

Leitwort Detection, Quantification and Discernment

Racheli Moskowitz, Moriyah Schick, Joshua Waxman

Stern College for Women, Yeshiva University

New York, NY

joshua.waxman@yu.edu

Abstract

English. A *Leitwort* (“leading word”) is a word or word-root deliberately repeated as a literary technique, for the sake of emphasis or to establish an underlying theme. *Leitworte* occur in many different works of literature, and traditional humanists who work to interpret those texts apply their literary expertise to discern those occurrences and explain their function. There is an element of subjectivity involved, and so, working from a digital humanities perspective, we developed an algorithm to detect potential *Leitworte* and rate their significance, by counting repetitions within a moving window and calculating tf-idf scores for candidate words. This reflects a convergence of humanistic and computer science methodologies. By compiling a list of *Leitworte* identified by an expert traditional literary scholar and then comparing it to the list produced by our computational approach, we attempt to (a) evaluate the subjective work of a particular human scholar against an objective standard, (b) consider whether the algorithmic approach we chose was sophisticated enough to match honed scholarly discernment, and (c) reflect on the difference between a traditional and digital humanities approach. In particular, we examine the Hebrew Bible and the *Leitworte* identified by Umberto Cassuto, an Italian Biblical scholar.

Italiano. Una *Leitwort* (“parola guidante”) è una parola o parola-radice deliberatamente ripetuta come tecnica letteraria, allo scopo di porre enfasi o stabilire un tema sottostante. Le *Leitworte* si presentano in molte diverse opere di letteratura, e gli umanisti tradizionali che lavorano per interpretare questi testi applicano la propria competenza letteraria per discernere tali occorrenze e spiegare la loro funzione. Si coinvolge un elemento di soggettività, e quindi, lavorando alla prospettiva degli studi umanisti digitali, abbiamo sviluppato un algoritmo per rilevare potenziali *Leitworte* e valutare la loro significatività, contando le ripetizioni all’interno di una finestra mobile e calcolando punteggi tf-idf per le parole candidate. Questo riflette una convergenza di metodologie umanistiche e informatiche. Compilando una lista delle *Leitworte* identificate da parte di uno studioso di letteratura tradizionale esperto e poi comparandola alla lista prodotta dal nostro approccio computazionale, tentiamo di valutare il lavoro soggettivo di un particolare studioso umano in confronto ad uno standard oggettivo; considerare se l’approccio algoritmico da noi scelto è abbastanza sofisticato da tenere testa ad un discernimento affinato dagli studi; e riflettere sulla differenza fra un approccio tradizionale e uno digitale agli studi umanistici. In particolare, esaminiamo e le *Leitworte* identificate da Umberto Cassuto, studioso biblico italiano.

1 Introduction

1.1 The *Leitwort* in Literature

In music, a *Leitmotif* is a recurring musical phrase (a “motif”) which is used to “lead” or guide the listener to recall something or make a connection. In literature, the *Leitwort* (“leading word”) plays a similar function. A certain word, or word root, is unusually repeated several times in a passage, in a way that jumps out at the reader, in order to establish a theme. Optionally, once established, it might then be echoed in a later passage to recall that theme. Here is one of the many *Leitworte* which Pinault (1986) identifies in his analysis of stylistic features in *The Arabian Nights*. In the story “The City of Brass”, over the course of a few consecutive pages, in poetry and prose, there is repetition of words with the Arabic root موت / *mawt*, meaning death.

أبادهم الموت / abadahum **mawt**, "Death destroyed them".

كأس الموت / ka's al-**mawt**, "the cup of doom".

إن كان موتي صلاة سجال / in kana **mawti** mahtuman cala cajal, "when my death was decreed all at once".

بإسم الحي الذي لا يموت / bi-ism al-hayy alladhi la **yamut**, "in the name of the Living, who dies not".

يُزخرفها الشيطان للإنسان إلى الممات / yuzakhrifuha al-shaytan lil-insan ila al-**mawt**, "Satan adorns it for man to lead him to death". (MacNaghten edition)

This literary technique has been discussed by scholars as it is found in sacred texts such as the Hebrew Bible (Alter, 1981) and the Koran (Wansbrough, 1978). It is similarly employed in the works of Goethe, Nietzsche, Heidegger and others. Then, one task of those who analyze these texts is to discern the use of a *Leitwort* and explain its purpose.

1.2 *Leitworte* and the Hebrew Bible

Various interpreters of the Hebrew Bible within the past century have taken note of the use of the *Leitwort*, though they differ as to its parameters and purpose. In Buber (1927), and in Buber and Rosenzweig (1936), the purpose of a thematic repetition in a passage is to reveal or clarify a meaning in the text, or to emphasize that meaning. They select *Leitworte* based on their subjective estimation of the word's significance, rarity, and the degree of repetition. These are all factors which contribute to a word capturing the attention of the reader. The repetitions can occur densely in a single passage, or can be distributed throughout the text. They draw connections between passages in which the same *Leitwort* occurs, as indicating an allusion or thematic echo. In one famous example from Buber, the seven scenes of revelation that compose the Abraham story arc are all tied together by the use of the term ראה / *ra'ah* / "see" in each passage, and unlike other translations which obscure this connection, Buber and Rosenzweig's German translation preserves the *Leitwort* by translating it consistently throughout.

At around the same time, Umberto Cassuto also took an interest in *Leitworte* in the Hebrew Bible. Cassuto was first the chief rabbi of Florence and subsequently a professor at University of Florence and at the University of Rome La Sapienza. Like Buber and Rosenzweig, Cassuto (1961, published posthumously) considered that *Leitworte* in the Bible served the purpose of establishing the theme of a passage or emphasizing a point. However, he was more selective in what types of repetitions he would consider to be bona fide *Leitworte*. In Cassuto's view, repetitions are only interpretable if they occur within a coherent passage and occur in a multiple of 3, 7, or 10. As illustrated below, he discusses the threefold repetition as a way of emphasizing a point within or between passages. The numbers seven and ten are chosen because they are particularly significant to an ancient Israelite author from a ritual and spiritual perspective. For instance, the number seven is echoed from the Creation narrative to a weekly cycle of seven days, a seven-year cycle before each Sabbatical year appears, and seven Sabbatical years leading up to the Jubilee. Similarly, the number ten brings the Ten Commandments immediately to mind.

We present an extended example to illustrate the position of *Leitworte* in Cassuto's textual analysis. In *Exodus* 2:2-6, in the background of Egyptian governmental decrees to drown all firstborn Hebrew boys, Moses is born. This story, Passage A, contains the word-root ראה / *ra'ah* / "saw", repeated three times:

Moses' mother "saw [וַתֵּרֶא / *vateire*'] him that he was a goodly child", hid him for as long as she could, and then placed him in an ark on the riverside. When Pharaoh's daughter visited the river, "she saw [וַתֵּרֶא / *vateire*'] the ark among the flags, and sent her handmaid to fetch it. (6) And she opened it, and saw it [וַתֵּרֶאֱהוּ / *vahir' eihu*], even the child; and behold a boy that wept. And she had compassion on him, and said: 'This is one of the Hebrews' children.'" (Jewish Publication Society Translation)

A few verses later (*Exodus* 2:11-15), Moses, after being raised as a prince in Pharaoh's house with his biological mother as his nursemaid, decides to check on his enslaved brethren, sees their suffering, and reacts. Once again, the thrice-repeated index root is *ra'ah*, though there are others marked. We label this Passage B.

יא ויהי בנמים ההם, ויגדל משה אל-אֶתְהוֹ, ויֵרָא, בסבלתם; ויֵרָא אִישׁ מִצְרַיִ, מִפְּהַ אִישׁ-עֲבָרִי מֵאֶתְיוֹ. יב ויִפְּן כֹּה וְכֹה, ויֵרָא כִּי אִין אִישׁ; ויֵרָא, אֶת-הַמִּצְרַיִ, ויִטְמְנֶהוּ, בַּחֹל. יג ויֵצֵא בַיּוֹם הַשְּׁנַיִ, וְהָיָה שְׁנַיִ-אֲנָשִׁים עֲבָרִים נֹצִים; ויֵאמֶר, לְרִשָׁע, לָמָּה תִכְהָ, רָעָה. יד ויֵאמֶר מִי שָׁמְךָ לְאִישׁ שָׂר וְשֹׁפֵט, עָלֵינוּ--הַלְהַרְגֵנִי אֵתָּה אִמֶּר, כַּאֲשֶׁר הִרְגֵת אֶת-הַמִּצְרַיִ; ויֵרָא מֹשֶׁה וַיֵּאמֶר, אָכֵן נֹדַע הַדְּבָר. טו וישמע פרעה את-הַדְּבָר הַזֶּה, ויִבְקֹשׁ לְהַרְגוֹ אֶת-מֹשֶׁה; ויִכְרַח מֹשֶׁה מִפְּנֵי פְרֵעָה, וישב בְּאֶרֶץ-מִדְיָן וישב על-הַבְּאֵר.

(11) And it came to pass in those days, when Moses was grown up, that he went out unto his brethren, and looked [*vayar*'] on their burdens; and he saw [*vayar*'] an Egyptian smiting [*makeh*] a Hebrew, one of his brethren. (12) And he looked this way and that way, and when he saw [*vayar*'] that there was no man, he smote [*vayakh*] the Egyptian, and hid him in the sand. (13) And he went out the second day, and, behold, two men of the Hebrews were striving together; and he said to him that did the wrong: 'Wherefore smitest [*takeh*] thou thy fellow?' (14) And he said: 'Who made thee a ruler and a judge over us? thinkest thou to kill me [*haleharegeni*], as thou didst kill [*haragta*] the Egyptian?' And Moses feared, and said: 'Surely the thing is known.' (15) Now when Pharaoh heard this thing, he sought to slay [*laharog*] Moses. But Moses fled from the face of Pharaoh, and dwelt in the land of Midian; and he sat down by a well. (Jewish Publication Society Translation)

Before we turn to Cassuto's analysis, a brief discussion of Passage B can help us understand the general phenomenon of *Leitwort*. Note that while ראה / *ra'ah*, "saw", is repeated three times, this is evident only when reading the passage in the original Hebrew, rather than in the English translation above. It appears twice in verse 11; the first time it is rendered as "looked" and the second time as "saw". Meanwhile, in verse 12, the word "looked" appears translating a different word-root, פנה / *panah*, while the word "saw" translates ראה / *ra'ah*.

These threefold repetitions seem deliberate. Firstly, an author can include or omit details while still advancing the narrative, so many times, a repeated word didn't truly need to appear. Secondly, Hebrew has synonyms. An author can select from an inventory of terms. For instance, in the same passage, the roots *nakhah* ("smite", "strike") and *harag* ("kill") are used to refer to Moses' killing of the Egyptian taskmaster.

In analyzing Passage B, Cassuto notes the threefold repetition of the root *ra'ah*. He says it is for emphasis. It matches the threefold repetition of the same root in passage A. The parallel is not accidental, but it is to stress that just as Moses' mother, and Pharaoh's daughter saw and had mercy on him, so did Moses take pity and have mercy on his brethren. Cassuto also notes the threefold repetitions of smiting and killing.

1.3 The Problem of *Leitwort* Operationalization

Despite the value of *Leitworte* as a literary technique that unifies a text and enriches the experience of the reader, any attempt to accurately identify *Leitworte* is somewhat problematic, particularly in a large and varied corpus such as the Hebrew Bible. The problem, essentially, is that language is repetitive by nature. Zipf (1945) discusses the relative frequencies of all words in a corpus, repetitions within clusters, as well as intervals between clusters, and these phenomena are observed in the absence of deliberate stylistic repetition.

Subject matter or narrative concerns can require a certain word to appear more than once in a story. For instance, the first chapter of Carlo Collodi's *Le Avventure di Pinocchio* contains repetitions of the Italian word *legno*, "wood" and *pezzo*, "piece". This is because it is where we first encounter a talking, weeping, and laughing piece of wood. Even if Collodi had no intention of drawing the reader's attention to these words, the narrative would be senseless without them. Luhn (1958) demonstrates that this occurs in non-narrative texts as well, and establishes the content of a text based on the non-deliberate repetition of words or word roots as an author advances his arguments or elaborates on an aspect of a subject. Additionally, certain words, such as function words (e.g., the articles "a" and "the"), are extremely common in language because they assist communication. These words will necessarily occur numerous times throughout a text, and will be totally unrelated to any theme that the writer wishes to emphasize. In fact, Luhn suggests that words that occur above a certain frequency threshold can generally be considered insignificant.

In the midst of a sea of non-significant word repetitions, identification of meaningful *Leitworte* for further study poses a challenge which can be addressed in one of two ways. The first solution is to maximize subjectivity, simply relying on a human interpreter to name which words should be considered *Leitworte*. This traditional approach has the benefit of embracing the nuances of human insight and the finely-honed skills of expert analysis, but can also be considered arbitrary and subjective by its very nature. An alternative solution, which we implement in this paper, is to introduce objective measurement tools in an attempt to quantify significance of repeated words in a given passage. Of course, our algorithm cannot truly appreciate a text, and standardized rules do not take intuitive understanding into account. However, our program has the advantage of working systematically to find every candidate *Leitwort*, in contrast to human experts who are subject to the limitations of their scanning and matching capability and who may be biased by selective interest in certain terms. This novel approach allows us to move towards greater objectivity in analysis, and is a great tool for scholars who want to consider every potential *Leitwort* in the text.

In the next section, we detail our programmatic approach. Our goals were to devise a quantitative measure of repetition significance, and to compare output of our program to the list of *Leitworte* identified by Cassuto, a traditional expert with a relatively systematic approach. To that end, our operational definitions are modeled on Cassuto's definitions but strip out the subjective components of his expert analysis.

2 Our Approach

For clear and consistent *Leitwort* identification, several practical questions must be answered. First, what constitutes a "word"? Second, what is a qualifying number of repetitions? Third, how closely spaced must the repetitions be? Fourth, how is significance of candidate words defined? Here we explain how we addressed each of these questions in designing our program.

2.1 Reduction to Lexemes or Roots

Since Semitic languages such as Hebrew are inflected, and the typical *Leitwort* is based on repetition of the word's root, we first reduce the words in the corpus to their lexemes. For the Hebrew Bible, we use the ETCBC dataset described in Roorda (2015), which contains manual marking of rich linguistic features by human experts. One such feature is the lexeme, which is a close approximation to the root. In Hebrew, most words contain a trilateral root which conveys a core meaning. For instance, אור / `or has the meaning of “light”. In the ETCBC dataset, words with this root are divided into two separate lexemes, אור / `or (“light”) and מאור / ma`or (“luminary”, thing which gives light). The lexemes are stemmed versions of the full word, stripping out definiteness, gender, number and person. Thus, the full word לְמֵאֹרוֹת / lim`orot / “as luminaries” in *Genesis* 1:15 is marked with the lexeme מאור / ma`or while the word לְהַאִיר / leha`ir / “to give light” in the same verse is marked with the lexeme אור / `or. The ETCBC dataset does not have a root feature.

We differ here from Cassuto, who primarily considers repetitions of roots. However, it is noteworthy that many of these lexemes are also the simple root (such as the אור example above). Of the 788 sevenfold lexeme-based *Leitwort* candidates our algorithm discovered across the Pentateuch, 81% consisted of trilateral roots. Many of the non-root lexeme candidates are names of nations or places.

2.2 Counting Repetitions

Following Cassuto, our candidate *Leitworte* must occur a multiple of 3, 7, or 10 times in a passage. We gravitate towards Cassuto's definition of *Leitwort* for a few reasons. Many modern Biblical interpreters (such as Elchanan Samet of Yeshivat Har Etzion) employ both Buber and Cassuto-type *Leitworte* in their analyses, but consider the more rigorously defined Cassuto-type *Leitworte* as especially significant (Grossman, 2011). Further, as discussed above, Cassuto shows that these are meaningful numbers for an ancient Israelite author, and he consistently demonstrates that thematic words are repeated this precise number of times, or a multiple thereof. Indeed, we are treating this threefold and sevenfold repetition as a mark of authorial deliberateness – that the author has set out to employ the *Leitwort* style. If a word were repeated by chance, simply because it is the topic of a passage (see the *legno* and *pezzo* examples above) or because it is a commonly occurring word (such as “said”), then it mostly would not occur specifically as a sevenfold repetition.

2.3 Scanning for Repetitions

We scan for repetitions in the text, in a moving window. For face validity, we require a certain minimum density of repetition. Buber did not require close proximity; the seven Abraham scenes that he connects with the root ראה / ra`ah / “see” span 178 verses over 8 chapters. Cassuto only identified *Leitwort* occurring within self-contained passages, but personally determined section and paragraph boundaries based on his own close reading analysis. Neither of these approaches is appropriate to our method, as both rely on subjective expert judgement to decide whether a given set of repetitions occurs within an acceptable space.

To define objective limits, we turned to the historical Jewish segmentation scheme of the *sidra*: the entire Pentateuch is chanted by a reader in synagogues over the course of a lunar year, one portion each week, on the Jewish Sabbath, though there are modifications due to holidays. The Pentateuch was divided into 54 such portions, or *sidrot*. While the calendar influenced the number of portions, scholars segmented the text at appropriate positions, such that there is often a consistency in the narrative or legal codes within the text. To make use of this narrative consistency, we only count repetitions within a *sidra*. Another Biblical segmentation scheme, of Christian origin, is the well-known series of chapter divisions (e.g. *Genesis* 1, *Genesis* 2), which breaks up the full text into chapters of about 30 verses each. We further require our repetitions to occur within a maximum window of 60 verses (approximately 2 chapters). Thus, if a word randomly occurs 4 times at the start of a *sidra* and much later has a sevenfold repetition, for a total of 11 occurrences, the sevenfold repetition will still remain a candidate.

Once a qualifying repetition has been found, we continue scanning along two pathways: one in which the passage stops with the most recent verse (and can thus be far smaller than 60 verses in length), and one in which it continues and allows for higher multiples to be identified.

Because our sections have flexible starting and ending points, a separate method must ensure that word appearances from other sections are kept distinct. In line with Cassuto's numerical definition of *Leitworte*, an eighth appearance of a candidate word in the same passage should disqualify the word. However, we would not wish to incorrectly disqualify a candidate merely because it occurs in an unrelated passage later in the

sidra. Using his idiosyncratic paragraph divisions, Cassuto would find a sevenfold repetition within a paragraph and ignore an unassociated occurrence one or two paragraphs earlier. Lacking such boundary lines, we create a buffer zone around each of our identified *Leitwort* passages. This zone is defined as $\frac{1}{4}$ the number of sentences of the passage span, and we require a total absence of the candidate word within that zone. Thus, a word will qualify as a *Leitwort* candidate if it occurs 7 times within a 16-verse span but does not appear at all in the 4 preceding and 4 subsequent verses. By requiring the word to appear in this “island,” we create de facto passage boundaries in a flexible way.

2.4 Filtering for Significance

Finally, we rate the candidate words for significance. In making estimation of significance the last step in our process, we diverge from the traditional expert-reader model. Scholars such as Cassuto and Buber would start with an impression that a word was significant, in the sense of meaningful and important. To Buber, if such a word was relatively rare (an undefined term) and also repeated within a story, it was a *Leitwort*. Cassuto required a precise number of repetitions and did not restrict based on rarity, but only examined words that he deemed especially significant rather than identifying every threefold or sevenfold repetition. Indeed, it would be simplistic to say that all of them are significant; these numbers can occur by chance just like any other.

Therefore, after systematically compiling a list of all sevenfold repetitions within our corpus, we employ a tf-idf measure to weed out the most clearly insignificant of them. The term frequency (tf) is the number of times a word appears in a given document, while the inverse document frequency (idf) is the log of the total number of documents N divided by the number of documents that contain the word. If a word is frequent in the current document and infrequent elsewhere, then the product of the tf and the idf will be high. The “documents” we use for this computation are the 54 aforementioned *sidrot*, since the text in each such division will typically be of a consistent genre (e.g. genealogy, legal code, narrative) and topic (e.g. trials in the wilderness).

We stress that the purpose of this tf-idf ranking is not to discover the emphatic and thematic words. The specific numerical repetition establishes that. Rather, our aim was to filter out common yet highly insignificant words, which will occur in sevenfold repetition (along with eightfold repetition, ninefold repetition, etc.) purely by chance. For this reason, we set our tf-idf threshold very low, at 0.07. After examining a small portion of unfiltered candidates, we chose this value because it could retain words that appeared thematically relevant, while excluding common words with high frequency throughout the corpus but no discernable relevance to the specific passage. We did not use a simple stoplist of frequent words since a common word might be extremely significant in a given context. For instance, in *Genesis* 1-2, in which God creates the Universe in a sequence of speech acts, the lexeme אָמַר / *amar* / “said” occurs 28 times (a multiple of 7) and has a tf-idf score of 0.13, above our threshold of significance. It also occurs seven times in *Deuteronomy* 5-7 with a non-significant tf-idf score of 0.04.

3 Results

In the five books of the Hebrew Bible, we discovered a total of 788 potential *Leitwort* candidates that appeared a multiple of seven times in an island of text. Of these, 332 (or 42%) exceeded our tf-idf threshold and were counted as significant. Passage span ranged from 5 to 60 verses; and candidate lexemes were repeated within these passages 7, 14, 21, 28, 35, 42, or 49 times. As would be expected, threefold repetitions had shorter passage spans on average, and many fewer of them were deemed significant. We compiled a comprehensive list of Cassuto’s *Leitworte* and compared them against the output of the program. Cassuto wrote commentary on the first 13 chapters (out of 50) of *Genesis* and on the entire (40 chapter) book of *Exodus*, identifying 164 *Leitworte*, of which 142 were of simple word or root repetitions. Of these root repetitions, 59 represented a sevenfold recurrence. For the same group of chapters, we found 207 potential candidates appearing a multiple of seven times, of which 102 (49%) exceeded our tf-idf threshold.

Table 1 cross-tabulates our results with Cassuto’s. Twenty words were deemed significant by our program and also discussed by Cassuto, 82 are marked at *Leitworte* by our program only, 39 by Cassuto only, and 105 words that do not appear in Cassuto’s work were originally flagged by our program but fell below our significance threshold.

		Algorithm	
		Yes	No
Cassuto	Yes	<i>Total: 20</i> <i>Genesis: 8</i> <i>Exodus: 12</i>	Total: 39 Genesis: 18 Exodus: 20
	No	Total: 82 Genesis: 17 Exodus: 65	<i>Total: 105</i> <i>Genesis: 23</i> <i>Exodus: 82</i>

Table 1: Cross-tabulation of the results of Cassuto and our algorithm. “Yes” means that it appears in the list (and, for the algorithm, deemed significant). Cells representing agreement of the two sources are italicized.

A few facts are apparent from these results. We see that our algorithm identified many more potential *Leitworte* overall than Cassuto. Also, Cassuto and the algorithm agreed about 50% of the time, and were much more likely to agree that a word was non-significant than that it was significant. Cassuto discussed many *Leitworte* that were not accepted by the algorithm, and vice versa. Finally, it is noteworthy that the results differ substantially based on specific text. Cassuto described almost as many *Leitworte* in the first 13 chapters of Genesis as in the 40-chapter Exodus. Meanwhile, the algorithm flagged repetitions with similar density across the two books and consistently identified about half of them (52% in Genesis, 48% in Exodus) as potentially significant. Cassuto and the algorithm therefore find about the same number of Genesis *Leitworte*, with few of Cassuto’s appearing in the computer-generated list, whereas the algorithm finds more than twice as many Exodus *Leitworte* as Cassuto does.

If Cassuto’s work is held up as the gold standard, one can say that the algorithm achieved 19.6% precision (32.0% in Genesis, 15.6% in Exodus) and 33.9% recall (30.8% in Genesis, 37.5% in Exodus). This suggests that it is able to catch about a third of the *Leitworte* discerned by an expert, and introduces a high number of spurious candidates, particularly in Exodus. Valid *Leitworte* can be missed by the program either because they are never identified or because they are rejected as insignificant. Close inspection of the data reveals that only 5 of Cassuto’s *Leitworte* that were flagged by our algorithm fell below our tf-idf significance threshold. Most did not meet the algorithm’s criteria for being a sevenfold repetition. This may be because we were restricted to using lexemes while Cassuto primarily used roots or was more flexible about linguistic features, or because we lacked his sharp boundaries of paragraph and story. Therefore, we may have inadvertently cut off our “passages” before the end of a scene, or disqualified a true *Leitwort* because it re-appeared in an unrelated context within our buffer zone. Further work can address some of these issues.

If, on the other hand, the objective algorithmic approach is considered the ideal, one can say that Cassuto obtained 33.9% precision (30.8% in Genesis, 37.5% in Exodus) and 19.6% recall (32.0% in Genesis, 15.6% in Exodus). This suggests he found about a fifth of possible *Leitworte* in his chosen text, and that about a third of his self-defined *Leitworte* are valid. Due to the limits of human attentional capacity, it would be practically impossible for a person to manually identify all existing *Leitworte* in such a complex text. Humans can be biased by their own interests to overlook many details, which can lead both to false positive and false negative detection errors. Notably, Cassuto’s list for Genesis, the text in which he first perceived *Leitworte* and which evoked tremendous enthusiasm for the task, has the highest recall and lowest precision compared against the algorithm.

The truth probably lies somewhere between these extremes. Dismissing 80% of the algorithm’s suggestions as invalid merely because Cassuto did not talk about them ascribes omniscience to the human expert, which is absurd. Similarly, it is ridiculous to say that Cassuto’s analysis is only meaningful if its tf-idf score falls above our program’s cut-off. The low overlap between Cassuto’s results and the algorithm’s is evidence that the two methodologies bring different perspectives and different strengths. Only a human being can explain the meaning of a *Leitwort* in context and weave it into a consistent tapestry with other methods of literary analysis. However, the ability to systematically evaluate every instance, and to apply objective criteria undiluted by personal bias, are core benefits of computerized *Leitwort* detection. The best use of such digital tools will be to allow merging of these two approaches by using algorithms before or after the human eye. Modern scholars of Biblical literature might use our program to systematically generate a list of repetitions to consider in their analyses, or consult its quantitative information (e.g. tf-idf scores) to consider whether their initial impressions might be distorted and in need of further scrutiny. Thus, objective methodology can become a thread woven into the subjective tapestry.

References

- 1917, *The Holy Scriptures According to the Masoretic Text: A New Translation*, The Jewish Publication Society of America, Philadelphia
- Nevin Reda Al-Tahri, 2012. Textual Integrity and Coherence in the Qur'an: Repetition and Narrative Structure in Surat al-Baqara. Doctoral Thesis, University of Toronto.
- Robert Alter, 1981. *The Art of Biblical Narrative*, Basic Books, New York, NY
- Ronald D. Anderson, "Lietworter in Helaman and 3 Nephi," in *The Book of Mormon: Helaman Through 3 Nephi 8, According to Thy Word*, ed. Monte S. Nyman and Charles D. Tate, Jr. (Provo, Utah: Religious Studies Center, Brigham Young University, 1992) 241–250.
- Martin Buber, 1927 lecture, Leitwort Style in Pentateuch Narrative.
- Martin Buber and Franz Rosenzweig, 1936. *Die Schrift und ihre Verdeutschung*, Schocken, Germany.
- Umberto Cassuto, 1964, *A Commentary on the Book of Genesis (Part II): from Noah to Abraham* (Israel Abrahams, Trans.). Magnes Press, Jerusalem, Israel
- Umberto Cassuto, 1967, *A Commentary on the Book of Exodus* (Israel Abrahams, Trans.). Magnes Press, Jerusalem, Israel
- Umberto Cassuto, *Peirush al Sefer Bereishit*, 1965. Jerusalem, Israel.
- Umberto Cassuto, *Peirush al Sefer Shemot*, 1974. Jerusalem, Israel.
- Umberto Cassuto, 1934, La creazione del mondo nella Genesi, *Annuario di studi ebraici*, vol. i, pp. 47–49
- Umberto Cassuto, 1961, *A Commentary on the Book of Genesis (Part I): from Adam to Noah* (Israel Abrahams, Trans.). Magnes Press, Jerusalem, Israel
- Yonatan Grossman, 2011. Literary Study of Biblical Narrative, Lecture 11, Leitwort, Part I, The Israel Koschitzky Virtual Beit Midrash, <https://www.etzion.org.il/en/leitwort-part-i>
- Hans Peter Luhn, 1958. The automatic creation of literature abstracts. *IBM Journal of Research and Development*, 2, 159-165
- William Hay MacNaghten, 1839-1842. *The Alif Laila or Book of the Thousand Nights and One Night*, 4 vols, W. Thacker and Co., Calcutta
- David Pinault, 1986, Stylistic Features in Selected Tales From "The Thousand and One Nights", Doctoral Dissertation, University of Pennsylvania
- David Pinault, 1992, *Story-Telling Techniques in the Arabian Nights. Studies in Arabic Literature*, volume xv. Brill, Leiden, NY
- Dirk Roorda, 2015. The Hebrew Bible as Data: Laboratory - Sharing – Experiences, <https://arxiv.org/pdf/1501.01866.pdf>
- John Wansbrough, 1978. *The Sectarian Milieu: Content and Composition of Islamic Salvation History*, Oxford University Press, London.
- George Kingsley Zipf, 1945. The Repetition of Words, Time-Perspective, and Semantic Balance. *The Journal of General Psychology*, 32(1), 127–148.