Version Variation Visualization (VVV): Case Studies on the Hebrew Haggadah in English
Tom Cheesman, Avraham Roos

To cite this version:
Tom Cheesman, Avraham Roos. Version Variation Visualization (VVV): Case Studies on the Hebrew Haggadah in English. 2017. <hal-01307217v2>

HAL Id: hal-01307217
https://hal.archives-ouvertes.fr/hal-01307217v2
Submitted on 3 Jun 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Version Variation Visualization (VVV): Case Studies on the Hebrew Haggadah in English

Tom Cheesman (Swansea University) and Avraham Roos (University of Amsterdam)

The ‘Version Variation Visualization’ project has developed online tools to support comparative, algorithm-assisted investigations of a corpus of multiple versions of a text, e.g. variants, translations, adaptations (Cheesman, 2015, 2016; Cheesman et al., 2012, 2012-13, 2016; Thiel, 2014; links: www.tinyurl.com/vvvex). A segmenting and aligning tool allows users to 1) define arbitrary segment types, 2) define arbitrary text chunks as segments, and 3) align segments between a ‘base text’ (a version of the ‘original’ or translated text), and versions of it. The alignment tool can automatically align recurrent defined segment types in sequence. Several visual interfaces in the prototype installation enable exploratory access to parallel versions, to comparative visual representations of versions’ alignment with the base text, and to the base text visually annotated by an algorithmic analysis of variation among versions of segments. Data can be filtered, viewed and exported in diverse ways. Many more modes of access and analysis can be envisaged. The tool is language neutral. Experiments so far mostly use modern texts: German Shakespeare translations. Roos is working on a collection of approx. 100 distinct English-language translations of a Hebrew text with ancient Hebrew and Aramaic passages: the Haggadah (Roos, 2015)

1. The Haggadah

On the evening before Passover (Pesach), Jews gather at home to celebrate a festive ceremony and meal with family and friends, to commemorate the biblical Exodus of the Jewish people out of Egypt. They eat the traditional matza, drink the prescribed four glasses of wine, and read from the Haggadah. This is a Hebrew text with instructions for a 15-phase ceremony: what to say or sing, which acts to perform, in what order, when to eat or drink what, etc. All participants hold a printed copy of the Haggadah. Typically, many different versions (plural: Haggadot) are present in the room.

The Hebrew Haggadah text is a compilation of Bible quotes, excerpts from traditional rabbinal teachings (Mishnah, Midrash), Exodus narrative, explanations of the festival’s history, and Passover ‘laws’. The text probably dates back to 200-300 CE. The oldest complete manuscript dates to the 10th century CE. Thousands of variant Hebrew-language versions are extant, in manuscript and print. There are translations in over 40 different languages (Yudlov, 1997). The first English-language version appeared in London in 1770. Countless more have since appeared. Yudlov (1997) catalogues 823 English-language editions to 1960. The rate of production of new ones has since been accelerating exponentially. Most of these are retranslations, variously dependent on precursors.

Roos is compiling a digital corpus of English-language Haggadah translations, and using digital tools to compare them and visualize the differences. He aims to explain the differences in terms of their cultural historical contexts, and so shed light on translators’ minds and motives.
2. Version Variation Visualization (VVV): Eddy and Viv Algorithms

VVV compares multiple retranslation documents at segment level, and visualizes the similarities and differences, in order to facilitate overviews, close reading, and navigation among versions. An algorithm called ‘Eddy’ (‘$\Sigma D$’) quantifies variation among versions of a base text segment, in order to distinguish more and less predictable or distinctive versions. An algorithm called Viv (‘variation in variation’) aggregates Eddy metrics, and projects the result onto the base text segment, in order to distinguish more and less variously translated segments. The algorithms can be applied to the aligned corpus or any selected sub-corpus.

2.1 Eddy

The Eddy value assigned to a particular version or section indicates its "strangeness" as compared to other versions. Essentially, Eddy assigns lower metrics to wordings which are closer to the notional average, and higher metrics to more distant ones. So, Eddy ranks versions on a cline from low to high distinctiveness, or originality, or unpredictability. It sorts common-or-garden translations from interestingly different ones.

Eddy can be implemented in various ways. Our standard approach is:

Each word in the corpus word list [where corpus means the corpus of aligned segment versions] is considered as representing an axis in N-dimensional space, where N is the length of the corpus word list. For each version, a point is plotted within this space whose co-ordinates are given by the word frequencies in the version word list for that version. (Words not used in that version have a frequency of zero.) The position of a notional ‘average’ translation is established by finding the centroid of that set of points. An initial ‘Eddy’ variation value for each version is calculated by measuring the Euclidean distance between the point for that version and the centroid. (Flanagan in: Cheesman, Flanagan, and Thiel, 2012-13)

No stop words are excluded; no stemming, lemmatisation or parsing is performed. Users can also select a more primitive arithmetical formula, and one using Dice’s co-efficient.

In the VVV ‘Eddy and Viv’ view, when a base text segment is selected, the segment-versions are displayed in a scrollable list in Eddy order, with associated metrics, and with a visual representation of relative Eddy value. The list can be re-ordered to display by date, translator name, or segment length in characters. Eddy values can also be displayed, explored, and exported in the form of charts and tables.

Examples of Eddy use will be provided in section 4.1.

2.2. Viv

Viv aggregates the Eddy values for a segment. In our standard approach, Viv is the average of Eddy values of version-segments. Users can also select Viv as the standard deviation of Eddy values. Viv indicates where translators differed most or least, in relation to the base
text. (This function is comparable with the amount of layering associated with a word or string of words in the TRAViz visualization: Jänicke et al., 2015.)

In the VVV ‘Eddy and Viv’ view, Viv is represented on the base text by a tonal underlay, varying with the relative value of each segment. Metrics can be viewed by brushing a segment. Floor and ceiling values can be altered to facilitate surveying the base text.

Segments can be filtered in various ways (text search, Eddy/Viv ranges, segment lengths, etc), in the base text and in the version corpus or subcorpora, and texts exported in CSV tables with associated Eddy and Viv metrics.

Examples of Viv use will be provided in section 4.3.

As one reviewer commented, Eddy/Viv is not the only possible approach to comparing differing translations/versions. Measuring the overlap of words (or lemmas) among segments would achieve the same effect. Such a method would also need to calculate a centroid and distances from it.

VVV is specifically created to compare numerous retranslations of the same source text, making it ideal for research into Haggadah version variation. It can help a researcher identify variations, and present them to an audience.

3. Related Work

There has been some digital work on larger retranslations corpora, involving works of wide intrinsic interest, but none designed to facilitate access to multiple translations, and the translated work, together with algorithmic analyses. Janicke et al. (2015) take an in some ways similar approach, but their ‘TRAViz’ interface offers a very different mode of text visualization, is monolingual (shows no translated text), and works best with more limited variation and shorter texts. Similarly, Juxta-Commons, CollateX and Stemmaweb are monolingual, do not rate the "strangeness" of variants in comparison to all others (Eddy value) and do not create a heat map in the source text (Viv Value) revealing which ST instances generated most target text variants, two of the most powerful VVV features. It is especially these comparisons between the ST and the TT that this research focuses on. Whereas the above software highlights differences in the versions, VVV highlights how these differences are connected to the ST, thus helping us the explain the reasons for the variants.

Lapshinova-Koltunski (2013) describes a parallel multi-translation corpus designed to support computational linguistic analyses of differences between professional translations, student translations, MT outputs and edited MT outputs. Shei and Pain (2002) proposed a similar parallel corpus, with an interface designed for translator training. These projects only offer access to filtered segments of the text corpus, and do not envisage exploring variation among retranslations. Altintas, Can, and Patton (2007) used two time-separated (c.1950, c.2000) collections of published translations of the same seven English, French or Russian literary classics into Turkish, in order to quantify aspects of language change. This raises the question whether such translations ‘represent’ their language. Corpus-based Translation Studies (Baker, 1993; Kruger et al., 2011) has established that translated language differs from untranslated language. We also know from decades of work in Descriptive Translation Studies (Morini, 2014; Toury, 2012) that retranslations vary for complex genre-, market-, subculture-specific and institutional factors, and individual psychosocial factors, involving
the translators and others with a hand in the work (commissioners, editors), and their uses of resources including source versions and prior (re)translations.

4. Using VVV with Haggadah Samples

One section of the Haggadah concerns four sons who represent four different attitudes to Judaism. Each asks a question which characterizes their attitude, and the text suggests how to respond to these questions. The Hebrew text has 126 words and is divided into six parts: (1) introduction; (2) characterization of the four sons; (3)-(6) one paragraph for each son, with his question and the response.

60 different translations were uploaded to VVV, segmented and manually aligned with the Hebrew base text. Each translation contains between five and twelve manually defined segments: units of meaning.

4.1 Exploring with Eddy

In part (2), characterizing the four sons, most translators use straightforward terms: ‘wise’, ‘wicked’, ‘simple’, and ‘one who does not know how to ask’. Some are more creative. Eddy highlights certain translations as ‘strange’. VVV automatically rates each version, thus ranking all 60 versions from most common to most exotic. In Table 1, the left column gives the original Hebrew and the commonest translation (lowest Eddy value). The second column gives the five translations with highest Eddy values (rounded Eddy figures given in column 3): outliers in comparison to all other versions in the corpus, worth further exploration.

<table>
<thead>
<tr>
<th>Base text and lowest Eddy translation</th>
<th>5 highest Eddy translations</th>
<th>Eddy value</th>
<th>Version reference</th>
<th>Version date</th>
</tr>
</thead>
<tbody>
<tr>
<td>One wise</td>
<td>One is intelligent</td>
<td>1.28</td>
<td>REGFORST2</td>
<td>1952</td>
</tr>
<tr>
<td></td>
<td>One is understanding</td>
<td>1.30</td>
<td>POLYHH</td>
<td>1974</td>
</tr>
<tr>
<td></td>
<td>The intelligent child</td>
<td>1.38</td>
<td>WILROS</td>
<td>1906</td>
</tr>
<tr>
<td></td>
<td>A clever son</td>
<td>1.45</td>
<td>TCH</td>
<td>1954</td>
</tr>
<tr>
<td></td>
<td>The first is sensible</td>
<td>1.52</td>
<td>MSAM</td>
<td>1942</td>
</tr>
<tr>
<td>One wicked</td>
<td>The second mean</td>
<td>1.38</td>
<td>MSAM</td>
<td>1942</td>
</tr>
<tr>
<td></td>
<td>The rebellious child</td>
<td>1.38</td>
<td>ANIM</td>
<td>2005</td>
</tr>
<tr>
<td></td>
<td>One is ill-mannered</td>
<td>1.50</td>
<td>REGFORST2</td>
<td>1952</td>
</tr>
<tr>
<td></td>
<td>and one who is stubborn</td>
<td>1.63</td>
<td>GUT</td>
<td>1956</td>
</tr>
</tbody>
</table>
A translation’s relative Eddy varies, as we will see in the next section.

<table>
<thead>
<tr>
<th>son1</th>
<th>REGFORST2</th>
<th>POLYHH</th>
<th>TCH</th>
<th>WILROS</th>
<th>MSAM</th>
<th>...</th>
</tr>
</thead>
</table>
TABLE 2. Highest Eddy Scorers from Table 1

4.2 Eddy Variation Chart

The poet Abraham Regelson published several *Haggadot*. Roos’s collection includes one from 1944 (REGFORST1) and another from 1952 (REGFORST2). VVV’s Variation Chart (‘Eddy Overview’) helps us compare these two translations (see Figure 1). This chart plots each version’s Eddy values on the y-axis, for segments in sequence on the x-axis. The user can select which versions’ graphs to display or hide, select an area to zoom in on, and hover over a node to display base text and version (as is shown in Figure 1). In Figure 1 we see Regelson using higher Eddy-value language (more distinctive language in relation to the corpus) in 1952 than in 1944. One exception is highlighted, in part (2) of the passage (discussed in Section 3.3).

![Figure 1. Comparing REGFORST1 and REGFORST2 in the Eddy Variation Chart](image_url)

Let us compare (close read) these two Regelson versions:

<table>
<thead>
<tr>
<th>Table 2</th>
<th>REGFORST1 Regelson 1944</th>
<th>REGFORST2 Regelson 1952</th>
</tr>
</thead>
<tbody>
<tr>
<td>son2</td>
<td>REGFORST2 POLYHH MSAM ANIM GUT</td>
<td>REGFORST2 TCH NAH LEHM HOS</td>
</tr>
<tr>
<td>son3</td>
<td>REGFORST2 TCH</td>
<td>REGFORST2 POLYHH TCH WILROS GLATZ NSEX</td>
</tr>
</tbody>
</table>

One does not break up the Passover ceremony by announcing: "To the afternoon entertainment!"
Blessed is the All-Including, blessed is He who had given a Torah to his people. The Torah has spoken concerning four sons –

one wise, and one wicked, and one simple, and one who wits not to ask.

What says the Wise Son? - What are the testimonies and laws and behaviors, which the Lord, our God, has commanded you? Do thou, then, tell him precisely the practices of the Passover: One does not break up the Passover ceremony by announcing: "To the aftermeal entertainment!"

What says the Wicked Son - "Of what use is this service to you?" To you, and not to himself! By excluding himself from the Community, he has denied the Deity. Do thou, then, set his teeth on edge! Say to him: "This is on account of what the Lord did for me when I went forth from Egypt." For me, and not for him - had he been there, he would not have been redeemed.

What says the Simple Son? - "What is all this about?" Therefore, say to him: "With might of hand, the Lord hath taken us out of Egypt, from the house of slavery."

But the One Who Wits Not To Ask-sit is for thee to open talk with him, as it is said: "And thou shalt tell thy son on that day, saying: 'This is on account of what the Lord did for me when I went forth from Egypt.'"

Praised be God, praised be He. Blessed be He who gave the Torah to His people, Israel. Blessed be He. On the subject of the Passover service the Torah speaks of – FOUR SONS

One is intelligent, one is ill-mannered, one is indifferent, and one is not even able to ask a question.

1. The INTELLIGENT son asks: “What is the meaning of all the Passover customs and ceremonies, the rules and rites which God has commanded?” You will explain to him all the traditions of Pesach down to the last detail of the Afikoman.

2. The ILL-MANNERED one asks: “What’s the sense of all this business of yours?” Yours, he says; and none of his. By refusing to identify himself with his people he denies a basic principle of our religion. You may fling this in his teeth: “I do this because of what the Lord did for me when He rescued me from Egypt.” Me not him. Let him know that had he been there, by denying his brothers he could not have been saved.

3. The INDIFFERENT one merely asks: “What is this?” Tell him: “With a strong hand God took us out of Egypt where we were slaves.”

4. The INCOMPETENT one-get him started by quoting the words from the Bible: “In that day you shall tell your son saying: (Point to the ceremonial dishes.) “All of this is because of what the Lord did for me when I came out of Egypt.”
It makes sense to assume that, having in 1944 already translated the Haggadah in a quite straightforward manner, Regelson decided for the 1952 translation to try out different translation techniques, more off the beaten path. That would explain why his 1952 translation scores higher Eddy values. Examples of this can be seen by the alliterated names of the four sons (Intelligent, ill-mannered, indifferent, and incompetent), something not found in any of the other versions. On the other hand, having used a etymology-based final comment for the wise son in 1944, Regelson opts for a very specific Jewish-sources based closing for his 1952 intelligent son. We will comment on this further in the next section.

4.3 *Viv: Variation in Variation*

In VVV’s ‘Eddy and Viv’ interface, base text segments are highlighted according to their Viv value: the higher the value, the darker the underlay tone. We can visually identify which base text segments produced the most variant translations, or ‘read the original by the light of the translations’ (Cheesman, 2015).

Figure 2 depicts a selection from an ‘Eddy and Viv’ view of the ‘Four Sons’ passage. The six parts are shown as paragraphs. Viv underlay tones indicate that the segments with the highest Viv value are within part (3), the ‘wise’ son, and part (4), the ‘wicked’ son. Most of part (4) has very low Viv value. Part (2), giving the sons’ names, also has very low Viv value.

![Figure 2. Partial Screenshot of Eddy and Viv view of Haggadah 'Four Sons' Passage](image)

This focuses our attention on the highest Viv segment: the answer to the wise son’s question. Here the Hebrew text of part (3) is followed by a recent, scholarly English version, including commentary:

ךכם המ אמרו? מה הכותדים והקשרים ['-ך' ה' אלוהים ]

אתם. אם אתם אמרו 'ךכם המ קשרים'? א' מipients אמו קשרים.

וככן. אם אתה אמרו 'ךכם המ קשרים'? א' מipients אמו קשרים.
What does the wise [son] say? ‘What are these testimonies, statutes and judgments that the Lord our God commanded you?’ (Deuteronomy 6:20) And accordingly you will say to him, as per the laws of the Pesach sacrifice, ‘We may not eat an afikoman [a dessert or other foods eaten after the meal] after [we are finished eating] the Pesach sacrifice.’ (Sefaria website, www.sefaria.org, 2014)

The segment with highest Viv value is the one beginning: ‘We may not eat an afikoman…’. This is a quote from the Mishna (oral laws compiled about 200 CE by Rabbi Judah HaNasi, the basis for the later Talmud). Already in the Talmud (c. 500 CE) the correct meaning of the term afikoman had become obscure and was disputed. In Talmudic traditions, afikoman (Hebrew אפיקומן) is said to derive from Greek epikomen or epikomion (ἐπί Κομός), ‘that which comes after’, variously interpreted as (A) ‘dessert’, or (B) ‘after-dinner entertainment/revelry’, and additionally as (C) a metaphor.

There are five common interpretations in the context of the wise son: (A1) any proscribed dessert; (A2) the specific prescribed dessert at stage 13 of Passover (the piece of a matza which is broken in two during stage 4); (B1) proscribed excessive subsequent entertainment (distinguishing Passover from pagan celebrations); (C1) prescribed teaching of all of the (Passover) law: because the afikoman is the last law in the section on Pesach, so “We don't leave anything until after the afikoman” could mean, "we don't stop studying until we have learned everything"; (C2) prescribed sacrifice of a Passover lamb. There is also a sixth option for translation: leaving afikoman to stand in the target text, uninterpreted.

This range of options explains the segment’s high Viv value. Some of the variant English versions, low in the Eddy value list, are shown in the VVV ‘Eddy and Viv’ view in Figure 2.

Figure 2. Eddy and Viv view of Haggadah Four Sons passage
Recall Figure 1. In this segment, Regelson first (in 1944) opted for ‘afternoon entertainment’ (interpretation B). That has quite high Eddy value: it’s a less popular translation. In 1952 he switched to the version seen in Figure 2: ‘down to the last detail of the afikoman’ (interpretation C1). This has a much lower Eddy value than the 1944 option: in this instance, Regelson’s later Haggadah made a commoner translation choice. It might be that this decision was influenced by the fact that more and more translators by that time interpreted afikoman in a metaphoric manner.

Viv and Eddy values are calculated according to manually set segments (meaningful units). These can be single words, phrases, sentences, paragraphs or even the whole text. By creating a whole text segment, one could easily aggregate data for a quantitative comparison of whole texts. This way, although no one version consistently deviates from the norm for all 4 sons (as shown above), it is possible to ascertain which translation as a whole is furthest from the base texts. It should be noted, however, that by doing this, the Viv value becomes irrelevant. The version that scored the highest overall Eddy value was the 1974 Polychrome Haggadah by Jacob Freedman, whose translation is extremely verbose and elaborate.

The Viv

4.4 Parallel View Visualization: Alignment Maps

Parallel view visualizations include a distant overview of segment alignments between base text and versions: an ‘alignment map’. Successive segments of the base text are represented as a vertical ‘barcode’: the thickness of a bar represents segment length in words. Segments of a version are represented in the same format. Alignments are represented by lines connecting base text and version. This enables rapid identification of translators’ editing decisions: omission, addition, reduction, expansion, and transposition.

Figure 3 (created from screenshots) shows ten examples of the ‘Four Sons’ passage. The unchanging base text is on the left, the version on the right of each ‘map’. The afikoman segment is highlighted.
Figure 3. Alignment Maps of the Four Sons Passage

Evidently most translations are much longer than the original. Hebrew is a very concise language; many translations expand, expound and explain. We can see that the very first English version (1770) is almost word for word, but omits the *afikoman* segment. So do the 1906 and 1993 versions. The 1983 and 1993 versions (both associated with Reform Judaism) cut other parts of the text. The 1974 version (POLYHH, an outlier in Table 1) expands to an extraordinary degree. The 1955 and 1967 versions are also expansive. They did not appear in Table 1. Figure 3 now explains why: both omit the segments which name the four sons, shown in Table 1.

5. Conclusions

Using VVV can yield valuable insights when comparing multiple variants, and is also useful for presenting findings visually. Manually comparing different versions becomes difficult with larger corpora. When Viv is highlighted in base text segments, even researchers with no knowledge of a language (in this case Hebrew) can identify the parts that warrant closer inspection.

VVV offers a useful range of visualization modes, but many more can be developed. Future research planned on the *Haggadah* includes comparing the language use of translators when translating Hebrew and Aramaic text passages, comparing the translations of biblical Hebrew versus more modern Hebrew text passages, and translated transitive versus intransitive verb forms.

6. Thanks

We would like to thank VVV developer Kevin Flanagan for continuing technical support, and Eran Hadas for his endeavors to streamline uploading and segmenting texts in VVV.

7. Funding

The VVV project was supported in 2011-13 by Swansea University (Research Incentive Fund and Bridging the Gaps). The main phase of software development was funded by a 6-month Research Development Grant under the Digital Transformations theme of the Arts and Humanities Research Council (UK), reference AH/J012483/1, in 2012.

8. References


