Davide Saponaro - Emiliano Giovannetti - Flavia Sciolette

FROM RELIGIOUS SOURCES TO COMPUTATIONAL RESOURCES: APPROACH AND CASE STUDY ON HEBREW TERMS AND CONCEPTS

1. *Introduction*

Flavia Sciolette

In recent years, a growing attention for digital humanities has interested the sector of social and religious studies. Researchers have begun adopting approaches that combine traditional methodologies with technologies grounded on digital resources. In the following, we take into account resources based on knowledge data, with a particular focus on ontologies, and their application in religious studies.

Though "ontology" is a term borrowed from metaphysics, where it represents the philosophical study of being, in the context of computer science an ontology can be defined as a formal representation of a domain, consisting of concepts, individuals (representing entities) and relations holding between them. An ontology is also conceived to provide a shared vocabulary to be adopted and used between different communities: as a matter of fact, one of the most widespread definitions has been provided by Gruber,[1] who defined an ontology as a "specification of a shared conceptualization".

In the area of Jewish religion, we can rely on three extensive and complex resources: i) the ontology developed by Foundation Contemporary Jewish Documentation Center (CDEC);[2] ii) JudaicaLink, the resource of Fachinformationsdienst (FID) Jüdische Studien;[3] iii) the ontology of digital library Sefaria.[4]

The first resource is a domain ontology, focussed on Shoah, in particular about persecutions and deportations that occurred in Italy between 1943 and 1945. The second contains structural domain-based information (concepts and relations) about Jewish culture and history derived from various sources, including geographical datasets, image collections, databases about cultural heritage, libraries, and links to DBpedia for named entities.[5] Both resources are available as Linked Open Data (LOD).[6] The third resource represents the index of themes of Sefaria,[7] the widest collection of digitised texts about Jewish culture and relative tools to analyse them, such as dictionaries and digitised lexicons. The ontology of Sefaria is based on BFO (Basic Formal Ontology),[8] with a perspective of interoperability. In addition to the aforementioned resources we also cite the Ontology of Masters of

[1] T.R. Gruber, *Toward principles for the design of ontologies used for knowledge sharing?*, «International Journal of Human-Computer Studies» 43, 5-6 (1995), pp. 907-928, https://doi.org/10.1006/ijhc.1995.1081.

[2] <http://dati.cdec.it/indiceEN.html>.

[3] <http://web.judaicalink.org/>.

[4] <https://github.com/Sefaria/Sefaria-Project/wiki/Documentation-for-Topics-Hierarchy>.

[5] L. Brazzo, S. Mazzini, *Linked Open Data per l'analisi dei dati e lo sviluppo della ricerca sulla Shoah in Italia*', in *Quaderni di Umanistica Digitale*, *AIUCD2017 - Book of Abstracts*, 2017, doi: 10.6092/unibo/amsacta/5885; K. Eckert, M. Dadvar, *JudaicaLink: A Knowledge Base for Jewish Culture and History*', «Umanistica Digitale» 4 (2019), doi: http://doi.org/10.6092/issn.2532-8816/9047.

[6] A resource as LOD adheres to a series of principles, in particular: i) data is released under open licences, ii) each element is uniquely identified by means of a URI (Uniform Resource Identifier), and iii) it can refer to and be linked by other resources <https://lod-cloud.net/>.

[7] <https://www.sefaria.org/>.

[8] <https://github.com/bfo-ontology/BFO/wiki/>.

the Babylonian Talmud, a knowledge base, the construction of which is still in progress, about the rabbis cited in babylonian Talmud.[9]

In the field of Christian studies, we first mention New Testament Names (NTN),[10] a knowledge base describing named entities (such as names of people, cities, organisations, etc.) appearing in the New Testament. Additionally, we also cite some recent works on Information and Knowledge Retrieval for Biblical Texts, in particular: Popa and Goga for the use of Text2Onto[11] for automatic extraction of concepts from semi-structured data, and the work of Varghese and Punithavalli[12] for the application of text mining techniques on religious texts. It is also noteworthy to mention Logos,[13] a proprietary software for browsing and searching the Biblical Text and a large set of linked resources (commentaries, dictionaries, etc.) supported by tools for linguistic analysis; amongst the features offered by the tool, we mention Bible Sense Lexicon (BSL), a lexicalised ontology for Greek and Hebrew terms, organised by concepts appearing in the New Testament.

Finally, we cite the resources relative to Islamic studies. Two ontologies, concerning themes from the Quran, have been developed by the University of Leeds. Quran Analysis[14] is a platform containing the Quran Ontology[15] (902 concepts and 1628 relations), while the Quranic Arabic Corpus[16] is a collection of richly annotated texts, with an ontology of base-concepts of the Koranic domain; additionally, we can mention the work of Taa *et alii* for the Al-Quran ontology,[17] developed with a process of automatic extraction of knowledge from the Quran. Instead, the Hadith Ontology described in Fairouz *et alii*[18] represents a lightweight ontology based on knowledge extracted from the *ḥadīṯ*, the source of normative prescriptions for the Islamic religion. This ontology was built with METHONTOLOGY[19] a vocabulary for the construction of knowledge bases.

All the above-mentioned initiatives and projects are relative to knowledge-related resources in the religious domain and the tools used to exploit them. However, the present work is methodological, and it is conceived to illustrate an approach for the creation of such types of resources. Another point of difference lies in the nature of resources themselves: the proposed building approach, as described in the following section, involves the integration of textual, linguistic and conceptual information in the very same resource, allowing to link to each other and manage different kinds of data representing a case study in the religious domain.

[9] E. Giovannetti, D. Albanesi, A. Bellandi, D. Dattilo, A. Del Grosso, S. Marchi, *An ontology of masters of the Babylonian Talmud*, «Digital Scholarship in the Humanities» 37, 3 (2021), pp. 725-737, doi: 10.1093/llc/fqab043. The resource is available on https://licodemo.ilc.cnr.it/resources/rabbiontology/#.

[10] <https://www.semanticbible.com/index.html>

[11] R.C. Popa, M.N. Goga, *Extracting Knowledge from the Bible: A Comparison between the Old and the New Testament*', in *International Conference on Automation, Computational and Technology Management (ICACTM)*, 2019, doi: 10.1109/ICACTM.2019.8776828. For Text2Onto, a framework for ontology creation from textual resources, cf. P. Cimiano, J. Völker, *Text2Onto. A Framework for Ontology Learning and Data-driven Change Discovery*, in A. Montoyo, R. Muńoz, E. Métais, (eds.), *Natural Language Processing and Information Systems. NLDB 2005. Lecture Notes in Computer Science*, Springer, Berlin-Heidelberg, https://doi.org/10.1007/11428817_21.

[12] N. Varghese, M. Punithavalli, *Lexical And Semantic Analysis Of Sacred Texts Using Machine Learning And Natural Language Processing*, «International Journal of Scientific & Technology Research» 8 (2019), pp. 3133-3140.

[13] <https://www.logos.com/>.

[14] <http://www.qurananalysis.com/>.

[15] <http://www.quranontology.com/>.

[16] <http://corpus.quran.com/ontology.jsp>.

[17] A. Taa, M. Ahmad, Q.A. Abed, *Al-Quran ontology based on knowledge themes*, «Journal of Fundamental and Applied Sciences» 9, 5S (2017), pp. 800-817.

[18] B. Fairouz, T. Nora, A.A. Nouha, *An Ontological Model of Hadith Texts. Semantic Representation of Hadith*, «International Journal of Advanced Computer Science and Applications» 11, 4 (2020), pp. 367-371.

[19] M. Fernandez, A. Gomez-Perez, N. Juristo, *METHONTOLOGY: from ontological art towards ontological engineering*', in *AAAI-97 Spring Symposium Series*, EEUU, Stanford University 1997, pp. 33-40.

This work is structured as follows: the next section introduces the model and the proposed methodology; the third section concerns the selected case study; the fourth section presents the computational resource and its components; the fifth section illustrates the tools employed for the development of the resource; the last section discusses the advantages of the proposed approach and outlines future works.

2. *Model and methodology*

Emiliano Giovannetti

The basic aim of this work is the introduction of a model[20] and a methodology for the integrated representation of textual, terminological and conceptual entities in the context of religious studies. The idea is focussed on the creation of a resource representing the computational representation of all the relevant information. This resource is conceived with the precise purpose of providing an innovative way for the analysis of texts by *expliciting* their content.

In this context, we define as *textual information explicitation* the process that allows to associate a set of textual elements present in a corpus to formal entities of a computational resource belonging to specific classes. To bring an example, a single textual occurrence, such as "face", can be annotated as a "domain term" and linked to an entity belonging to the class of "Terms", the instances of which represent terms of a particular domain having specific semantic features. The very same textual occurrence, however, may be also annotated as a morphological form and linked to an instance of the class representing all lexical forms of a particular language. More details about this "multi level" structuring will be provided below.

This process of explicitation aims at the discovery of all the informative entities without separating them from the text they belong to (as is usually done in "extraction" processes). The process, as somehow the name implies, is not limited to the explicit information mentioned in the text, but also includes all the "contextual", implicit data which is essential to fully represent the content of a text.

As anticipated, the explicitation process is conceived to classify each textual element into a specific set of the resource. In general, when studying a text, a number of different "levels" of analysis must be taken into account (documental, linguistic, historical, social, etc.) in order to find the relevant information pertaining the intended research purposes.

Accordingly, a computational resource conceived to represent the information conveyed by a text in a structured and integrated way must be organised in a number of distinct but interrelated "areas". In other words, the resource must be grounded on a properly expressive model. Such a model must provide the general framework necessary to formalise and sort the various textual entities in specific classes and to link them to each other. The proposed model is divided into a series of *systems*, each of which in charge of representing a specific aspect of text:[21] in this paradigm, the text takes the form of a "set of systems". The systems that will be taken into account in this approach are the *linguistic*, the *conceptual*, and the *graphic* systems: the overall model includes also a *discursive* system (designed to represent formulaic data, such as dates, idiomatic expressions, etc.) and a *documental* system (to represent metadata and structural information), which, however, will not be considered in this work. In addition, each single system is organised into *dimensions*, conceived to collect data of homogeneous nature inside a system. For example, the linguistic system is composed of several dimensions, among which the *terminological* dimension (made up by terms belonging to a specific domain) and the *lexico-semantic* dimension (represented by the set of lexical senses).

[20] F. Sciolette, E. Giovannetti, *Un modello per domarli tutti: verso una rappresentazione del testo come esplicitazione di documento, lingua e contenuto*, in *Fare linguistica applicata con le digital humanities*, a cura di J. Saturno - L. Spreafico, AItLA, Milano 2022, pp. 145-157.

[21] For a use of the concept of "system" in a modelisation of a text cfr. T. Orlandi, *Informatica testuale: teoria e prassi*, Laterza, Roma 2010.

In more detail, the graphic system is constituted by elements that can be represented as sequences of alphanumeric characters. In this system, returning to our previous example, "face" is not to be considered a word, but an ordered sequence of the characters "f", "a", "c", and "e".

On the other hand, the linguistic system is constituted by linguistic elements, such as morphemes, words, lexemes, terms, lexical senses, and so on. The sequence "face" must be here interpreted as an English word belonging, specifically, to the *lexical* dimension. At the same time, in a particular domain (in this case, the religious domain) "face" can also belong to the *terminological* dimension, being a domain term. In this sense, each *token* (to be intended as a textual unit) of the graphic system, can be analysed and interpreted in many ways, depending on the various "glasses" a reader chooses to wear. On the basis of this analogy, the token "face" can be interpreted in different ways when "wearing" the glasses of the *semantic* dimension, since the meaning of "face" in the two sentences "*the face of the woman*" and in "*the face of the house*" refer to distinct lexical senses.

In our model, these lexical senses belonging to the linguistic system must be eventually linked to concepts formalised in the *conceptual* system. Following our example, this system will have to include a concept describing an entity of anatomical nature, located in the front of a human head, etc. that, in English, can be lexicalized with the word "face" and in Italian with words "volto", "viso", and "faccia".

Now that our model of reference has been briefly introduced, we are going to describe the analytical methodology which will serve as a guideline to the construction of the computational resource. The methodology can be decomposed in the following steps:

1. *definition of the case study*, which, in turn, can be subdivided into the following steps
   a. definition of the research context (in our case, the study of terms and concepts of the religious domain);
   b. definition of the corpus of reference (for our purposes, the religious texts introduced in Section 3);
   c. definition of the specific case study to analyse and formalise in the form of computational resource;
2. *selection of the most significative textual fragments* conveying informative elements useful for the predetermined case study; all the entities populating the linguistic system will refer to these textual portions containing their attestations; depending on the availability of data already digitally formalised, it will also be possible to provide automatic suggestion mechanisms for the selection of pertinent textual parts;
3. *structuring of the selected informative elements in the computational resource* grounded on the aforementioned model; as exemplified in Section 4, each element of the resource (portion of text, term, sense, concept) will have to refer, directly or indirectly, to a specific place of the corpus defined in 1b); the formalisation of these relations is particularly important to:
   a. document the provenance and guarantee the reliability of the formalised data (taking into account, of course, the subjective and interpretative components introduced by the scholar);
   b. allow a scholar to precisely access the corpus through advanced (linguistic and conceptual) searches fully exploiting the information encoded in the resource;
   c. allow a natural language processing algorithm to take advantage of all the available information when analysing a specific textual passage.

The construction of the computational resource, achieved by following the proposed methodology and the relative steps, can be greatly facilitated by the availability of digital tools a scholar can rely on. As a matter of fact, as will be detailed in Section 5, we are also working at the development of an environment for the construction of resources capable of integrating, in the very same workspace, the management of corpora, lexicons, and ontologies.

In the two following sections the approach will be exemplified through a case study. An analysis of the Hebrew word *panim* in some of its different meanings and appearing in different contexts will be provided, followed by some comparative considerations with analogous words in other languages. This "traditional" approach to the study of words will be followed by its translation in the relative computational representation, described in Section 4.

3. *A case study*

Davide Saponaro

3.1. Panim: *many faces of meaning*

The word *panim*, basically meaning "face", is widely present in the Hebrew language throughout its history, either in its basic form or combined with prefixes and suffixes, covering an array of meanings stretching far beyond the idea of "the front of the head". Thus, this term and its polysemy were a natural source of interest for our work. In this part of our article, we will try to briefly summarise some of the meanings this term and its constructs can assume in the texts we examined. At the end, a short comparison will be made with some attestations of polysemy of the word "face" in Christian and Muslim sacred texts.

The triliteral root of *panim* is *peh-nun-yod*. The corresponding verb is *panah*, whose basic meaning is "to turn oneself". The word is a noun, can be masculine or feminine and only appears in the plural form (a phenomenon known elsewhere in Hebrew, for example *ḥayyim*, "life", or *mayim*, "water"). The word's construct form is *pene*. We also find that the same combination of radical letters, when vocalised *penim* (*peh* vocalised with *shewa* instead than *qamaṣ*), means "inside", "interior". The word *panim* could also have some (not totally clear) connection to *peh*, "mouth". The same root is already attested in Akkadian as *panu*, and it's highly polysemic.[22]

As we will see later an equally wide, or maybe even wider, range of meanings of this term is displayed throughout the whole history of Hebrew language. In particular, combining the word with prepositions can give birth to an extremely wide semantic palette.

We would now like to spend a moment to see how the word *panim* is explained in some Hebrew dictionaries. These dictionaries' entries can give us an idea about how this word's meaning expanded during the long history of Hebrew language, and how it was perceived by scholars of different ages. In first place, we would like to clarify that the following examples are neither aimed at providing a complete list of the meanings the term *panim* assumed during its history, nor at drawing any scientific conclusion. Their aim is just to provide us, from a bird's eye view, a glance of the rich history of this term's polysemic character. The dictionaries entries were not reported literally: they have in many cases been shortened for space reasons, and in at least one case translated from Hebrew.[23]

Marcus Jastrow's *A Dictionary of the Targumim, Talmud Babli and Yerushalmi and Midrashic Literature* (1926) explains the term as follows: **1** front, face, countenance, person. **2** *panim šel maṭṭah*: pudenda. **3** aspect, manner, way of interpretation. **4** *lifne*: for appearance sake, formally. b) in front (of time), before this, in the past. *-lifne ʿiwwer*: the law prohibiting an act that may bring a person to sin; - *mipenè*: on account, for the sake of. *mipenè še-* because. *-bifne*: in the presence of.[24]

The *PESHAT* database, created by the University of Hamburg and mainly devoted to premodern Hebrew lexicon and terminology, lists 7 meanings: **1** reason; ground. **2** way; as-

[22] In the *Assyrian Dictionary of the Oriental Institute of the University of Chicago* the term is listed as follows: *panu* A s.; 1. front, front part, 2. surface, 3. looks, appearance, 4. ranking position, 5. past, past time, 6. wish, choice, intention, plan, purpose, concern, consideration, opinion, 7. reciprocal (math. term), 8. *panū* (pl.) face, visage, 9. *panū* (pl.) dignity, prestige (*The Assyrian Dictionary of the Oriental Institute of the University of Chicago*, volume 12 (P), Chicago 2005, available online at <https://oi.uchicago.edu/sites/oi.uchicago.edu/files/uploads/shared/docs/cad_p.pdf>). It could be worth noting that the last two entries show the word in its plural form, that in Hebrew will become exclusive. The word is also attested in Phoenician (*pnm*), Moabite (*mpny*, *lfny*), Punic (*pn*), Ugaritic (*pnm*) (E. Klein, *A Comprehensive Etymological Dictionary of the Hebrew Language* - Carta Jerusalem 1987).

[23] Friedrich Wilhelm Gesenius' *Hebrew and Chaldee Lexicon to the Old Testament* (1846) devotes more than four whole pages just to this lemma. This is far too much to be reported here, but can give us an idea about the sheer quantity of meanings and examples provided in one of the first modern Hebrew dictionaries. The *Lexicon*, first published in German in 1815, had a high number of editions. We consulted the one available at <http://www.tyndalearchive.com/TABS/Gesenius/index.htm>.

[24] <http://www.tyndalearchive.com/TABS/Jastrow/>.

pect; manner. **3** face. **4** metaphorical, knowledge of the Divine; Metaphysics **5** cheek. **6** decan, a subsection of a zodiac sign equal to ten degrees. **7** front side.[25]

Ernest Klein's *A Comprehensive Etymological Dictionary of the Hebrew Language* (Carta Jerusalem, 1987) lists 6 basic meanings for *panim*: **1** face, countenance. **2** forepart, front part, front. **3** surface, level. **4** appearance. **5** manner, way. **6** wrath, anger.[26]

According to the online dictionary of the *Academy of the Hebrew Language* (shortened in *AHL*; the translation from Hebrew is ours), the meanings are as follows: **1** the front part of the head. **2** front side. **3** appearance. **4** outer or upper surface, such as "the surface (lit. "face") of water". **5** mode, side, and also modes, sides (of a subject or problem). **6** in conjunction with prepositions, it is used to express relationships of time, place, cause, etc.[27]

From this brief exposition of dictionaries' entries, we can already obtain an idea of how polysemic the term has been over the course of the centuries, and how it acquired and lost some of its meanings over the ages. The listing order of the various meanings in dictionaries could also be an interesting field of research. However, exploring these fields goes well beyond the scope of the present work and its possibilities.

On our part, with the purpose of showing the approach we defined and without the ambition of arriving at definitive conclusions – a task we gladly leave to specialised scholars – we would like to present a short and non-exhaustive list of some of the senses the word *panim* can acquire. In this case, we chose to limit our research field to the Torah. A decision that was taken precisely because, as we stated earlier, this work is not aimed at carrying out an exhaustive research, but rather at providing a sample of our approach's application: a task better carried out on a limited field of action. As an introduction, we would like to show one verse we found particularly inspiring: Gen. 32:21.[28] The context is the following: Jacob, coming back to the land of Cana'an after the years spent in Ḥaran, is preparing to meet his brother Esau, that wanted to kill him last time they met. The verse in Hebrew states:

וְאַמַרְתֶּם גַּם הִנֵּה עַבְדְּךָ יַעֲקֹב אַחֲרֵינוּ כִּי־אָמַר אֲכַפְּרָה פָנָיו
בַּמִּנְחָה הַהֹלֶכֶת לְפָנָי וְאַחֲרֵי־כֵן אֶרְאֶה פָּנָיו אוּלַי יִשָּׂא פָנָי׃

In current English[29] the word "face" appears just once:

And say moreover, Behold, thy servant Ya'aqov is behind us. For he said, I will appease him with the present that goes before me, and afterwards I will see his face; perhaps he will accept me.

However, a more literal translation can highlight the widespread use of the word *panim*, with several meanings, here:

And say moreover, Behold, thy servant Ya'aqov is behind us. For he said, I will atone his face with the offering going before my face and after that I will see his face; perhaps he will lift my face.

This Biblical verse can well serve as a sort of access point to the many meanings of the word *panim*. In the present work, we defined five main senses of the term. The list does not

[25] <https://www.peshat.org/display/peshat_lemmas_00003532>.

[26] <https://www.sefaria.org/Klein_Dictionary%2C_%D7%A4%D6%B8%D6%BC%D7%A0%D6%B4%D7%99%D7%9D.1?lang=bi&with=all&lang2=en>.

[27] <https://hebrew-academy.org.il/keyword/%D7%A4%D6%BC%D6%B8%D7%A0%D6%B4%D7%99%D7%9D>. One cannot avoid mentioning the *Ma'agarim* website, also by the Academy of Hebrew Language, and its Historical Dictionary: an enormously useful resource for researching Hebrew throughout its historical phases (in Hebrew: <https://maagarim.hebrew-academy.org.il/Pages/PMain.aspx>).

[28] We owe this inspiration to the late Rabbi Lord Jonathan Sacks in his book "Not in God's Name - Confronting Religious Violence" (Schocken Books, New York 2015; published in Italian as "Non nel nome di Dio", Giuntina, Firenze 2017).

[29] The translation we used for all Hebrew texts in this article is from the Koren Jerusalem Bible, available on <sefaria.org>.

have the pretension of covering all the meanings the term can have; rather, it defines a limited field of action to work upon. For each sense, we signalled where in the dictionary entries cited above it appears or is alluded to:

1) anatomic face of a person (see Jastrow point 1; PESHAT 3; Klein 1; *AHL* 1);

2) the front part or surface of something (see PESHAT 7; Klein 2; *AHL* 2 and 4). This sense originates from the fact that the front part or surface of an object is in some way its "face", its exposed part;

3) presence, countenance, appearance. It's worth mentioning that this sense in particular, when the term is in conjunction with particles such as *mi-*, *le-*, *bi-*, *el*, *'al* or their combinations, can generate a quite wide range of spatial indications (see Jastrow 3 and 4; Klein 1 and 4; *AHL* 3 and especially 6);

4) with the preposition *mi-*, and frequently followed by the particle *asher/she-*, "because of" (see Jastrow 4; PESHAT 1; *AHL* 6);

5) an indicator of emotions or behaviours (see Klein 6).

Following, a short selection of verses from the Torah illustrating each of the five groups of senses. They are listed according to their order of appearance, and below each one we listed the sense that, according to our opinion, it fits in.

1	וְהָאָרֶץ הָיְתָה תֹהוּ וָבֹהוּ וְחֹשֶׁךְ עַל־פְּנֵי תְהוֹם וְרוּחַ אֱלֹהִים מְרַחֶפֶת עַל־פְּנֵי הַמָּיִם׃	And the earth was without form and void; and darkness was on the **face** of the deep. And a wind from God moved over the **surface** of the waters. (Gen. 1:2) **Sense: 2**
2	וְאֶל־קַיִן וְאֶל־מִנְחָתוֹ לֹא שָׁעָה וַיִּחַר לְקַיִן מְאֹד וַיִּפְּלוּ פָנָיו׃	but to Qayin and to his offering He had not respect. And Qayin was very angry, **and his face fell**. (Gen. 4:5) **Sense: 5**
3	וַיֹּאמֶר ה׳ אֶמְחֶה אֶת־הָאָדָם אֲשֶׁר־בָּרָאתִי מֵעַל פְּנֵי הָאֲדָמָה מֵאָדָם עַד־בְּהֵמָה עַד־רֶמֶשׂ וְעַד־עוֹף הַשָּׁמָיִם כִּי נִחַמְתִּי כִּי עֲשִׂיתִם׃	And the Lord said, I will destroy man whom I have created from **the face of the earth**: both man, and beast, and creeping things, and the birds of the air; for I repent that I have made them. (Gen. 6:7) **Sense: 2**
4	וַתֹּאמֶר רִבְקָה אֶל־יִצְחָק קַצְתִּי בְחַיַּי מִפְּנֵי בְּנוֹת חֵת אִם־לֹקֵחַ יַעֲקֹב אִשָּׁה מִבְּנוֹת־חֵת כָּאֵלֶּה מִבְּנוֹת הָאָרֶץ לָמָּה לִי חַיִּים׃	And Rivqa said to Yiṣḥaq, I am weary of my life **because of** the daughters of Ḥet: if Ya'aqov take a wife of the daughters of Ḥet, such as these, of the daughters of the land, what good shall my life be to me? (Gen. 27:46) **Sense: 4**
5	וַתַּעַל שִׁכְבַת הַטַּל וְהִנֵּה עַל־פְּנֵי הַמִּדְבָּר דַּק מְחֻסְפָּס דַּק כַּכְפֹר עַל־הָאָרֶץ׃	And when the layer of dew was gone up, behold, **upon the face of the wilderness** there lay a fine flaky substance, as fine as the hoar frost on the ground. (Ex. 16:14) **Sense: 2**
6	וְהַר סִינַי עָשַׁן כֻּלּוֹ מִפְּנֵי אֲשֶׁר יָרַד עָלָיו ה׳ בָּאֵשׁ וַיַּעַל עֲשָׁנוֹ כְּעֶשֶׁן הַכִּבְשָׁן וַיֶּחֱרַד כָּל־הָהָר מְאֹד׃	And mount Sinay smoked in every part, **because** the Lord descended upon it in fire: and the smoke of it ascended like the smoke of a furnace, and the whole mountain quaked greatly. (Ex. 19:18) **Sense: 4**

7	לֹא־יִהְיֶה־לְךָ֙ אֱלֹהִ֣ים אֲחֵרִ֔ים עַל־פָּנָ֑י׃	Thou shalt have no other gods beside Me [lit. "**upon My face**"]. (Ex. 20:3) **Sense: 3**
8	וַיְהִי֙ בְּרֶ֣דֶת מֹשֶׁ֔ה מֵהַ֖ר סִינַ֑י וּשְׁנֵ֨י לֻחֹ֜ת הָעֵדֻ֗ת בְּיַד־מֹשֶׁ֛ה בְּרִדְתּ֥וֹ מִן־הָהָ֖ר וּמֹשֶׁ֣ה לֹא־יָדַ֗ע כִּ֤י קָרַן֙ עוֹר֙ פָּנָ֔יו בְּדַבְּר֖וֹ אִתּֽוֹ׃	And it came to pass, when Moshe came down from mount Sinay with the two tablets of Testimony in Moshe's hand, when he came down from the mountain, that Moshe knew not that the skin of **his face** shone while he talked with him. (Ex. 34:29) **Sense: 1**
9	וְטָבַ֤ל הַכֹּהֵן֙ אֶצְבָּע֔וֹ מִן־הַדָּ֖ם וְהִזָּ֣ה שֶׁ֑בַע פְּעָמִ֗ים לִפְנֵ֤י ה֙׳ אֵ֚ת פְּנֵ֣י הַפָּרֹֽכֶת׃	and the priest shall dip his finger in some of the blood, and sprinkle it seven times before [lit: "**to the face of**"] the Lord, before [lit. "**the face of**"] the veil. (Lev. 4:17) **Senses: 2 and 3**
10	לֹא־תַעֲשׂ֥וּ עָ֙וֶל֙ בַּמִּשְׁפָּ֔ט לֹא־תִשָּׂ֥א פְנֵי־דָ֖ל וְלֹ֣א תֶהְדַּ֑ר פְּנֵ֣י גָד֔וֹל בְּצֶ֥דֶק תִּשְׁפֹּ֖ט עֲמִיתֶֽךָ׃	You shall do no unrighteousness in judgement: thou shalt not respect the person [lit. "**lift the face**"] of the poor, nor honour the person [lit. "**the face**"] of the mighty: but in righteousness shalt thou judge thy neighbour. (Lev. 19:15) **Sense: 5**
11	דַּבֵּ֣ר אֶל־אַהֲרֹ֔ן וְאָמַרְתָּ֖ אֵלָ֑יו בְּהַעֲלֹתְךָ֙ אֶת־הַנֵּרֹ֔ת אֶל־מוּל֙ פְּנֵ֣י הַמְּנוֹרָ֔ה יָאִ֖ירוּ שִׁבְעַ֥ת הַנֵּרֽוֹת׃	Speak to Aharon, and say to him, When thou lightest the lamps, the seven lamps shall give light towards the body [lit. "**the face**"] of the candlestick. (Num. 8:2) **Sense: 2**
12	וַיָּבֹ֤א מֹשֶׁה֙ וְאַהֲרֹ֔ן מִפְּנֵ֖י הַקָּהָ֑ל אֶל־פֶּ֣תַח אֹ֔הֶל מוֹעֵ֖ד וַיִּפְּל֥וּ עַל־פְּנֵיהֶ֖ם וַיֵּרָ֥א כְבוֹד־ה֛׳ אֲלֵיהֶֽם׃	And Moshe and Aharon went **from the presence** of the assembly to the door of the Tent of Meeting, and they fell **upon their faces**: and the glory of the Lord appeared to them. (Num. 20:6) **Sense: 3 and 1**
13	וַיָּ֥גָר מוֹאָ֖ב מִפְּנֵ֣י הָעָ֑ם מְאֹ֕ד כִּ֥י רַב־ה֖וּא וַיָּ֥קָץ מוֹאָ֖ב מִפְּנֵ֥י בְּנֵ֖י יִשְׂרָאֵֽל׃	And Mo'av was sore afraid of [lit. "**from the face of**", "before"] the people, because they were many: and Moav was seized with dread **because** of the children of Yiśra'el. (Num. 22:3) **Sense: 3 and 4**
14	גּ֣וֹי עַ֔ז פָּנִ֖ים אֲשֶׁ֣ר לֹא־יִשָּׂ֑א פָּנִים֙ לְזָקֵ֔ן וְנַ֖עַר לֹ֥א יָחֹֽן׃	A ruthless [lit. "**hard-faced**"] nation, that will show the old no regard [lit. "**will not lift** the old's **face**"] and the young no mercy. (Deut. 28:50) **Sense: 5**
15	וְלֹא־קָ֥ם נָבִ֛יא ע֖וֹד בְּיִשְׂרָאֵ֑ל כְּמֹשֶׁ֕ה אֲשֶׁ֥ר יְדָעוֹ֖ ה֥׳ פָּנִ֖ים אֶל־פָּנִֽים׃	And there arose not a prophet since in Yiśra'el like Moshe, whom the Lord knew **face to face**. (Deut. 34:10) **Sense: 1**

Table 1 - Excerpts from the Torah.

3.2. *A short comparison: "face" in Christian and Islamic sacred texts*

Following, we will present a very short series of excerpts from Christian and Islamic sacred texts. In order to further clarify and explain our approach, we deemed it appropriate to introduce some reflections about the meaning, nuances and functions of the term "face" within sacred texts belonging to Christianity and Islam, the two other Abrahamic religions. Thus, we thought that linking elements belonging to more than one single cultural-religious background (in this case they are somehow interrelated, but the methodology could be applied to any other religious-cultural traditions) could be a fruitful approach. Regarding Christian sources, we focused on the New Testament, while Islamic examples were taken from the Quran.[30] We will notice that the term "face" in these contexts can easily be categorised as expressing some of the five groups of senses we defined for Hebrew. Some interesting features seem to appear: for example, Sense 4 (the sense of causal relationship) is the only one that does not appear in this short selection of texts. The sense of this little showcase is for the moment just providing some material selected according to our methodology; a methodology that it will be possible to apply elsewhere for different goals.

3.2.1 *Excerpts from the New Testament*

In the following examples, analogies with Biblical prose can be spotted pretty clearly. All of the five groups of senses we defined for Hebrew are present, with the exception of Sense 4 (causal relationship). In some cases, using expressions similar to those of the Torah is probably not casual at all, but on the contrary has the precise goal of conveying a well-determined message: both the Apostles and the Jewish communities they initially preached to had to be very familiar with Scripture, either in Hebrew or in the Greek translation. In Example 4, for instance, it could be easily assumed that the promise of seeing "face to face" constitutes a clear reference to Deut. 34:10, with the implication, from a Christian point of view, that the Torah's statement "And there arose not a prophet since in Yisra'el like Moshe, whom the Lord knew face to face" has been transcended with the advent of the new religion: not only, but seeing "face to face" would not be a faculty reserved to a special individual, such as the greatest prophets; on the contrary, all those who believe in Jesus' message will enjoy this blessing.

1	οἵτινες δίκην τίσουσιν ὄλεθρον αἰώνιον ἀπὸ **προσώπου** τοῦ κυρίου καὶ ἀπὸ τῆς δόξης τῆς ἰσχύος αὐτοῦ,	These will suffer the punishment of eternal destruction, separated from the presence [lit. "**the face**"] of the Lord and from the glory of his might (2 Thess. 1:9) **Sense: 3**
2	καὶ ἀποστέλλουσιν αὐτῷ τοὺς μαθητὰς αὐτῶν μετὰ τῶν Ἡρῳδιανῶν λέγοντες, Διδάσκαλε, οἴδαμεν ὅτι ἀληθὴς εἶ καὶ τὴν ὁδὸν τοῦ θεοῦ ἐν ἀληθείᾳ διδάσκεις, καὶ οὐ μέλει σοι περὶ οὐδενός, οὐ γὰρ βλέπεις εἰς **πρόσωπον** ἀνθρώπων.	So they sent their disciples to him, along with the Herodians, saying, "Teacher, we know that you are sincere, and teach the way of God in accordance with truth, and show deference to no one; for you do not regard people with partiality [lit. "you don't look at the **face** of people"]. (Mt. 22:16) **Sense: 5**

[30] In excerpts from the New Testament, the English translation we used is the New Revised Standard Edition (<https://www.biblestudytools.com/nrs/>); regarding the Quran, we used the Clear Quran translation by dr. Mustapha Khattab (<https://theclearquran.org/>).

3	ἐποίησέν τε ἐξ ἑνὸς πᾶν ἔθνος ἀνθρώπων κατοικεῖν ἐπὶ παντὸς **προσώπου τῆς γῆς**, ὁρίσας προστεταγμένους καιροὺς καὶ τὰς ὁροθεσίας τῆς κατοικίας αὐτῶν,	**From one ancestor he made all nations to inhabit the whole earth [lit. "the whole face of earth"], and he allotted the times of their existence and the boundaries of the places where they would live, (Acts 17:26)** **Sense: 2**
4	βλέπομεν γὰρ ἄρτι δι› ἐσόπτρου ἐν αἰνίγματι, τότε δὲ **πρόσωπον πρὸς πρόσωπον**: ἄρτι γινώσκω ἐκ μέρους, τότε δὲ ἐπιγνώσομαι καθὼς καὶ ἐπεγνώσθην.	**For now we see in a mirror, dimly, but then we will see face to face. Now I know only in part; then I will know fully, even as I have been fully known.** **(1 Cor. 13:12)** **Sense: 1**

Table 2 - Excerpts from the New Testament.

3.2.2 *Excerpts from the Quran*

In the Quran some analogies with the Hebrew Bible can be spotted, but they follow a somewhat different pattern. Differently from the New Testament, the Quran is not considered as a direct derivation or filiation of Hebrew Scriptures; on the other hand, being written in a Semitic language related to Hebrew, some linguistic similarities appear pretty frequently. For example, in Excerpt 2 the bright and gloomy faces (literally "black" and "white") indicate the joy or distress of souls on the Day of Judgement. Another interesting example is Example 4, the last one: Sense 2 (the spatial sense indicating a surface or exposed part of something, such as "the face of earth") is somehow transferred to the dimension of time.

1	ٱذْهَبُوا۟ بِقَمِيصِى هَٰذَا فَأَلْقُوهُ عَلَىٰ **وَجْهِ** أَبِى يَأْتِ بَصِيرًا ۖ وَأْتُونِى بِأَهْلِكُمْ أَجْمَعِينَ	**Go with this shirt of mine and cast it over my father's face, and he will regain his sight. Then come back to me with your whole family." (Q 12:93)** **Sense: 1**
2	يَوْمَ تَبْيَضُّ **وُجُوهٌ** وَتَسْوَدُّ **وُجُوهٌ** ۚ فَأَمَّا **ٱلَّذِينَ ٱسْوَدَّتْ وُجُوهُهُمْ** أَكَفَرْتُم بَعْدَ إِيمَٰنِكُمْ فَذُوقُوا۟ ٱلْعَذَابَ بِمَا كُنتُمْ تَكْفُرُونَ	**On that Day some faces will be bright [lit. "white"] while others gloomy [lit. "black"]. To the gloomy-faced it will be said, "Did you disbelieve after having believed? So taste the punishment for your disbelief."** **(Q 3:106)** **Sense: 1 and 5**
3	وَمَنْ أَحْسَنُ دِينًا مِّمَّنْ **أَسْلَمَ وَجْهَهُۥ** لِلَّهِ وَهُوَ مُحْسِنٌ وَٱتَّبَعَ مِلَّةَ إِبْرَٰهِيمَ حَنِيفًا ۗ وَٱتَّخَذَ ٱللَّهُ إِبْرَٰهِيمَ خَلِيلًا	**And who is better in faith than those who ˹fully˺ submit themselves [lit. "submitted his face"] to Allah, do good, and follow the Way of Abraham, the upright? Allah chose Abraham as a close friend.** **(Q 4:125)** **Sense: 3**
4	وَقَالَت طَّآئِفَةٌ مِّنْ أَهْلِ ٱلْكِتَٰبِ ءَامِنُوا۟ بِٱلَّذِىٓ أُنزِلَ عَلَى ٱلَّذِينَ ءَامَنُوا۟ **وَجْهَ ٱلنَّهَارِ** وَٱكْفُرُوٓا۟ ءَاخِرَهُۥ لَعَلَّهُمْ يَرْجِعُونَ	**A group among the People of the Book said ˹to one another˺, "Believe in what has been revealed to the believers in the morning [lit. "in the face of the day"] and reject it in the evening, so they may abandon their faith.** **(Q 3:72)** **Sense: 2 (in time, not in space)**

Table 3 - Excerpts from the Quran.

4. *The Computational representation*

Flavia Sciolette

In this section, we discuss the development of the resource, which represents the computational formalisation of the case study mentioned above. This resource is based on the model and methodology described in the second section; consequently, it includes the formalisation of the following systems:

- Linguistic System (domain of language): this system is modelled, as stated below, according to the OntoLex-Lemon model and its modules (core-OntoLex, Variations and Translations, Syntax and Semantics); lexical data is entered with the LexO editor (an overview of the used tools is included in Section 5);
- Conceptual System (domain of concepts): in this work, we use an ontology based on the BFO vocabulary, to be interoperable with other major religious resources. In addition to the categories of the BFO vocabulary, the ontology includes some custom classes related to the Hebrew domain;
- Graphic System (domain of sign): it represents the term in context. The strings form the attestation of the written representations, connected with the term.

The very first step for the construction of the resource is to collect and analyse contexts, as the first base of vital data related to the meanings of terms. The resource comprises two components: 1) the definition of the terms from a computational point of view, with the attestations and the relations among the different languages involved in the case study; 2) the ontological modelling of concepts, of which the terms represent the lexicalization.

The resource has to be compliant with the principles of LOD (Linked Open Data); the LExicon Model for ONtologies (OntoLex-Lemon), developed by the W3C workgroup Lexica,[31] is the standard *de facto* for the representation of computational lexicons. It allows to combine conceptual and linguistic information. Consequently, the model is a fundamental reference for the Linguistic Linked Open Data community.[32] OntoLex-Lemon provides the labels, in an "agnostic" way towards the linguistic theories, to model and formalise linguistic data. The description of the language is demanded to specific vocabularies; in this work, we consider the labels of the "core" module of OntoLex-Lemon, concerning the principal structures for entries and senses; the Syntax and Semantics module (synsem), related to an association between syntactic frame and ontological mapping; the Variations and Translations module (vartrans), for lexical, terminological, and translation relations amongst entries and senses.

In particular, the OntoLex module formalises the structure of linguistic data, such as lexical entries, their grammarly possible forms (including the lemma, as "canonical form"), their senses, and the relative conceptual references.

OntoLex-Lemon does not provide labels for morphological traits; according to the linguistically "agnostic" approach mentioned above and to LOD philosophy, morphological features are described by LexInfo,[33] an ontology created for the linguistic description of data (such as part of speech, number, gender, and so on).

This section is structured as follows: the first paragraph is dedicated to the formalisation of the Hebrew term, *panim* (פָּנִים) "face" with its senses, the ontological references and relations amongst forms; the second paragraph, reflecting the structure of the previous section, describes a formalisation with relations amongst terms of other languages, in a comparative perspective.

[31] J.P. McCrae, J. Bosque-Gil, J. Gracia, P. Buitelaar, P. Cimiano, *The OntoLex-Lemon model: development and applications*, in *Proceedings of eLex 2017 conference*, 2017, pp. 19-21.

[32] P. Cimiano, C. Chiarcos, J.P. McCrae, J. Gracia, *Linguistic linked open data cloud*, in P. Cimiano (ed.), *Linguistic Linked Data*, Springer, Cham, 2020, pp. 29-41.

[33] <https://lexinfo.net/>.

4.1 *Many (digital!) "faces" of a word: the case of panim*.

In this paragraph, we expose a practical example of formalisation of a term, according to the model, with the entry *panim* (פָּנִים) "face". Here we considered the formalisation of the sense for "face" as a single term; we will deepen the multiwords in the next paragraph.

Figure 1 describes *panim* as a graphic representation of the relations among the different elements that compose the entry.

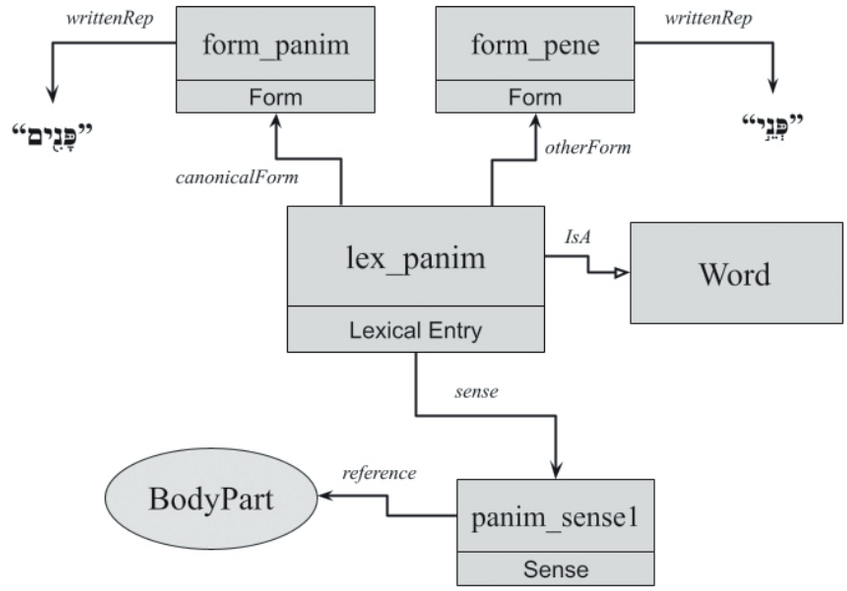

Fig. 1 - The entry *panim* modelled in OntoLex-Lemon (for ease of viewing just one sense is represented).

In this example, thanks to the labels of the OntoLex-Lemon model, we formalise *panim* as an object of type "Lexical Entry" (similarly to an entry in a dictionary) and, in particular, of type "Word" (separate from the class of "Affix", as part of a Word, and the class of "Multiword Expression"), with a sense (the semantic meaning of an entry), a certain number of forms – one of which is the lemma – with their written representations, and an ontological reference for the sense.

This simple representation can be rendered in OntoLex-Lemon with Turtle notation,[34] which is characterised by its simplicity and conciseness.

```
relex:lex_panim a ontolex:LexicalEntry,
ontolex:Word;
ontolex:canonicalForm relex:form_panim_
plural;
lexinfo:partOfSpeech lexinfo:noun;
ontolex:lexicalForm relex:form_panim_con-
struct;
ontolex:sense relex:panim_sense1.

relex:panim_sense1 a ontolex:LexicalSense;
skos:definition "The front part of a per-
son's head from the forehead to the chin,
or the corresponding part in an animal."@
en;
ontolex:reference <#bodyPart>.
```

[34] https://www.w3.org/TR/turtle/.

```
relex:form_panim_plural a ontolex:Form;
lexinfo:number lexinfo:plural;
ontolex:writtenRep "פָּנִים"@he; lexinfo:
transliteration "panim"@he.

relex:form_pene_definite a ontolex:Form;
lexinfo:definitess lexinfo:definite;
ontolex:writtenRep "פְּנֵי"@he; lexinfo:
transliteration "pene"@he.
```

"Relex" is an example of custom "namespace" for our resource, that defines the entries of our vocabularies; every data – lexical entries, senses, forms, and so on – is described as an object, with a proper URI (for example, *lex_panim* identifies the entry for "face" in Hebrew in our resource). The vocabulary "ontolex" defines the features of the entries according to the OntoLex-Lemon model; for example, the first assertion

```
relex:lex_panim a ontolex:LexicalEntry,
ontolex:Word;
```

means that the entry "lex_panim" of the vocabulary "relex" is a Lexical Entry according to the module "ontolex", belonging to the class "Word". In the same logic, we can represent all the linguistic information and link them to concepts and attestations. In the next figure, we describe these relations amongst data; the attestations are connected with the different forms associated with an entry. For brevity, we illustrate forms connected with attestations and senses, without mentioning, in every relation, the Lexical Entry of reference. We also created an ontology to gather together the lexicalisations belonging to the Hebrew cultural domain (represented with "HebrewThing" as the top-class and other subclasses describing the domain); this ontology is connected with the vocabulary of BFO (Basic Formal Ontology) for traits of common sense knowledge. We have chosen to employ this high-level ontology for different reasons: i) it is used in similar projects, such as Sefaria.org, and the use of a resource that has been already extensively employed in other projects is, in terms of LOD philosophy, a necessary good practice, especially in a perspective of interoperability; 2) it presents a solid and rigorous model, characterised by simplicity and clarity. The ontology presents a top-level class "Entity", representing anything that exists, with two subclasses: Continuant and Occurrent. Continuant represents "an entity that persists, endures, or continues to exist through time while maintaining its identity",[35] as an object; Occurrent, on the other hand, denotes "an entity that unfolds itself in time or it is the instantaneous boundary of such an entity (for example a beginning or an ending)".[36] In our examples we consider only the subclasses of Continuant: Independent Continuant and Dependent Continuant, in turn divided into specific subclasses "generically dependent continuant" and "specifically continuant". An independent continuant can contain the material entities; dependent continuant includes the qualities of an entity. For the sake of simplicity, figures 2 and 4 do not show the entire taxonomy of BFO, but only the high level concepts (e.g. "Continuant"). Figure 2 illustrates the case-study with the five groups of senses.

[35] M. Almeida, J. Bona, M. Brochhausen, W. Ceusters, M. Courtot, R. Dipert *et alii*, *Basic Formal Ontology 2.0*, https://github.com/BFO-ontology/BFO/blob/master/docs/bfo2-reference/BFO2-Reference.pdf, p. 21.

[36] *Ivi*, p. 66.

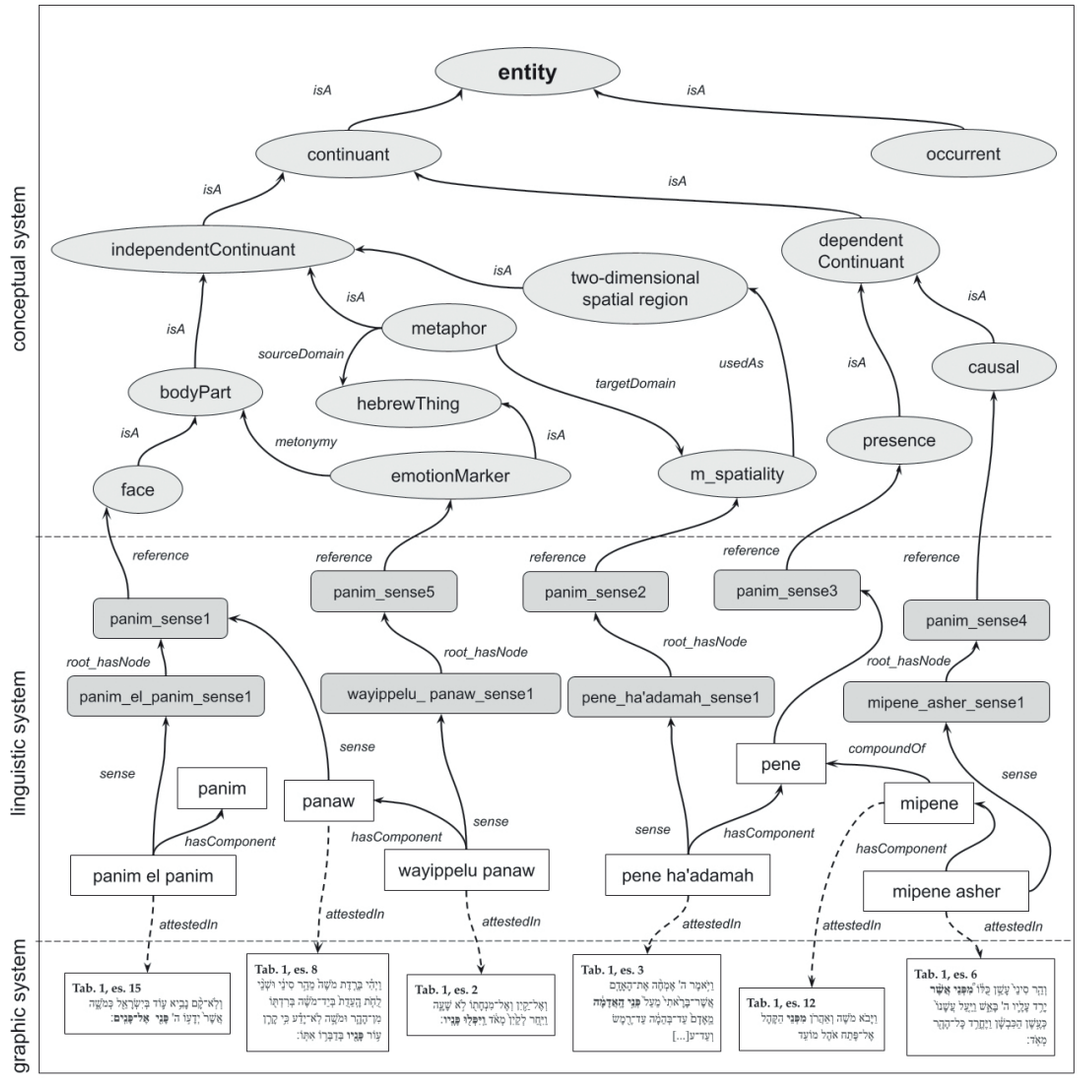

Fig. 2 - Formalisation in the three different systems (graphic, linguistic, and conceptual) of the Hebrew term "panim" and the five relative groups of senses.

The Graphic System contains the attestations; each form is connected with the attestation that represents the context of the word. For example, the attestation Ex. 34:29 (Table 1, n. 8) is linked with the form *panaw*, belonging to the entry *panim*.

In the Linguistic System, the form *panaw* presents two relations for different kinds of information; concerning the form in the context of Ex. 34:29, the word is connected with "Panim_sense1", with the definition of "Part of the body" (cfr. supra). However, *panaw* can appear in multiword expressions with different meanings, as in the case of *wayippelu panaw*. *Panaw* is a component of a multiword, and it is linked with the multiword "wayippelu panaw" through the relation "hasComponent", described in OntoLex-Lemon. The "Decomp module" of OntoLex-Lemon includes labels to formalise the traits of phrases; furthermore, the synsem module permits to describe the syntactical constraints and the semantic range for a certain expression. After the modelisation of the forms of a multiword, we can describe the semantics of expressions; in the case of *wayippelu panaw*, we indicate a specific sense for the entire string,

with the URI "wayippelu_ panaw_sense1". This object is linked to "panim_sense5": this sense is a "node", i.e. a part of the meaning of the multiword (see Section 4.2 for more details about multiwords and their components).

Both senses are connected with a class, in the Conceptual System, with the "reference" relation to two custom classes that point to high-level concepts of BFO; for the first sense of *panim*, we use the reference "bodyPart". Body parts are considered as concrete entities and thus, in the BFO vocabulary, refer to the class "Continuant". The other sense has as reference to the class "emotionMarker" (cfr. Sense 5 in Section 3.1), because the term *panim* can be used as a metonymy for physical manifestations of emotion. A similar process is used to define the third and fourth senses. The forms present specific morphological traits and the relation "compoundOf" models the compositionality of the expression, according to the pattern particle "*mi*" and the form "*pene*".

Regarding the second sense, we can model a peculiar use of the sense as a metaphor; in this case, we indicate a custom class "metaphor", that includes all the linguistic metaphors belonging to the Hebrew domain ("HebrewThing"), of which "penè haadama" is an example. In this case, the class "metaphor" is linked with other classes with two specific relations: *sourceDomain* links the class to the class that constitutes the prototypical concept for the term *panim* (the body part), while *targetDomain* points to the class of reference for the translated concept (the spatiality). This last is linked with a class of BFO that defines the spatial region in two dimensions (surfaces).

This formalisation allows us to describe different behaviours of a word in constructions, with different uses; in the next paragraph, we will show how this model can be used also to describe relations amongst terms in other languages.

A comparison of this resource obtained with the proposed approach to analogous resources appears particularly difficult. As a matter of fact, this resource was developed as a prototype and mainly conceived to test the methodology. At the same time, the purposes and nature of our resource differ considerably from those of the resources mentioned in the first section of the paper. However, a comparison with PESHAT, one of the considered dictionaries available online, might provide additional elements to better assess the characteristics of our approach.

PESHAT stands out among the dictionaries cited in Section 3.1, because it appears as a strongly structured resource that can be particularly useful to scholars. Though not yet compliant with the LOD paradigm (since data stored in a database cannot be formally linked to data belonging to other resources) its structure distinguishes definitions and attestations, associating a sense to all the citations of texts belonging to the considered domain. In fact, PESHAT is a specialised vocabulary, focused on premodern Hebrew philosophical and scientific terminology. The following figure shows the entry *panim* in PESHAT.

Fig. 3 - The entry *panim* in PESHAT.

The entry is described as a "lemma". The first line represents the root of the word, followed by the morphological traits and the external documentation (in this case, Jastrow and Historical Dictionary Project, cfr. Section 3.1). For each sense, all quotations of the word in context are mentioned, with metadata of texts. For some senses, equivalent translations in different languages (especially European) are available.

Although the methodology is not bound to the use of a specific application, in the next section, we present the collaborative tool for the development of this kind of resources.

4.2 *Comparative considerations in a computational resource*

In this paragraph, we describe a possible use of our methodology in a comparative perspective, following the examples mentioned in 3.2 and subsections. From a conceptual point of view, we decided to create one single ontology, but divided into three different parts, including the already mentioned "HebrewThing" and other two – ChristianThing and IslamicThing – as macro classes. These three branches of the ontology are not meant to replace existing ontological resources for these domains, but only to give

a basic conceptual infrastructure for the formalisation of our case study.

The concept of "face" is represented in all the three parts of the ontology. The Hebrew word *panim* and the Arabic word *wajh* both appear in the corpus of reference as part of different phrases with a metaphorical value. The conceptual system permits a representation of similarities among different domains (and lexicalisation of these concepts), as a reflection of the mental encyclopaedia of the members of the reference communities, while the vocabulary of vartrans module allows to model the relations between entries and senses in different languages and consequently in different domains. To model the case study considered with the equivalent terms in other languages, in particular we modelled three macro-categories of relations: translation relations (among terms chosen in different linguistic domains); relations of senses among terms belonging to the same linguistic domain; relations among terms of different linguistic domains to represent differences of meaning from religious or cultural point of view. In Figure 4, we show a practical example of this approach; for brevity, we consider only the second sense of "face" and its role as a component for specific multiwords.

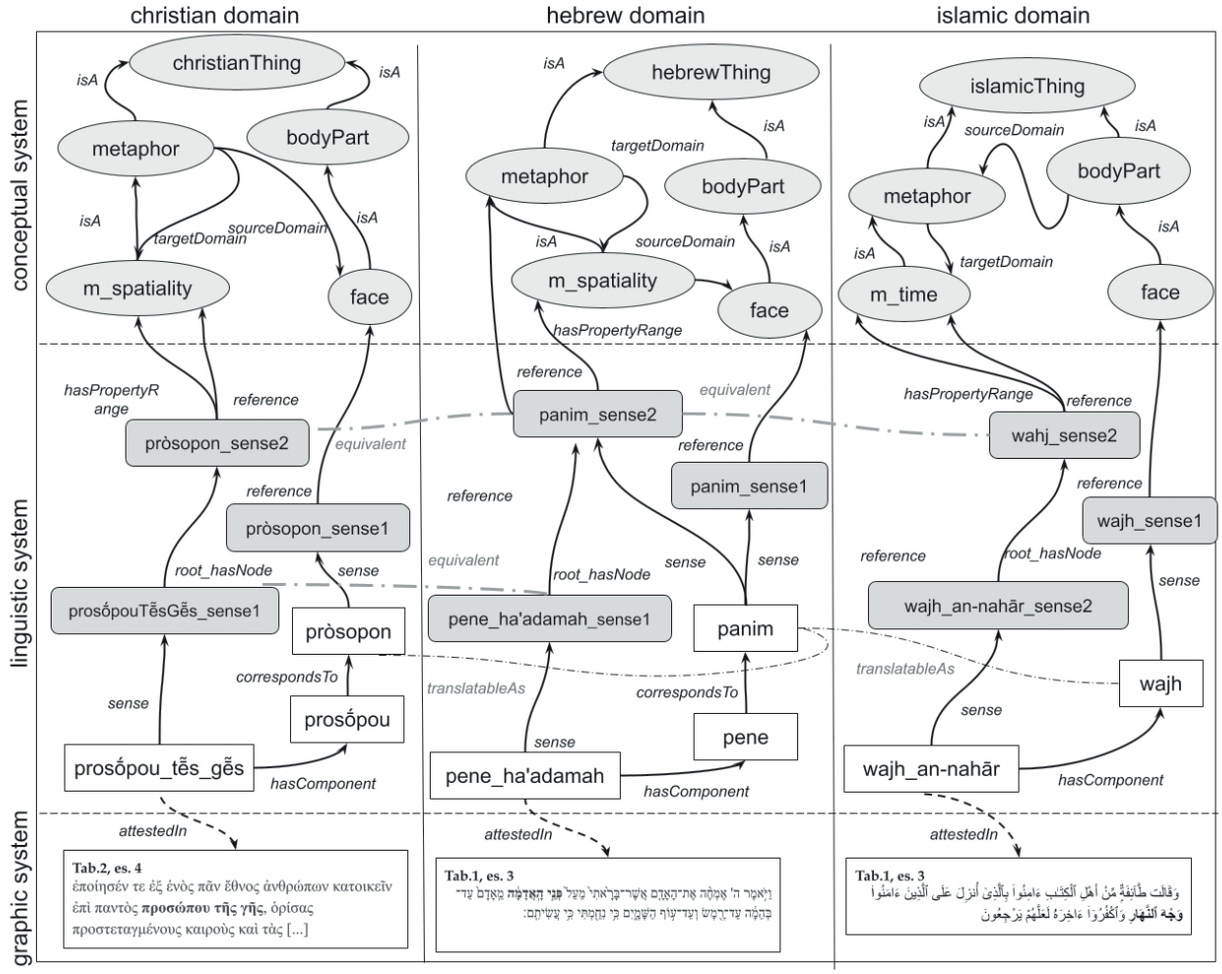

Fig. 4 - An example of formalisation in the three different systems (graphic, linguistic, and conceptual) of the term "face" in a comparative perspective.

In the "Hebrew domain" portion of the figure, at the bottom part of the "linguistic system" level, we show the expression *pene ha'adamah* (פְּנֵי הָאֲדָמָה) with the literal meaning of "the face of the earth" and consequently used to mean the whole world. In the OntoLex-Lemon model, *pene ha'adamah* is a Lexical Entry, belonging to the class of "Multiword Expression".

This expression is formed by two elements: the form "pene", that represent the construct state of the lemma *panim* (פָּנִים); the form "ha'adamah" from the feminine lemma *adamah* (אֲדָמָה), standing for "earth". *Adamah* and *panim* are two distinct lexical entries that we can codify in the lexical part of the resource; for each entry, we can associate different forms and their morphological traits. So, for *panim*, the described form is *pene*, with the trait of "genitive"; in the same way, for *adamah*, the form is "ha'adamah" with the trait "definite", according to LexInfo vocabulary. Furthermore, we can formalise the components of the multiword and their order in this linguistic construction; as other entries, we associate a reference to Lexical Entry, and create a relation "correspondsTo" between the component "pene" to entry "panim". The same relation binds the component "ha'adamah" with entry "adamah". The same formalisation is used for other multiwords in the selected corpus with the same meaning as in English "face of the earth", and as Greek "prosṓpou tẽs gẽs". All of these expressions are calques formed from the original expression in Hebrew. For this reason, we can formalise a translation relation of type "direct equivalent".

We have to formalise the use of "face" as "root" of a multiword with a metaphorical meaning. The starting point is the formalisation of the lexical entry *panim*, two senses of which are formalised: the first referring to a part of the human body; the second being a metaphorical use of the term, as figurative meaning for expressions in the semantic field of spatiality.

For the first sense, *panim*, gr. *pròsopon*, and ar. *wajh* (وَجْه) are associated to each other with a "TraslatableAs" relation. For the second sense, we can observe that *wajh* has the same syntactic and semantic behaviour of *panim*, as first element of a multiword with a metaphorical meaning; "panim_sense2" and "wajh_sense2" are associated with one sense of the corresponding multiword through the relation "root_hasNode". For this sense, *waih* and *panim* are associated to each other with a translation relation of type "direct equivalent". However, in Arabic, the term *wajh* can be used for metaphors in the semantic field of time, in expressions like *wajh an-nahār* (وَجْه ٱلنَّهَارِ), lit. "the face of the day", that is "morning". To the lexical entry *wajh an-nahār* we associate an appropriate reference ("m_time", a subclass of "metaphor", defined to represent the concept of metaphor about time), as we have previously done for *pene ha'adamah* (as in Figure 4, "m_spatiality", a subclass created as the concept of metaphor about spatiality). Instead, to panim_sense2 and wajh_sense2, we can associate the reference of "metaphor", but – to model the different uses – we need to formalise a "domain constraint". This is possible through the OntoLex-Lemon relation "propertyRange", which allows us to describe limitations in the use of a term, from a semantic point of view. In the conceptual system we can formalise an additional relation between the class "BodyPart". This formalisation permits to describe finely the syntactic and semantic behaviour of multiwords.

5. *Tools to support in the construction of resources*

Emiliano Giovannetti

As shown in the previous section, the process that scholars need to carry out to create the computational representation of their case studies can be particularly complex and time consuming if not adequately supported. At the time of writing of this paper, an editor called LexO, developed at ILC-CNR, is already available and can provide an easy way to create lexical or terminological resources.[37] We developed the multilingual lexicon to represent the case study introduced in Section 3 (and formalised in Section 4)

[37] A. Bellandi, *LexO: An Open-source System for Managing OntoLex-Lemon Resources*, «Language Resources & Evaluation» 55 (2021), pp. 1093-1126, https://doi.org/10.1007/s10579-021-09546-4.

using LexO, which can be accessed and browsed using the link indicated in the footnote.[38] LexO can be freely downloaded and installed using the instructions provided at GitHub.[39] In Figure 5 LexO's interface is shown, where details of the word "panim" (lemma, forms, and senses) appear inside boxes of different colours placed in the right side of the screen.

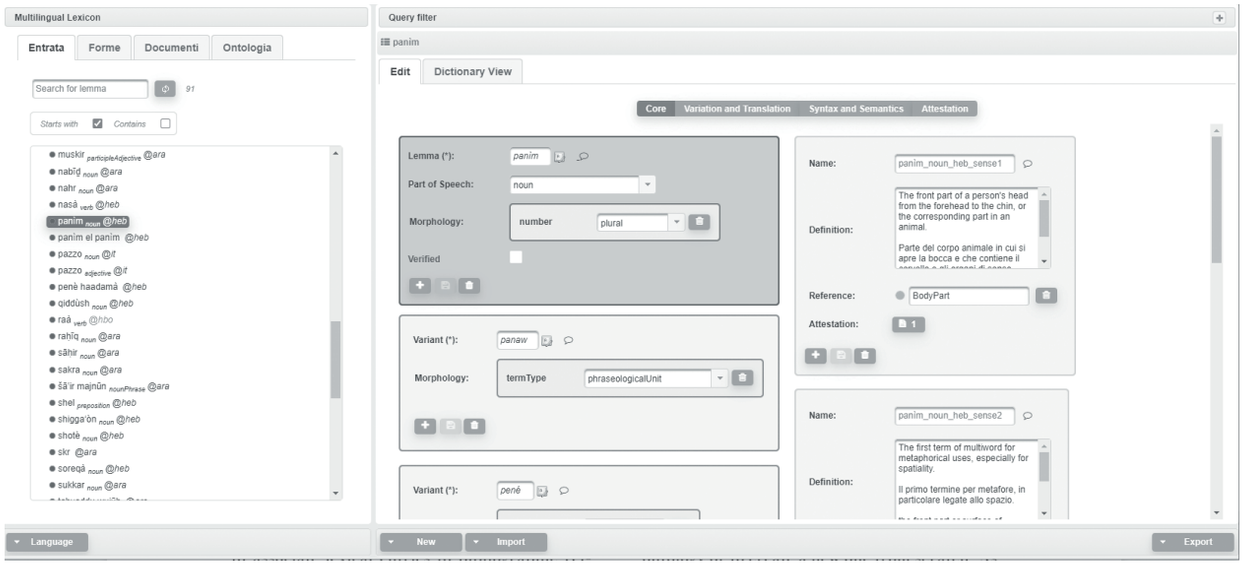

Fig. 5 - An example of lexical entry created in LexO.

Although LexO offers some basic features to associate lexical entries to bibliographic references, it does not allow uploading the texts constituting the corpus of reference. However, LexO permits to model the attestation (concordances and relative documentation) of forms of an entry through a dedicated "Dictionary View" section.[40] Similarly, it allows to browse the taxonomy of concepts constituting an imported ontology but it lacks the functionalities to edit the ontology or to create a new one from scratch. As a matter of fact, to populate the conceptual part of the resource we used Protégé,[41] a free tool developed by Stanford University for the creation and editing of OWL ontologies.[42] In particular, LexO allows to create all the lexical entries (in each of the desired languages) needed for the analysed case study along with all the lexical forms and senses. Then, each sense can be linked to an ontological reference and one or more attesta-

[38] <https://klab.ilc.cnr.it/LexO-PRIN> (username: "Prin2017", password: "diversita1").

[39] <https://github.com/andreabellandi/LexO-lite>

[40] The functionalities of the attestation tab in LexO do not consider the module "Attestation", the development of which is still in progress by the Lexica Group for OntoLex-Lemon. See C. Chiarcos, M. Ionov, J. de Does, K. Depuydt, A.F. Khan, S. Stolk, T. Declerck, J.P. McCrae, *Modelling Frequency and Attestations for OntoLex-Lemon*', in *Proceedings of the Globalex Workshop on Linked Lexicography, Language Resources and Evaluation Conference (LREC 2020)*, 2020, pp. 1-9.

[41] M.A. Musen, *The Protégé project: A look back and a look forward*, «AI Matters. Association of Computing Machinery Specific Interest Group in Artificial Intelligence» 1, 4, (June 2015), doi: 10.1145/2557001.25757003.

[42] <https://protege.stanford.edu/>.

tions, representing the lexical contexts. In this way, the lexicon can be connected both to the textual corpus and to the ontology, representing the formal and explicit representation of the knowledge the texts convey.

To provide an easy and integrated way to build up all the resources needed to create the computational representation of a case study, we are working at the development of a new tool, grounded on LexO's technology, which will allow scholars to set up their own personalised digital working environments to gather and organise all the textual, lexical, and ontological resources they need to carry out their research.

In particular, the tool will expose a series of functions for the use of scholars who will be enabled to:

1) create their own corpus by uploading and organising the documents they need to work with;
2) enrich the texts belonging to documents with the desired annotations on the basis of all the levels provided by the underlying model (briefly described in Section 2), such as, for example, morphological annotations (e.g. lemma, Part-of-Speech, morphological traits), semantic annotations (e.g. to attribute a specific sense to a polysemous word), and so on.
3) import, or create "ex novo", lexical and terminological resources, to formally structure, from the one hand, the lexical entries pertaining to one or more languages of interest and, on the other hand, the terms describing a specific domain (similarly to what LexO already allows to do);
4) import, or create "ex novo", ontologies representing the conceptualization of one or more particular domains of interest;
5) define links between elements belonging to the three classes of managed resources (i.e. texts, lexicons, ontologies), for example to link a linguistic annotation of a text to an entry of a lexicon, or to link a lexical sense of a lexicon to a concept of an ontology.

The interface of the new tool will be inspired to the design of modern web-based Integrated Software Environments (IDE), applications that are commonly used by software developers to organise a number of heterogeneous resources they need to accomplish their work, also by creating their own customised layouts and workspaces.

Another important feature of the new system will be related to collaborative work. The idea is to take full advantage of all the available technologies to enable a team of scholars (and not just one) to work in a collaborative way to the same resources. This feature has been already experimented successfully both with LexO and with Traduco[43] (Giovannetti *et al.*, 2017), the computer-assisted translation tool developed by ILC-CNR and used in the Babylonian Talmud Translation Project.[44]

6. *Discussion and future works*

Davide Saponaro

The work described in the present article is aimed at developing a methodology for the construction of a computational resource integrating texts, terms, and concepts concerning the religious domain. The resource is conceived to provide scholars with all the information they need to formulate their research hypotheses based on a specific textual corpus with a set of linguistic and conceptual entities of interest found therein and made explicit in the resource. The benefits of a common and shared digital environment where textual passages, terms, and concepts coexist and are formally linked to each other are particularly evident in a comparative perspective, to highlight similarities and differences that can be found in a textual corpus. The approach was also explained by presenting a case study centred on the theme of "face" and

[43] E. Giovannetti, D. Albanesi, A. Bellandi, G. Benotto, *Traduco: A collaborative web-based CAT environment for the interpretation and translation of texts*, «Digital Scholarship in the Humanities» 32, suppl_1 (2017), pp. i47-i62, doi: 10.1093/llc/fqw054.

[44] <https://talmud.it>.

describing its computational representation.

A digital resource built according to the suggested criteria, integrating portions of texts (referring to specific sources), codified terms (describing their linguistic characteristics in detail), and finally ontological entities formally representing the concepts evoked in the text, would be suitable for several purposes.

In first place, it would represent, in itself, an integrated and unambiguous formalisation of a set of heterogeneous data, that a scholar would be able to share as his or her interpretation of a specific case study, also using modalities complying with Linked Data standards. Such a resource could in fact be used to enrich pre-existing linguistic or conceptual resources (lexica, thesauri, terminologies, taxonomies, ontologies, etc.) as well as to assist tools designed for automatic language management or for an "intelligent" access to text. In fact, every part of the resource (term, concept, relation, etc.) will be able to act as a linguistic-semantic access key to texts.

Looking towards the future, we intend, in the first place, to go on in the development of the general model for a text's computational representation, analysing and detailing its various systems and the dimensions of its articulation. At the same time, the tools (including LexO) aimed at supporting scholars in the process of computational resource building will be further developed.

Regarding the relation with existing collections, the methodology (and consequently the nature of resources built in its context) will be enriched to also include data belonging to resources such as PESHAT. For example, senses of words and the relative definitions, as shown in Section 3.1, are already available and may be included in resources built following our approach.

We also plan to deepen the comparative aspects of our methodology in order to explore the theme of religious diversity, here briefly introduced in the context of the case study of "face". This will be done considering new semantic fields with the goal of formalising and highlighting similarities and differences, on a terminological and conceptual level, among the three Abrahamic religions. Starting from the Hebrew domain, a link between the defined conceptual part and the already mentioned hierarchy of Judaism-related subjects developed by Sefaria will also be experimented.

7. *Acknowledgment*

This work was carried out in the context of PRIN 2017 "Representing Religious Diversity in Europe: Past and Present Features" project.

Davide Saponaro
davide.saponaro@ilc.cnr.it

Emiliano Giovannetti
emiliano.giovannetti@ilc.cnr.it

Flavia Sciolette
flavia.sciolette@ilc.cnr.it

SUMMARY

This paper introduces a methodology for the creation of a digital representation of a religious case study integrating textual, linguistic, and conceptual entities. The description of a holistic model of text, to be used as the backbone of the computational resource that needs to be built, is provided. The proposed case study, focusing on the semantic field of "face" in Jewish religion, is first introduced from the point of view of the scholar and then translated, with the support of digital tools, into the relative computational representation.

KEYWORDS: Religious Studies; Hebrew Terminology; Ontology.