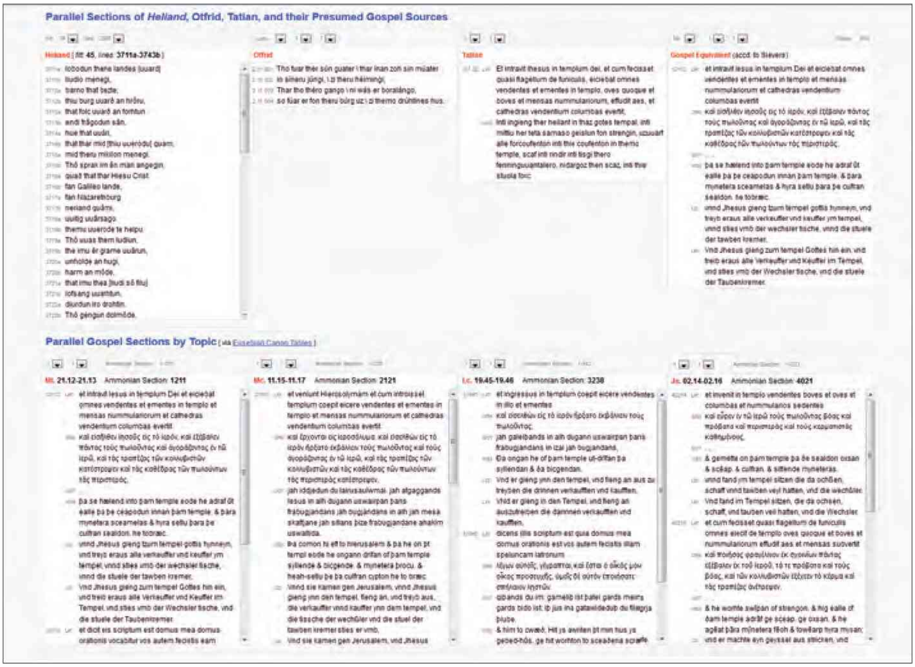
4. Representation

Access to the text groups is provided via a webpage,[3] shown here in Figure 1. This page is divided into two horizontal areas, each with four columns. In the top area are the three derived texts (the *Heliand*, Otfrid's *Evangelienbuch*, Tatian's *Diatessaron*) and their "Gospel Equivalent", i.e. the verse(s) from which the *Diatessaron* section is purportedly derived. The bottom half of the page delivers the contents of the Ammonian section to which the "Gospel Equivalent" belongs, as well as the text of any parallel Ammonian sections from the other Gospel books. This bottom grouping is ordered according to western tradition (Matthew, Mark, Luke, John). Wherever a text section from any source is too long for the page space provided, the column is scrollable:

Figure 1 reflects a search performed by inputting the group number 680 (of the 4,560 total inter-text groups) in a search field on a previous page. A similar search can be performed on structural divisions of any of the texts involved. Thus, searching for *Heliand* fitt[4] 34, line 2,885 will similarly bring up all the text subsections to which it has been related in the background database. Besides allowing for searches on book, chapter, and verse of the Gospels (lower half), one can also search according to Ammonian section numbers. The related parallel material of the other Gospels and the derived alignments will be called up automatically.

[3] http://titus.uni-frankfurt.de/database/diatessaron.

[4] The term historically used in reference to major subdivisions of the *Heliand*.



Parallel Sections of *Heliand*, Otfrid, Tatian, and their Presumed Gospel Sources

Heliand [fitt 45, lines: 3711a-3743b]

lobodun thane landes [uuard]
liudio menegi,
barno that bezte.
thiu burg uuarð an hrôru,
that folc uuarð an forhtun
endi frâgodun sân,
huue that uuâri,
that thar mid [thiu uuerodu] quam,
mid theru mikilon menegi.
Thô sprak im ên man angegin,
quað that thar Hiesu Crist
fan Galileo lande,
fan Nazarethburg
neriand quâmi,
uuitig uuârsago,
themu uuerode te helpu.
Thô uuas them liudiun,
the imu êr grame uuârun,
unholde an hugi,
harm an môde,
that imu thea [liudi sô filu]
lofsang uuarhtun,
diurdun iro drohtin.
Thô gengun dolmôde,

Otfrid

Tho fuar ther sún guater \ thar inan zoh sin múater
in álleru júngi, \ zi theru héimingi;
Thar tho thero gango \ ni wás er boralángo,
so fúar er fon theru búrg uz \ zi themo drúhtines hus.

Tatian

Et intrauit ihesus in templum dei. et cum fecisset quasi flagellum de funiculis, eiciebat omnes vendentes et ementes in templo. oves quoque et boves et mensas nummulariorum, effudit aes, et cathedras vendentium columbas evertit.
Inti ingieng ther heilant in thaz gotes tempal. inti mittiu her teta samaso geislun fon strengin, uzuuarf alle forcoufenton inti thie coufenton in themo temple, scaf inti rindir inti tisgi thero femninguuantalero, nidargoz then scaz, inti thie stuola forc

Gospel Equivalent (acc'd. to Sievers)

et intravit Iesus in templum Dei et eiciebat omnes vendentes et ementes in templo et mensas nummulariorum et cathedras vendentium columbas evertit
καὶ εἰσῆλθεν ἰησοῦς εἰς τὸ ἱερόν, καὶ ἐξέβαλεν πάντας τοὺς πωλοῦντας καὶ ἀγοράζοντας ἐν τῷ ἱερῷ, καὶ τὰς τραπέζας τῶν κολλυβιστῶν κατέστρεψεν καὶ τὰς καθέδρας τῶν πωλούντων τὰς περιστεράς,
...
þa se hælend into þam temple eode he adraf ût ealle þa þe ceapodun innan þam temple. & þara myneteta sceamelas & hyra setlu þara þe culfran sealdon. he tobræc.
vnnd Jhesus gieng tzum tempel gottis hynneyn, vnd treyb eraus alle verkeuffer vnd keuffer ym tempel, vnnd sties vmb der wechsler tische, vnnd die stuele der tawben kremer.
Vnd Jhesus gieng zum tempel Gottes hin ein, vnd treib eraus alle Verkeuffer vnd Keuffer im Tempel, vnd sties vmb der Wechsler tische, vnd die stuele der Taubenkremer.

Parallel Gospel Sections by Topic [via Eusebian Canon Tables]

Mt. 21.12-21.13 Ammonian Section: 1211

et intravit Iesus in templum Dei et eiciebat omnes vendentes et ementes in templo et mensas nummulariorum et cathedras vendentium columbas evertit
καὶ εἰσῆλθεν ἰησοῦς εἰς τὸ ἱερόν, καὶ ἐξέβαλεν πάντας τοὺς πωλοῦντας καὶ ἀγοράζοντας ἐν τῷ ἱερῷ, καὶ τὰς τραπέζας τῶν κολλυβιστῶν κατέστρεψεν καὶ τὰς καθέδρας τῶν πωλούντων τὰς περιστεράς,
...
þa se hælend into þam temple eode he adraf ût ealle þa þe ceapodun innan þam temple. & þara myneteta sceamelas & hyra setlu þara þe culfran sealdon. he tobræc.
vnnd Jhesus gieng tzum tempel gottis hynneyn, vnd treyb eraus alle verkeuffer vnd keuffer ym tempel, vnnd sties vmb der wechsler tische, vnnd die stuele der tawben kremer.
Vnd Jhesus gieng zum tempel Gottes hin ein, vnd treib eraus alle Verkeuffer vnd Keuffer im Tempel, vnd sties vmb der Wechsler tische, vnd die stuele der Taubenkremer.
et dicit eis scriptum est domus mea domus orationis vocabitur vos autem fecistis eam

Mc. 11.15-11.17 Ammonian Section: 2121

et veniunt Hierosolymam et cum introisset templum coepit eicere vendentes et ementes in templo et mensas nummulariorum et cathedras vendentium columbas evertit
καὶ ἔρχονται εἰς ἱεροσόλυμα. καὶ εἰσελθὼν εἰς τὸ ἱερὸν ἤρξατο ἐκβάλλειν τοὺς πωλοῦντας καὶ τοὺς ἀγοράζοντας ἐν τῷ ἱερῷ, καὶ τὰς τραπέζας τῶν κολλυβιστῶν καὶ τὰς καθέδρας τῶν πωλούντων τὰς περιστερὰς κατέστρεψεν,
jah iddjedun du Iairusaulwmai. jah atgaggands Iesus in alh dugann uswairpan þans frabugjandans jah bugjandans in alh jah mesa skattjane jah sitlans þize frabugjandane ahakim uswaltida.
Þa comon hi eft to hierusalem & þa he on þt templ eode he ongann drifan of þam temple syllende & bicgende. & myneteta bordu. & heah-setlu þe þa culfran cypton he to bræc.
Vnnd sie kamen gen Jerusalem, vnnd Jhesus gieng ynn den tempel, fieng an, vnd treyb aus, die verkauffer vnnd kauffer ynn dem tempel, vnd die tissche der wechßler vnd die stuel der tawben kremer sties er vmb,
Vnd sie kamen gen Jerusalem, vnd Jhesus

Lc. 19.45-19.46 Ammonian Section: 3238

et ingressus in templum coepit eicere vendentes in illo et ementes
καὶ εἰσελθὼν εἰς τὸ ἱερὸν ἤρξατο ἐκβάλλειν τοὺς πωλοῦντας,
jah galeiþands in alh dugann uswairpan þans frabugjandans in izai jah bugjandans,
Ða ongan he of þam temple ut-drifan þa syllendan & ða bicgendan.
Vnd er gieng ynn den tempel, vnd fieng an aus zu treyben die drinnen verkaufften vnd kaufften.
Vnd er gieng in den Tempel, und fieng an auszutreiben die darinnen verkaufften und kauffen.
dicens illis scriptum est quia domus mea domus orationis est vos autem fecistis illam speluncam latronum
λέγων αὐτοῖς, γέγραπται, καὶ ἔσται ὁ οἶκός μου οἶκος προσευχῆς. ὑμεῖς δὲ αὐτὸν ἐποιήσατε σπήλαιον λῃστῶν.
qiþands du im: gameliþ ist þatei gards meins gards bido ist; iþ jus ina gatawideduþ du filigrja þiube.
& him to cwæð. Hit ys awriten þt min hus ys gebed-hûs. ge hit worhton to sceaðena scræfe

Jo. 02.14-02.16 Ammonian Section: 4021

et invenit in templo vendentes boves et oves et columbas et nummularios sedentes
καὶ εὗρεν ἐν τῷ ἱερῷ τοὺς πωλοῦντας βόας καὶ πρόβατα καὶ περιστερὰς καὶ τοὺς κερματιστὰς καθημένους,
...
& gemette on þam temple þa ðe sealdon oxan & scêap. & culfran. & sittende myneteras.
vnnd fand ym tempel sitzen die da ochßen, schaff vnnd tawben veyl hatten, vnd die wechßler.
Vnd fand im Tempel sitzen, die da ochsen, schaf, und tauben veil hatten, und die Wechsler
et cum fecisset quasi flagellum de funiculis omnes eiecit de templo oves quoque et boves et nummulariorum effudit aes et mensas subvertit
καὶ ποιήσας φραγέλλιον ἐκ σχοινίων πάντας ἐξέβαλεν ἐκ τοῦ ἱεροῦ, τά τε πρόβατα καὶ τοὺς βόας, καὶ τῶν κολλυβιστῶν ἐξέχεεν τὸ κέρμα καὶ τὰς τραπέζας ἀνέτρεψεν,
...
& he worhte swipan of strengon. & hig ealle of ðam temple adrâf ge scêap. ge oxan. & he agêat þâra mýnetera fêoh & towêarp hyra mysan;
vnd er machte eyn geyssel aus stricken, vnd

Figure 1: Screenshot showing parallel texts belonging to group 680

The HTML code draws its information from a centralized database with a main table containing information about the 4,560 inter-text groups. These in turn reference the major subdivisions of each text by their various systems: the *Heliand* by Taeger's page numbering,[5] Otfrid by Piper's analysis, Tatian via Sievers' reprint, and the "Gospel Equivalents" by Sievers' alignment of the Latin Tatian version back to the Vulgate. The Gospel sections are referred to via their associated Ammonian Sections.

Beyond this, the database contains tables with information for the various divisional structures used for each text. Through a Structured Query Language (SQL) search of either these tables directly, or via the main inter-text table, a third layer of tables is queried from which the text is pulled. In the case of the *Heliand* and Otfrid, this third layer contains a single column containing the

[5] I.e., not by *fitts*, since the device used to compare between the *Heliand* and the *Diatessaron* provided by Taeger occurs at the top of each page of his *Heliand* edition. As any given *fitt* can vary in the number of pages it covers in his edition, it has been more practical to use Taeger's pages ad hoc as the unit of subdivision, though the text presented also indicates the given fitt and line number traditionally assigned to the *Heliand*.

text, while the other texts contain multiple columns separating the languages in which the text occurs.

5. Post-project directions

Due to the database structure chosen, each text can be accessed individually, and each language version of the multilingual texts can be accessed or hidden independently. Certain features were not implemented due to time constraints. These include: a set-up page in which the user can limit which of the Gospel languages he wishes to view. Also not realized was the ability to upload one's own Gospel translations and/or Gospel-derived texts, e.g. the Arabic *Injil*, which could theoretically still be made possible, as such translations/texts can be subdivided into comparably similar chunks as has been done for the *Heliand*, the *Evangelienbuch*, and the *Diatessaron*.

At the end of the research period, the results described in this paper were handed over for further development to Prof. Christian Chiarcos, also of the Frankfurt "Digital Humanities" group. Prof. Chiarcos has continued to build on these results by developing the means to automatically compare the linguistic material of the inter-text groups, in part by taking advantage of comparative semantic information delivered by another sub-project, namely 'Historische Sprachdatenbank ‹Simplex›', for which I was also responsible. For this sub-project, I collated digitized etymological and translation dictionaries of the languages dealt with above.

Using the resulting network of semantic and formal relationships, Prof. Chiarcos has succeeded in being able to automatically highlight the semantic relationships between the texts under concern here, as delivered by the inter-text subdivision. Thus, for example, by clicking on "intravit" of the Latin Tatian section shown in Figure 1, related words ("ingieng" in OHG Tatian, "quam" in *Heliand* line 3734a, "geng" in Otfrid, "eode" in OE [WSX],[6] etc.) appears highlighted, directing the user's eye to more specifically, semantically related areas between the texts than what the current divisions provide.

Also under development by Prof. Chiarcos is the ability to compare multi-word units and *n-grams*, where such exist. A goal of this step is to reveal similar syntactic strings with semantic relations, e.g. "intravit Ihesus in templum dei" (Latin Tatian), "ingieng ther heilant in thaz gotes tempal" (OHG Tatian

[6] I.e., Old English (Wessex).

117: 2), "Thô he an thene uuîh innen, / geng an that godes hûs" (*Heliand* 3733b-3734a,), "fúar er [...] \ zi themo drúhtines hus" (Otfrid Ev. 2 11:4), "se hælend into þam temple eode" (OE [WSX] Matthew 21:12), as well as the similar text from the other Gospels, e.g. "vnnd Jhesus gieng ynn den tempel" (L22 Mark 11:15), "ah galeiþands in alh" (GOT Luke 19:45), "et invenit in templo" (LAT John 2:14), even where this differs from the "direct" translation by Luther (L45): "Vnd fand im Tempel sitzen".

Digital resources[7]

Aland, Barbara/Aland, Kurt (2001): Novum Testamentum Graece. 27. Aufl. Stuttgart: Dt. Bibelges. Center for Computer Analysis of Texts (CCAT) database corrected and amplified by James Tauber. Accessed using E-Sword: www.e-sword.net/.

Behaghel, Otto/Taeger, Burkhard (1984): Heliand und Genesis. Tübingen: Max Niemeyer. Accessed at TITUS: http://titus.uni-frankfurt.de/texte/etcs/germ/asachs/heliand/helia.htm.

Luther, Martin (1522): Biblia: die gantze Heilige Schrifft. Deudsch. Wittenberg. Accessed at Wikisource: http://de.wikisource.org/wiki/Lutherbibel.

Luther, Martin (1546): Biblia: Die gantze Heilige Schrifft Deudsch. Wittenberg: Hans Lufft. Accessed at Wikisource: http://de.wikisource.org/wiki/Lutherbibel.

Piper, Paul (1884): Otfrids Evangelienbuch: mit Einleitung, erklärenden Anmerkungen und ausführlichem Glossar / 2 Glossar und Abriss der Grammatik. Paderborn: Schöningh. Accessed at TITUS: http://titus.uni-frankfurt.de/texte/etcs/germ/ahd/otfrid/otfri.htm.

Sievers, Eduard (1872): Tatian. Lateinisch und altdeutsch / mit ausführlichem Glossar. Paderborn: Schöningh. Accessed at TITUS: http://titus.uni-frankfurt.de/texte/etcs/germ/ahd/tatian/tatia.htm.

St. Jerome (405): Latin Vulgate with Deuterocanon using Gallican Psalter. Accessed using E-Sword: www.e-sword.net/.

Streiberg, Wilhelm (1919): Die Gotische Bibel: Der gotische Text und seine griechische Vorlage. Mit Einleitung, Lesarten und Quellennachweisen sowie den kleineren Denkmälern als Anhang. Heidelberg: Carl Winter. Accessed at the Wulfila Project: www.wulfila.be.

Wessex Gospels (ca. 990): Part of the York-Toronto-Helsinki Parsed Corpus of Old English prose. Accessed at the Oxford Text Archive: http://ota.ahds.ac.uk.

[7] All URLs have been checked and found valid as of late January 2015.

References

Fenlon, John Francis (1910): The Catholic Encyclopedia. Vol. 7. New York: Robert Appleton Company.

Miller, Stephen M./Huber, Robert V. (2004): The Bible: A History: The Making and Impact of the Bible. Intercourse, PA: Good Books.