



**HAL**  
open science

## Version Variation Visualization (VVV): Case Studies on the Hebrew Haggadah in English

Tom Dr. Cheesman, Avraham A. I. R. Roos

### ► To cite this version:

Tom Dr. Cheesman, Avraham A. I. R. Roos. Version Variation Visualization (VVV): Case Studies on the Hebrew Haggadah in English. 2016. hal-01307217v1

**HAL Id: hal-01307217**

**<https://hal.science/hal-01307217v1>**

Preprint submitted on 26 Apr 2016 (v1), last revised 3 Jun 2017 (v2)

**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/vvvox](http://www.tinyurl.com/vvvox)). 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 rabbinical 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' ( $\sum D$ ) 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

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.

### 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.

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. 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.

#### 3.1 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'. 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, worth further exploration.

| Base text and lowest Eddy translation | 5 highest Eddy translations                             | Eddy value | Version reference | Version date |
|---------------------------------------|---|------------|-------------------|--------------|
| אחד חכם<br>One wise                   | One is intelligent                                      | 1.28       | REGFORST2         | 1952         |
|                                       | One is understanding                                    | 1.30       | POLYHH            | 1974         |
|                                       | The intelligent child                                   | 1.38       | WILROS            | 1906         |
|                                       | A clever son  | 1.45       | TCH               | 1954         |
|                                       | The first is sensible                                   | 1.52       | MSAM              | 1942         |
| ואחד רשע<br>One wicked                | The second mean   | 1.38       | MSAM              | 1942         |
|                                       | The rebellious child                                    | 1.38       | ANIM              | 2005         |
|                                       | One is ill-mannered                                     | 1.50       | REGFORST2         | 1952         |
|                                       | and one who is stubborn                                 | 1.63       | GUT               | 1956         |
|                                       | one is recalcitrant and scornfully insolent             | 1.81       | POLYHH            | 1974         |
| ואחד תם<br>One simple                 | One Artless   | 1.17       | NAH               | 2012         |
|                                       | One is indifferent                                      | 1.32       | REGFORST2         | 1952         |
|                                       | A simpleton   | 1.32       | LEHM              | 1972         |
|                                       | The naïve son   | 1.33       | HOS               | 2009         |
|                                       | A dull son  | 1.45       | TCH               | 1954         |
| ואחד שאינו יודע לשאול                 | and a fourth, a child that does not yet know how to ask | 1.90       | GLATZ             | 1989         |
|                                       | and the fourth incapable of even asking a question      | 1.91       | POLYHH            | 1974         |

|                                      |  |      |        |      |
|--------------------------------------|--|------|--------|------|
| and one who does not know how to ask | and one who is too young to ask any questions about the things he sees   | 1.92 | TCH    | 1954 |
|                                      | and the child who does not know enough to make inquiry of his own accord | 1.94 | WILROS | 1906 |
|                                      | and the child still too young to even inquire of the Pesakh              | 2.13 | NSEX   | 1983 |

**TABLE 1. Names of the Four Sons: Translations with Lowest and Highest Eddy**

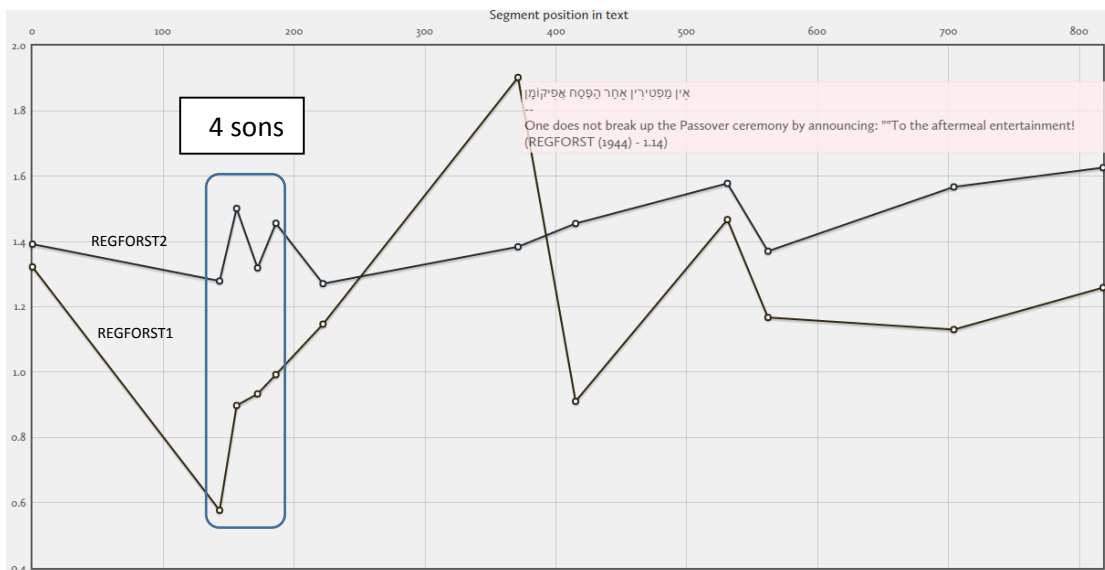
The corpus includes C18 and C19 versions, but none appear in this table. Almost all high Eddy versions date from the 1940s and after. The general retranslation trend is towards greater variation. The 1906 translation (WILROS) is an early outlier, worth further investigation. It is also intriguing that no version is consistently in the highest 5 for all four sons (see Table 2). A translation's relative Eddy varies, as we will see in the next section.

|      |           |        |     |        |      |      |     |     |      |     |       |      |
|------|-----------|--------|-----|--------|------|------|-----|-----|------|-----|-------|------|
| son1 | REGFORST2 | POLYHH | TCH | WILROS | MSAM |      |     |     |      |     |       |      |
| son2 | REGFORST2 | POLYHH |     |        | MSAM | ANIM | GUT |     |      |     |       |      |
| son3 | REGFORST2 |        | TCH |        |      |      |     | NAH | LEHM | HOS |       |      |
| son4 |           | POLYHH | TCH | WILROS |      |      |     |     |      |     | GLATZ | NSEX |

**TABLE 2. Highest Eddy Scorers from Table 1**

### 3.2 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 brush a node to display base text and version. 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**

### 3.3 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**

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:

חכם מה הוא אומר? מה העדות והחקים והמשפטים אשר צוה ה' אלהינו אתכם. ואף אתה אמור לו כהלכות הפסח: אין מפטירין אחר הפסח אפיקומן.

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](http://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* (επί 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 proscribed 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) proscribed teaching of *all* of the (Passover) law; (C2) proscribed 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.

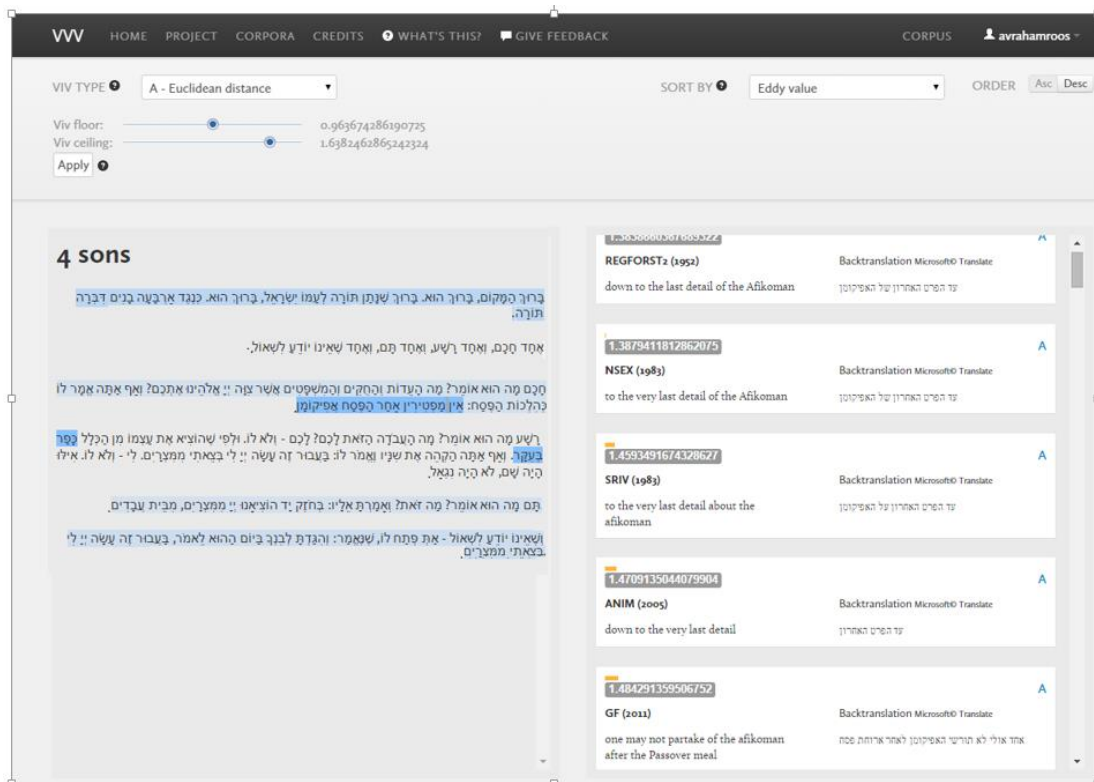


Figure 2. Eddy and Viv view of *Haggadah* Four Sons passage

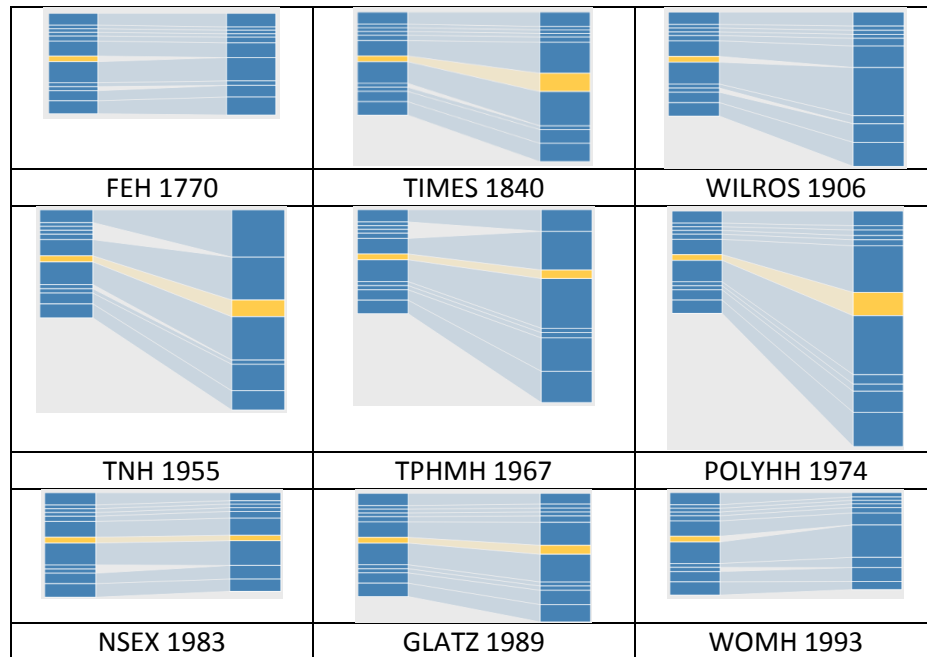
Recall Figure 1. In this segment, Regelson first (in 1944) opted for 'afternoon entertainment' (interpretation B). That has quite high Eddy value: it's a less popular translation. In 1952 he switched to the version seen in Figure 2: 'down to the last detail of the afikoman' (interpretation C1). This has a much lower Eddy value than the 1944 option: in this instance, Regelson's later *Haggadah* made a commoner translation choice.

### 3.3 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.

#### 4. 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.

## 5. 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.

## 6. 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.

## 7. References

- Cheesman, T. (2015). Reading Originals by the Light of Translations. *Shakespeare Survey*, 68: 87-98.
- Cheesman, T. and the VVV Project Team (2012). Translation Sorting with Eddy and Viv. <http://www.scribd.com/doc/101114673/Eddy-and-Viv> (accessed 16 January 2016).
- Cheesman, T., Flanagan, K., and Thiel, S. (2012-13). Translation Array Prototype 1: Project Overview. <http://www.delightedbeauty.org/vvvclosed/Home/Project> (accessed 16 January 2016).
- Cheesman, T., Flanagan, K., Thiel, S., and Rybicki, J. (2016), Six Maps of Translations of Shakespeare. In: Wiggins, B. and Macleod, C. (eds), *Un/Translatables: New Maps for Germanic Literatures*. Evanston, Illinois: Northwestern University Press, forthcoming.
- Geng, Z., Laramée, R.S., Cheesman, T., Ehrmann, E., and Berry, D.M. (2011). Visualizing Translation Variation: Shakespeare's *Othello*. *Advances in Visual Computing. Lecture Notes in Computer Science*, 6938: 653-663.
- Geng, Z., Laramée, R.S., Flanagan, K., Thiel, S., and Cheesman, T. (2013). ShakerVis: Visual Analysis of Segment Variation of German Translations of Shakespeare's *Othello*. *Information Visualization*, 12(3-4): 273-288.
- Jänicke, S., Geßner, A., Franzini, G., Terras, M., Mahony, S. and Scheuermann, G. (2015). TRAViz: A Visualization for Variant Graphs. *Digital Scholarship in the Humanities*, 30 (suppl 1): i83-i99. [http://dsh.oxfordjournals.org/content/30/suppl\\_1/i83](http://dsh.oxfordjournals.org/content/30/suppl_1/i83) (accessed 16 January 2016).
- Roos, A. (2015). Making a Clean Breast of English Passover Haggadah Translations: Data Visualization of Bowdlerization in Haggadah Translations of Ezekiel 16:7. Unpublished chapter of PhD dissertation, University of Amsterdam.

Sefaria: a Living Library of Jewish Texts Online. [www.sefaria.org](http://www.sefaria.org) (accessed 16 January 2016).

Thiel, S. (2014). Visualizing Translation Variation: Designing Tools for Literary Scholars in Translation Studies and Linguistics. Masters Dissertation. Bauhaus University Weimar.

Version Variation Visualization. [www.tinyurl.com/vvex](http://www.tinyurl.com/vvex) (accessed 16 January 2016).

Yudlov, Y. (1997). *'Otzar ha-haggadot: Bibliografiah šel hagadot pesaḥ me-rešit ha-defus ha-ivri 'ad šenat tš"k [1960]*. Yerušalayim: Magnes.