Resilient Septuagint Between Borges and Asimov: A State-of-the-Art Case of uBIQUity

"Resilient Septuagint"[1] between Borges and Asimov: A State-of-the-Art Case of uBIQUity

1. Introduction

Over the past two decades, there has been significant reflection within the two main centers of what is referred to in Italian as *informatica umanistica*[2] — the Association for Computers and the Humanities (ACH) and the Association for Literary and Linguistic Computing (ALLC, now the European Association for Digital Humanities EADH) — regarding both the nature of the discipline and the criteria by which that nature is to be assessed and defined. More precisely, we appear to be witnessing a shift from an informational concern to an epistemological need. In the beginning (since approximately the 1960s) there was a palpable desire to make linguists—and humanists in general–aware of the advances and frontiers of computation. This led to the founding of "Computers and the Humanities" (later, until its closure, "Language Resources and Evaluation"), i.e. the journal that, in 1978, gave rise to the ACH.[3] This periodical remained the guiding spirit that animated the joint conferences (beginning in 1989) of the ACH and ALLC, which the journal "Literary and Linguistic Computing" (now "Digital Scholarship in the Humanities") reported on. By contrast, starting in 2006, this initial need has gradually been overtaken by what can only be described as an epistemological turn.[4] This shift can be traced to a specific moment: 2004, the year Blackwell published its *Companions* to the *Digital Humanities* series.[5]

[1] "Resilient Septuagint. An Initial Exploration of the Semantics of Killing and Healing in the Septuagint and its Reception in Patristic and Late Antique Sources (3rd cent. BCE-5th cent.CE)" is a PRIN 2022 grant, a project funded by the European Union – Next Generation EU Mission 4 Component 1 CUP: J53D23013060001 Prot. 20229E83B3. The PRIN Resilient Septuagint research team articulates three research units: University of Bologna (PI research unit, whose members are Davide Dainese as PI, Laura Bigoni and Gianluca Scatigno as researchers, Giorgia Sampò and Marco Zanella as consultants), University of Catania (deputy-PI research unit, whose members are Arianna Rotondo as chief, and Isabella PIgnocco as a researcher), University of Bari (lead by Laura Carnevale and whose other member is Luca Arcari of the University of Naples Federico II). Paragraphs 1, 2, and 3 were written by Davide Dainese; sections 4.1 and 4.2 by Laura Bigoni; section 4.3 by Marco Zanella. The concluding section (4.4) is the result of an equal contribution by all authors.

[2] On this denomination see F. Ciotti, "Introduzione. La galassia delle Digital Humanities," in F. Ciotti (ed.), *Digital Humanities. Metodi, strumenti, saperi* (Roma: Carocci, 2023) 19-23 and D. Buzzetti, "Alle origini dell'informatica umanistica. Humanities Computing e/o Digital Humanities?"*, Rendiconti Lincei – Scienze Morali Storiche e Filologiche* 9-30 (2019) 71–103; with this I follow the definition of the Italian tradition, see F. Tomasi in T. Orlandi/F. Tomasi, "Una storia dell'informatica umanistica in Italia," in F. Ciotti (ed.), Digital Humanities. *Metodi, strumenti, saperi* (Rome: Carocci, 2023) 35-47, on p. 46 which tends to distinguish humanities computing (that is, Father Busa's approach and the inescapability of computation discerned in recent years) from DH or digital culture which, in this last quarter century, has cultivated interest in forms of both digital representation/communication of research results and and computational nature of the processes that produced them (e.g. through the semantic web and linked data).

[3] It aimed to fill the gap of an "international newsletter serving the community of humanistic scholars who employ computers" and to provide "support [...] by concentrating on news" (see "Prospect," Computers and the Humanities 1 [1966] 1-2, on p. 1). This paradigm is the framework in which the Text Encoding Initiative (TEI) was established, see D.T. Barnard, C.H. Fraser, G.M. Logan, "Generalized Markup for Literary Texts," *LLC* 3 (1988) 26-31.

[4] See D. Archer, "Digital Humanities 2006: When Two Became Many", *LLC* (2007) 103–8.

[5] A few months after the *Oxford Handbook of Computational Linguistics* (Oxford: Oxford University Press, 2003), edited by Ruslan Mitkov.

Today, the discursive construction of the so-called Digital Humanities paradigm as a discipline[6] revolves entirely around the search for a coherent epistemological perspective,[7] or the convergence of various heterogeneous practices.[8] Yet among the multiple attempts to circumscribe the nature of DH, those that turn to bibliometrics as a method of inquiry are particularly noteworthy, as they have revealed a growing interest on the part of various fields of *informatica umanistica* in a range of topics traditionally cultivated by computer science and computational linguistics, among which that of so-called AI stands out[9]. The goals of the Resilient Septuagint project lie precisely within this context.

2. Specific features of Italian scholarly tradition

Somewhat surprisingly, especially considering one of the founders of the so called *informatica umanistica* (once again, I find it more appropriate to adhere to the Italian definition) was the Jesuit Roberto Busa, Italy has always been on the sidelines of scientific debates sparked by Digital humanities in Biblical studies. Two main points can summarize the Italian context. The first—though less central to the present discussion—is the pioneering nature of many good practices at their inception, which, however, often proved short-lived.[10] The CNR, in particular, has managed to promote a number of initiatives: part of the legacy of Father Busa (1946),[11] with his student Antonio Zampolli (appointed in Pisa to the first university chair of Computational linguistics), who founded the *Istituto di Linguistica Computazionale* in 1998; the legacy of Tullio Gregory and Tullio de Mauro with the Lessico Intellettuale Europeo (1964, which sought to make the specificities of the use of a given lexicography computable),[12] The CNR also launched its own major projects, such as the work

[6]Alternatively, in the face of the difficulties encountered, proposals have been made in terms of "community" (L. Spiro, "'This Is Why We Fight': Defining the Values of the Digital Humanities," in M.K. Gold [ed.], *Debates in the Digital Humanities* [Minneapolis/London: University of Minnesota Press, 2012] 16-35) or "community of practice" (A. Burdick/J. Drucker/P. Lunefeld/T. Presner/J. Schnapp, *Digital Humanities* [Cambridge MA/London: The Mit Press, 2012] and C. Ross, "Social Media for Digital Humanities and Community Engagement," in C. Warwick/M. Terras/J. Nyhan [ed.], *Digital Humanities in Practice* [Cambridge: Cambridge University Press, 2012]), or "movement" (M. Kirschenbaum, "What Is Digital Humanities and What's It Doing In English Departments," in M.K. Gold [ed.], *Debates in the Digital Humanities* [Minneapolis/London: University of Minnesota Press, 2012] 3-11 and D. Parry, "The Digital Humanities or a Digital Humanism," in M.K. Gold [ed.], *Debates in the Digital Humanities* [Minneapolis/London: University of Minnesota Press, 2012] 429-37), or even as "experiments in interdisciplinarity" (I. Damian/M. de Almeida/T. de Mello/P. Rodrigues, "Convergências entre as Humanidades Digitais e a Ciência da Informação," *Ibersid* 9 [2015], 79-82).

[7] T. Koltay, "Library and Information Science and the Digital Humanities: Perceived and Real Strengths and Weakness", *Journal of Documentation* 72 (2016) 781–92.

[8] J. Schnapp/T. Presner, "Digital Humanities Manifesto 2.0", 2009 (retrieved from https://jeffreyschnapp.com/wp-content/uploads/2011/10/Manifesto_V2.pdf, March 29, 2025).

[9] See G. Spinaci/G. Colavizza/S. Speroni, "A Map of Digital Humanitites Research Across Bibliographic Data Sources", *DSH* 37 (2022) 1254–68.

[10] D. Buzzetti, "Prefazione. Oltre il limite istituzionale", in F. Ciotti (ed.), *Digital Humanities. Metodi, strumenti, saperi* (2023: Carocci, Roma), 15–8 and, for the following, see Orlandi in T. Orlandi/F. Tomasi, "Una storia dell'informatica umanistica in Italia", in F. Ciotti (ed.), *Digital Humanities. Metodi, strumenti, saperi* (Roma: Carocci, 2023), 35–47, on pp. 41–42.

[11] The other – substantial – part belongs to the CIRCSE (Centro Interdisciplinare di Ricerche per la Computerizzazione dei Segni dell'Espressione) of the Università Cattolica del Sacro Cuore in Milan, now directed by Marco Passarotti (PI of the ERC-consolidator grant LiLa project https://lila-erc. eu/#page-top last viewed March 9, 2025), an evolution of Father Busa's GIRCSE (Gruppo Interdisciplinare di Ricerche per la Computazione dei Segni dell'Espressione).

[12] These insights predate the findings of English- and French-language historiography by about thirty years; see V.B.Y. Ooi, *Computer Corpus Lexigraphy.Edinburgh Textbookks in Empirical Linguistics* (Edinburgh: Edinburgh University Press,

of Giuseppe Savoca, an Italianist from Catania, who developed the CLIPON, i.e. *Concordanze della Lingua Italiana Poetica di Otto-Novecento* (1977), and the work of the Crusca dictionary, edited by Silvio Avalle (1979), which lemmatized a series of concordances of the Italian language aiming at constructing a historical dictionary of the Italian language. Aside from some intrinsic limitations of certain projects—stemming partly from the so-called technological turn that tipped the balance toward representation at the expense of computation, and partly from the persistent hesitation within the humanists toward the digital realm—the main obstacle to meaningful progress in terms of structure and systematization has been the lack of academic recognition within the Italian university system.[13]

The second feature is more closely connected to our project. It can be read between the lines of the final essay published by the late Dino Buzzetti—his preface to the volume *Digital Humanities 2023*, edited by Fabio Ciotti—which, in some respects, stands as his intellectual legacy. As early as 1993, drawing on how archaeological data can be described, French archaeologist Jean-Claude Gardin proposed a binary classification of languages: on one side, logical-scientific language; on the other, literary, symbolic, or everyday language. These correspond to two distinct ways of engaging with information technology—either as guardians or as arbiters of computational techniques.[14] Buzzetti traced the entire Italian landscape back to these two halves.[15] one of which, however, is predominant[16] and stems from the "rifiuto della visione strumentalista dell'informatica," based instead, as Tito Orlandi has consistently argued, on methodology and epistemology. According to Buzzetti, the challenges posed by deep learning may offer a genuine opportunity to reconcile these two visions—not only of humanities computing, but also of the human dimension of science itself. In other words, for Buzzetti, deep learning represented a transposition to a new level of what Gardin, back in 1993, had described as a third way between rigorous deduction and narrative argumentation. Gardin had identified this middle ground in the logical formalization of narrative, referring to it as *logicisme*. Buzzetti went a step beyond Gardin, urging the Italian tradition to take a leap forward and view deep learning as the most fitting heir to Tito Orlandi's teachings on the impact of information technology on humanistic research—specifically, the profound transformation in the nature and foundations of rigorous humanistic inquiry through formalization, modeling, and semiotic representation.[17]

From this perspective, the challenge posed by Buzzetti implies that Ciotti—a scholar who has significantly shared and expanded upon Orlandi's insights—must now pursue four parallel paths. Talking about formalization means relating the "due grandi macroclassi di metodi computazionali applicabili in contesti umanistici: quelli qualitativi e simbolici […] e quelli quantitativi".[18] Speaking of modeling means reflecting on the concept of model and data structures and the human role in validating both. Representing these complex data means designing the appropriate infrastructure.

1998) and D. Piotrowski (ed.), *Lexicographie et Informatique: Autour de l'informatisation du Trésor de la Langue Française* (Paris: Didier Erudition, 1998).

[13] In Italy, there is no specific SSD, and while it is true that the new ministerial declarations (2024) introduce the so-called "digital humanities" on many fronts, in doing so they also exponentially increase the fragmentation of a discipline that continues to exist only in its potential.

[14] See J. Gardin, "*Points de vue logicisties sur les méthodologies en sciences sociales*" *Sociologie et sociétés* 25 (1993), 11–22, on p. 13 and p. 19; p. 14.

[15] See Buzzetti, "Prefazione", 17–18—to which Ciotti actually gives a precise name, on p. 26, respectively "versione a forte caratterizzazione epistemologica" of the "'paradigma metodologico'" of digital humanities and "versione 'strumentalista'".

[16] See Ciotti, "Introduzione", 28.

[17] See Ciotti, "Introduzione", 26 and 29 and Orlandi in Orlandi/Tomasi, "Una storia", 40.

[18] See Ciotti, "Introduzione", 31.

Gino Roncaglia has re-proposed this same fourfold division in a history of the interaction between Computer Science engaged in artificial intelligence and humanistic informatics. Moreover, for Tito Orlandi, the principles of formalization, modeling, and representation (for computer scientists, coding) have always been at the basis of both Computer science and the Humanities.

3. The road to take

The history of the interaction between AI and DH, as outlined by Roncaglia, can be summarized in a framework that brings together at least five stages:[19]

- hardware: from mainframes to smartphones;
- software: from interaction with the machine via command line interfaces to videoconferencing systems;
- economic: from the anarchic mentality of the first Silicon Valley companies of the 1970s to Big Tech companies, which, operating at the supranational level of *big data*, are a law unto themselves, thus effectively reincarnating the canonical concept (coined by Innocent IV and Cardinal Ostiense) of *potestas ordinaria*;[20]
- socio-evolutionary: from the era of scarce resources and fragmented content to the era of commerce and trade, characterized by abundant resources and complex content;
- gnoseological-epistemological: from the first prototypes of digital encyclopedias in the 1950s to the first online catalogs (early 1980s) and the integration of semantic web and generative AI.

The advantage of Roncaglia's approach is that it allows us to understand with extraordinary clarity where we are and where we are going. Roncaglia puts forward a further evolutive proposal that summarizes all five previous periodizations: the idea we have today of artificial intelligence is the evolution of an encyclopedic organization of knowledge, or rather a library system,[21] and follows the suggestions that come from great works of science fiction (especially Isaac Asimov, the *Foundation* series; Neal Stephenson, *Snow Crash*; Ted&Bob Rockwell, *Virtual Librarian*). It seems to me that, in doing so, Roncaglia gives real substance to Buzzetti's intuition: Buzzetti's proposed leap calls for an intertwining—or rather, a symbiotic development—between the literary sciences and computer science within the domain of generative AI.[22] But Roncaglia goes further and revisits what he had already discussed in his *L'età della frammentazione*,[23] arguing that we are

---

[19] Cfr. G. Roncaglia, *L'architetto e l'oracolo. Forme digitali del sapere da Wikipedia a ChatGPT* (Roma/Bari: Laterza, 2023), 5–21.

[20] This reflection was captured in William Gibson's cyberpunk novel *Neuromancer* (1984), which created the concept of cyberspace to convey the idea of a virtual reality that colonizes the physical one. See W. S. Haney, *Cyberculture, Cyborgs and Science Fiction: Consciousness and the Posthuman* (Amsterdam/New York: Rodopi, 2006).

[21] Hence the reference to Borges in the title.

[22] In addition to the arguments put forward by Roncaglia (and Ciotti), there are further indications that support the impression of a growing convergence between narrative approaches and formal-deductive logic. One such sign is the recent introduction—alongside MA programs in Digital Humanities—of LM-39 programs in linguistics with a strong computational focus (at institutions such as the Università Cattolica and the University of Bologna). This shift reflects a broader rethinking and re-foundation of classical humanistic disciplines (such as nineteenth-century philology) from an *ab initio* digital perspective rooted in linguistics—an approach exemplified by Buzzetti's concept of Knowledge Design (see Buzzetti, "Prefazione", 17), as well as by data analysis models that incorporate not only statistically based artificial intelligence but also machine learning (see F. Ciotti, "L'analisi del testo", in F. Ciotti [ed.], *Digital Humanities. Metodi, strumenti, saperi* (Roma: Carocci, 2023), 91–113, on pp. 97–112).

[23] See G. Roncaglia, *L'età della frammentazione. Cultura del libro e scuola digitale*, Laterza, Roma-Bari, 2020[2] (2018).

moving towards complex granularity (tools that manage atoms that are highly articulated and structured internally) that interacts with the main Large Language Models.

In an era like ours — still lagging behind in the digitization of large corpora — it is clear that the *conditio sine qua non* for any progress along the path outlined by Roncaglia cannot be a merely theoretical workflow, to be implemented only once the appropriate IT tools arrive.[24] The research process must necessarily move between high-level humanistic inquiry and state-of-the-art research in the field of Computer Science.

Although it has not yet made a significant impact in the field of artificial intelligence, biblical studies represent an area where digital transformations have developed early and in a well-structured way. It is precisely in continuity with these experiences—and with the already established infrastructure for the digital processing of sacred texts—that our project aims to position itself, proposing a possible extension toward integration with AI tools and methods. By biblical studies, I do not refer only nor so much to the contributions of theologians or historians of religions or Christianity.[25] Rather, I am referring to two significant results that emerged from decades of scholarly exchange, research, and collaboration—activities that have, directly or indirectly, revolved around the Society of Biblical Literature since the 1980s. The first, linked to the figure of John Abercrombie and dating to the early 1980s, involved the digitization of Greek literary *corpora* for the Thesaurus Linguae Graecae. By the time of Rahlfs' edition of the Septuagint (1935), there was a need to align the Greek of the Septuagint Bible and the Hebrew of the Masoretic Text.[26] This led to the CATSS project (1986), which sought to add a database preserving every single variant reading of the First Testament in Greek to the existing alignment, as well as a morphological analysis of all words of both the LXX and the Masoretic Text.[27] From the technological point of view, this meant aligned texts that were digitized, searchable (with a David Packard IBYCUS personal computer), and morphologically analyzed.[28] In practical terms, writing a simple binary algorithm operating within a table for words and a tree structure for desinences proved decisive.[29]

In a more general way, we might say that the late 1980s and early 1990s were fortunate times for computational linguists interested in the sacred scriptures of the Judaic and Christian traditions.[30] In 1998, CATSS was integrated into Accordance, the second of the so-called biblical

---

[24] See D. Hamidovic, “An Introduction to Emerging Digital Culture”, in C. Clivaz/P. Dilley/D. Hamidović (ed.), *Ancient Worlds in Digital Culture* (Leiden/Boston: Brill, 2016), 1–16, on p. 4.

[25] As far as the digital turn is concerned, theology (as well as "historical-religious knowledge") is somewhat like Hegel’s owl of Minerva, and nevertheless theologians have offered an authoritative review on the ethical dilemmas of certain issues (especially those related to digital platforms) in the pages of *TL* 149 (2024) 4-23. This was noted, moreover, by Claire Clivaz, in a 2017 essay she recently republished at the opening of the third volume of De Gruyter's series *Introductions to Digital Humanities - Religion*: *The Bible in the Digital* Age: multimodal Scriptures in Communities , see C. Clivaz, “The Bible in the Digital Age: Multimodal Scriptures in Communities,” in T. Hutchings/C. Clivaz (ed.), *Digital Humanities and Christianity. An Introduction* (Berlin/Boston: De Gruyter, 2021) 21-46. I also note this essay for mapping academic efforts to digitize Scripture *corpora* from the Jewish and Christian traditions.

[26] See E. Tov, *The Greek and Hebrew Bible. Collected Essays on the Septuagint* (Leiden/Boston/Köln: Brill, 1999) 31.

[27] See http://ccat.sas.upenn.edu/rak/catss.html (retrieved March 29, 2025). See E. Tov, *A Computerized Database for Septuagint Studies: The Parallel Aligned Text of the Greek and Hebrew Bible* (Stellenbosch: JNSL, 1986).

[28] In that case, thanks to a program David Packard designed in 1973 for IBM Mainframe 390/91 at Nasa's Goddard Flight Center in Washington DC (1966) and supervision by a team of experts.

[29] See D.W. Packard, “Computer-Assisted Morphological Analysis of Ancient Greek”, in A. Zampolli/N. Calzolari (ed.), *Computational and Mathematical Linguistics: Proceedings of the International Conference on Computational Linguistics* (Firenze: L.S. Olschki, 1977–1980), 343–55, on p. 348.

[30] On the New Testament see: M.E. Davison, “New Testament Greek Word Order”, *LLC* 4 (1989) 19–28; H.H. Greenwood, “St Paul Revisited—A Computational Result”, *LLC* 7 (1992) 43–7, “St Paul Revisited—Word Clusters in Multidimensional Space”, *LLC* 8 (1993) 211–9, and “Common Word Frequencies and Authorship in Luke’s Gospel and Acts”, *LLC* 10 (1995)

software[31] created to handle ancient Greek (after BibleWorks, created in 1992 as support for preaching and pastoral activities).[32] Today, the digital side of biblical studies offers a wide array of source collections that, from a technological standpoint, are more concerned with data representation (and knowledge) than with data processing.[33] These platforms manage the copyrights of various dictionaries and editions for a fee, including Olive Tree, Accordance, Logos, BibleWorks, Brill Dead Sea Scrolls Concordance (which uses and refines Abegg's database—through texts edited by the Oxford series *Discoveries in the Judean Desert*—and now offers it integrated into Brill's Electronic Library), or open source (such as STEP Bible).

The second contribution of biblical scholars is editorial. It has primarily taken shape within the framework of the International meetings of the Society of Biblical Literature and the activities of the European Association of Biblical Studies. These two groups of researchers, which began their dialogue between 2012 and 2013, should be credited with the launch—starting in 2016—of Brill's *Digital Biblical Studies* series,[34] initially coordinated by two Swiss professors, David Hamidović and Claire Clivaz, along with Andrew Gregory. From the very beginning the series has had a strong international orientation—even prior to the appointment of Garrick Allen and Paul Dilley as editors in 2019. From Volume I, scholars such as Marco Büchler, Joseph Verheyden, and Daniel Apollon were already involved, and by Volume II (2018), the French scholar Laurence Mellerin had also joined.

3.1. The state of the art in semantic analysis of the Greek Bible

A semantic analysis for research on the Greek Bible (or on the Greek "Bibles," if we consider the plurality of originally circulating versions)[35] must be conducted on two levels: first, the way the text was reconstructed; and second, the coexistence of multiple parallel versions, each of which reflects the individuals who produced it and the broader social and human context from which it emerged. These are two distinct ontological levels, both historical in nature. The first is that of the event itself, expressed in acts such as copying manuscripts or critically reconstructing a text—forms of

183–87; G. Ledger, "An Exploration of Differences in the Pauline Epistles using Multivariate Statistical Analysis", *LLC* 10 (1995) 85-97; D.L. Mealand, "Correspondence Analysis of Luke", *LLC* 10 (1995) 171-82, "Measuring genre differences in Mark with Correspondence Analysis", *LLC* (1997) 227–45; A.J.M. Linmans, "Correspondence Analysis of the Synoptic Gospels", *LLC* 13 (1998) 1–13; G.K. Barr, "A Computer Model for the Pauline Epistles", *LLC* 16 (2001) 233–50; G.K. Barr, "Interpolations, Pseudographs, and the new Testament Epistles", *LLC* 17 (2002) 439–55; A. Wilson, "Developing Conceptual Glossaries for the Latin Vulgate Bible", *LLC* 17 (2002), 413–26; G.K. Barr, "Two Styles in the New Testament Epistles", *LLC* 18 (2003) 235–48. On the Old Testament it is worth mentioning the third section of the Historical Dictionary of the Hebrew Language (see R. Merkin/Z. Busharia/E.Meir, "The Historical Dictionary of the Hebrew Language", *LLC* 4 [1989] 271–3) and the application of the famous TUSTEP (Tübingen System von Textverarbeitungsprogrammen) to the Book of Daniel (W. Bader [ed.], *Und die Wahrheit wurde hinweggefegt: Daniel 8 linguistisch interpretiert* [Tübingen: Franke, 1994]).

[31] Collections of digitized biblical versions and biblical commentaries began to circulate in 1980.

[32] Accordance would do this by merging Bibles from Roy Brown's biblical software, i.e., ThePerfectWord, later "MacBible," and the Dead Sea Scrolls word database of another Mac-user, i.e., Martin G. Abegg. See Tov, *Greek and Hebrew Bible*, 43.

[33] See Buzzetti, "Alle origini dell'informatica umanistica", 74–77.

[34] See C. Clivaz/S. Bowen Savant, "The Dissemination of the Digital Humanities within Research on Biblical, Early Jewish and Christian Studies", in D. Hamidović/C. Clivaz/S. Bowen Savant, *Ancient Manuscripts in Digital Culture. Visualisation, Data Mining, Communication* (Leiden/Boston: Brill, 2019), 1–12, on pp. 1–5.

[35] A. Mambelli and I have already pointed out this in D. Dainese/A. Mambelli, "Intertestualità tra *Bibbie* e antichi commentary cristiani: l'esempio di *simul* nel *De Genesi ad litteram* di Agostino", *Lexicum Philosophicum* 11 (2023-2024) 39–65. See also *infra*.

temporality measurable in human lifespans, or portions thereof. The second concerns the *conditions of possibility*—what Reinhart Koselleck referred to as historical structures—which possess their own temporalities and units of measurement that are not commensurable with those of individual human lives, ranging from the subjectivity of specific moments to the span of entire generations. The first plane corresponds to a type of knowledge that can be formalized in binary terms (true/false), representable—not surprisingly—in graph form (as in the case of a a *stemma codicum*) and manageable by inferential engines. Yet between the dimension of the event and that of its condition of possibility, there lies a space that cannot be formalized through binary combinations alone. The 'condition of possibility' has its own ontological status, distinct from that of the actual happening—that is, it defines the realm of what could happen but did not, what remains thinkable, traceable, or reconstructable beyond actual events. Conceptually, it aligns more closely with the Platonic idea or with Aristotle's notion of the 'verisimilar' in his *Poetics*. The range of cases it encompasses cannot be reduced to just two categories (T/F). If an event could have unfolded differently from how it actually did (true), then its alternatives are not limited to a single 'false' counterpart, but rather comprise a potentially infinite spectrum.

This logic applies, to a certain extent, also to texts and to the history of their transmission. Critical apparatuses that document different versions of a text can be seen as traces of its true nature: plural, and deeply dependent on the multiplicity of contexts in which it was copied and in which it has concretely 'lived'. These are precisely the kinds of issues familiar to those who study biblical reuse and intertextuality in patristic literature—forms of reuse that, it is worth recalling, may also influence textual reconstruction in cases where manuscript data are insufficient.

In fact, the very distinction between true and false proves inadequate in such cases, as it imposes a hierarchy shaped by the judgment of the modern editor, at the expense of the text's real and irreducible plurality. At the same time, neither intending nor being able to overturn entire philological traditions, it is clear that information concerning the reconstruction of the text—and the relative primacy assigned to a given version—must in some way be preserved.

It therefore seemed more appropriate to think in terms of a gradation, a scoring logic, or fuzzy logic—precisely the kind that is now managed by neural networks, but which had already been tested in the 1980s as a computational model for representing the semantics of natural language.[36] If, however, the relationship between the reconstructed text and its actual circulation eludes the constraints of binary logic, the way in which information about the text's natural and original plurality—and that concerning its reconstruction—is encoded in a critical edition is entirely manageable in binary terms. It is both possible and conceptually straightforward to link the words of a philological annotation to those in the edited text to which the note refers. And this is a starting point and a primary result to preserve.

3.1.1. Different Approaches to Textual Semantics

As should be clear by now, speaking of the 'semantics of natural language' when dealing with the relationship between 'historical structure' and manuscript production points to a problem that goes beyond the strictly linguistic dimension. It involves, rather, an inherently intertextual one—since the meaning of the biblical text is shaped by its reception. This leads us into a broad and, in some ways, theoretically and within the domain of literary theory, autonomous field of research—yet one for which there already exists a substantial tradition of digital and computational inquiry, which deserves brief consideration. Overall, we can identify two main approaches, which have at times

[36] See L.H. Zadeh, "Text-Score Semantics as a Basis for Computational Approach to the Representation of Meaning", *LLC* 1 (1986) 24–35.

coexisted and at times alternated over the years the classificatory/ontological approach[37] and the text-reuse approach.[38] Text-reuse algorithms in particular—heirs of the language agnosticism of machine translation in the 1940s[39] but still the most widespread technology for research in intertextuality within Ancient Greek literature[40]—are technologically classical and reliable old-school tools that pose a philological problem. Even when they allow for complex granularity—enriched with annotations on morphology or semantics—they do not enable the information provided by the critical apparatus of the main reference editions to be exploited. Working solely with edited texts, without any data on manuscript transmission, does not amount to real knowledge. It is a form of reduced, contextless information—something closer, hyperbolically speaking, to a social media post than to historical scholarship. A third approach may lie in statistical static embedding models (e.g., topic modelling, but various techniques have been tested),[41] i.e., based on words or sequences of words associated with vectors whose co-occurrences in linguistic contexts are counted,[42] using techniques that are still widely used by digital humanists[43]—they date back to the late 1990s until 2013 (the year of the transition to predictive models with word2vec), but they might perform quite well in some circumstances.[44]

---

[37] See most recently J. Horstmann/C. Lück/I. Normann, 'Systems of Intertextuality: Towards a Formalization of Text Relations for Manual Annotation and Automated Reasoning', *DHQ* 17 (2023) 1–74, but R. Hohli Trillini/S. Quassdorf, "A 'Key to All Quotations'? A Corpus-based Parameter Model of Intertextuality", *LLC* 25 (2010) 269–86 had already discarded classifications based on types considered classic but not based on formal logic. In general, this is undoubtedly a fruitful approach in terms of both building digital archives and long-term preservation, see T.L. Andrews/C. Macé (ed.), *Analysis of Ancient and Medieval Texts and Manuscripts: Digital Approaches* (2014: Brepols, Turnhout) (still a reference work – proof of how much can/must still be done in many areas of DH), exp. on pp. 183–244, i.e. the chapters by L. Spinazzè, S. Rubenson, C. Tupman and A. Jordanous, M. Romanov; G. Tomazzoli, "Intertextuality in Dante's Commedia: Hypermedia Dante Network", *Bibliotheca Dantesca* 5 (2022) 308–11. To a certain extent, research based on network analysis can be placed within this area, see M. Romanello, "Exploring Citation Networks to Study Intertextuality in Classics", *DHQ* 10 (2016) 1–46. Worth mentioning is D. Bamman/G. Crane, "The Logic and Discovery of Textual Allusion", in *Proceedings of the Second Workshop on Language Technology for Cultural Heritage Data [LaTeCH 2008]* at http://www.lrec-conf.org/proceedings/lrec2008/workshops/W22_Proceedings.pdf#page=31 last viewed March 2, 2025.

[38] In this respect, the best result to date remains Tracer by Marco Büchler, which can be adapted to any type of text. See the main publications at http://www.etrap.eu/research/tracer/ (last viewed March 29, 2025). Also in his PhD thesis one may find the best review of all of the text-reuse approaches: M. Büchler, *Informationstechnische Aspekte des* Historical Text Re-use (2013: [Universität Leipzig, Leipzig]), 58.

[39] See W. S. Bennett, "The Place of Semantics in MT Systems", *LLC* 4 (1989) 200–2.

[40] See T. Sommerschield/Y. Assael/J. Pavlopoulos, "Machine Learning for Ancient Languages: A Survey", *Computational Linguistics* 49 (2023) 703–47, on p. 723.

[41] I would mention, in particular, T. Merriam, "Intertextual Distances, Three Authors", *LLC* 18 (2003) 379–88, who adapted Labbé's metric to text strings with the same number of tokens.

[42] An approach with a deep-rooted basis in computer-aided research, see S. Gillmayr-Bucher, "A Computer-aided Quest for Allusions to Biblical Texts within Lyric Poetry", *LLC* 1 (1996) 1–8 (a scholar who is particularly attentive to the issue of intertextuality, see her "Intertextualität. Zwichen Literaturtheorie und Methodik", *PzB* 8 [1999] 5–20).

[43] See W. Scheirer/C. Forstall/N. Coffee, "The Sense of a Connection: Automatic Tracing of Intertextuality by Meaning," *DSH* 31 (2017) 204–2017 up to F. Ciotti (ed.), *Digital Humanities*. Encouraging results are being achieved in the field of semantic change, see B. McGillivray/S. Henchen/V. Lähteenoja/M. Palma/A. Vatri, "A Computational Approach to Lexical Polysemy in Ancient Greek" *DSH* 34 (2019) 893–907 and V. Perrone/M. Palma/S. Hengchen, *et al.* "GASC: Genre-Aware Semantic Change for Ancient Greek", in N. Tahmasebi/L. Borin/A. Jatwot *et al.*, *Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change* (Berlin: Language Science Press, 2021) 56–66.

[44] S. Stopponi/N. Pedrazzini/S. Peels/B. McGillivray/M. Nissim, "Evualuation of Distributional Semantic Models of Ancient Greek: Preliminary Results and a Road Map for Future Work", in *Proceedings of the Ancient Language Processing Workshop* (Shoumen: Incoma, 2023) 49–58 and T. Köntges, "Measuring Philosophy in the First Thousand Years of Greek

3.1.2. Back to the Future: Bridging Old-School and Innovation

In light of the above and particularly of the approach outlined by Roncaglia, we sought to explore a different path. We have tried to formalize the process of text reuse as a form of re-signification—one that corresponds to the construction of complex synonymies. In this framework, since we had to build a semantic tool that also involved a linguistic dimension, we started with WordNet, another area of NLP research, conceived by American linguists (George A. Miller) in the early 1980s but still active today.[45] WordNets are relational databases built using manual semasiological POS tagging (i.e., looking at the concept that a word expresses, aware that the same concept can be expressed by different words), but dictionary-like in nature. The specific case of Ancient Greek WordNet is interesting because, due to its structure, it articulates the entire semantics of ancient Greek as if it were a library, linking the semantic areas of human language to classes of the Dewey decimal system. This allowed us to start from a library, as brilliantly intuited by Roncaglia, who had pointed us toward the next step: its interaction with generative AI resources.

Computational linguists (in particular from KU-Leuven and the University of Heidelberg) have attempted to move in this way, with only partial success (technically it seems possible, but the results were insufficient in the automation phase).[46] That said, the key point for us lies elsewhere: while they certainly offered a coherent response by taking the leap that Buzzetti had envisioned, it remains a DH answer to a DH question (i.e. it stays within the field of computational semantics, oriented by the paradigm of knowledge representation).[47] The best result this research field has

---

Literature", *Digital Classics Online* 6 (2020), https://doi.org/10.11588/dco.2020.2.73197. In general: D.M. Blei/A.Y. Ng/M.I. Jordan, "Latent Dirichlet Allocation", *Journal of Machine Learning Research* 3 (2003) 993–1022.

[45] See C. Fellbaum (ed.), *WordNet. An Electronic Lexical Database* (London: MIT Press, 1998). Over the years, WordNets have proven effective for word sense disambiguation (E. Agirre/O. López de Lacalle/A. Soroa, "Random Walks for Knowledge-based Word Sense Disambiguation" *CL* 40 [2014] 57–84 and S. Melacci/A. Globo/L. Rigutini, "Enhancing Modern Supervised Word Sense Disambiguation Models by Semantic Lexical Resources", in N.Calzolari/K. Choukri/C. Cieri [ed.] *et al.*, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation [LREC 2018]* [Miyazaki: European Language Resources Association, 2018] 1012–17), summarization (A.R. Pal/D. Saha, "An Approach to Automatic Text Summarization Using WordNet", in *2014 IEEE International Advance Computing Conference [IACC]* [Washington DC: IEEE 2014] 1169–73; N. Xie/S. Li, H. Ren *et al.*, "Abstractive Summarization Improved by WordNet-Based Extractive Sentences", in *CCF International Conference on NLP and Chinese Computing* [Cham: Springer, 2018] 404–15), query expansion (M. Lu/X. Sun/S. Wang *et al.*, "Query Expansion via WordNet for Effective Code Search" *2015 IEEE 22nd International Conference on Software Analysis, Evolution, and Reengineering [SANER]* [Washington DC: IEEE, 2015] 545–49 and W. Li/S. Wang/Z. Yu, "Deep Learning and Semantic Concept Space Are Used in Query Expansion", *ACCS* 52 [2018] 175–83) and clustering (C. Wang/Y. Song/A. El-Kishky *et al.*, "Incorporating World Knowledge to Document Clustering via Heterogeneous Information Networks", in *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* [New York: Association For Computing Machinery, 2015] 1215–24; L. Stanchev, "Semantic Document Clustering Using Information from WorNet and DBPedia", in *2018 IEEE 12th International Conference on Semantic Computing [ICSC]* [Washington DC: IEEE, 2018] 100–7).

[46] See A. Keersmaekers/W. Mercelis/T. Van Hal, "Word Sense Disambiguation for Ancient Greek: Sourcing a Training Corpus through Translation Alignment", in *Proceedings of the Ancient Language Processing Workshop* (Shoumen: Incoma, 2023), 148–59. There is more production on this topic, which is still forthcoming. Similar results in F. Riemenschneider/A. Frank, "Exploring Large Language Models for Classical Philology", in *Proceedings of the 61st Annual Meeting of the Association for Computational Lingustics* (3vols.; New York: Curran, 2023) 1.15181–15199. As in the case of Latin, semi-automatic refinement techniques have not fully met expectations, see G. Franzini/A. Peverelli/P. Ruffolo *et al.*, "*Nunc Est Aestimandum*. Towards an Evaluation of the Latin WordNet," in *Proceedings of the Sixth Italian Conference on Computational Linguistics* (Turin: Associazione Italiana di Linguistica Computazionale, 2019) https://dx.doi.org/10.5281/zenodo.3518774.

[47] See P. Blackburn/J. Bos, *Representation and Inference for Natural Language. A First Course in Computational Semantics* (Stanford: CLSI, 2005)—to be compared with Buzzetti, "Alle origini dell'informatica umanistica", 74–77.

achieved is the significant growth of automatically annotated Greek *corpora*[48] as well as tools for automatic linguistic analysis[49] and language models (see *infra*). Semantics, however, is far from being satisfactory, especially for the specific needs of biblical research (see *infra*)—the main and most basic issue is that the number of available tokens is way below the threshold required for effective training of neural networks.

Since our research question lies upstream than the linguistic one, our approach takes a different direction: how can purely biblical-patristic research help advance semantic analysis—originally shaped by computational methods—toward a new stage of development, one generated through the interaction between the humanities (both digital and traditional) and generative AI? This calls for a return to the foundational phase of 'humanities computing' era, oriented by methods and their epistemological foundational tools.[50]

4. Needs and responses

4.1. Towards a Semantic of "killing" and "healing" in the Greek Bible

The methodological approach of the project on semantics is nestled within the reflections on the value of semantic research, and in particular lexical research, that have characterized Biblical studies

[48] E.g. Perseus under PhiloLogic by H. Dik, see https://perseus.uchicago.edu/ last viewed April 1, 2025; Diorisis by A. Vatri and B. McGillivray, see A. Vatri/B. McGillivray, "The Diorisis Ancient Greek Corpus. Research Data", *Journal for the Humanities and Social Sciences* 3 [2018], 55–65; GLAUX by Keersmaekers, see A. Keersmakers, "The GLAUx Corpus: Methodological Issues in Designing a Long-Term", in N. Tahmasebi/A. Jatowt/Y. Xu *et al.* [ed.], *Proceedings of the 2nd International Workshop on Computational Approaches to Historical Language Change 2021* [New York: Curran, 2021], 39–50; OGA by G.G.A. Celano, see G.G.A. Celano, *Opera Graeca Adnotata*, in https://zenodo.org/records/14206061). For manually annotated corpora see: PROIEL, by D.T.T. Haug and M.L. Jøhndal (D.T.T. Haug/M.L. Jøndal, "Creating a Parallel Treebank of the Old Indo-European Bible Translations", in C. Sporleder/K. Ribarov [ed.], *Proceedings of the Second Workshop on Language Technology for Cultural Heritage Data [LaTeCH 2008]* [available at http://www.lrec-conf.org/proceedings/lrec2008/workshops/W22_Proceedings.pdf#page=31, last viewed January 18, 2025]); AGDT by D. Bamman, F.Mambrini and G. Crane (see their "An Ownership Model of Annotation: Tha Ancient Greek Dependency Treebank", in M. Passarotti/A. Przepiórkowski/S. Raynaud *et al.* [ed.], *Proceedings of the Eighth International Workshop on Treebanks and Linguistic Theories* [Milan: EDUCatt, 2009] 5–16); Pedalion by A. Keersmaekers W. Mercelis T, C. Swaelens, T. Van Hal (see their "Creating, Enriching and Valorising Treebanks of Ancient Greek: the Ongoing Pedalion-project", in M. Candito/K. Evang/S. Oepen *et al.*, *Proceedings of the 18th International Workshop on Treebanks and Linguistic Theories* [Paris: Association for Computational Linguistics, 2019] 109–117); Gorman by V.B. Gorman (see her "Dependency Treebanks of Ancient Greek Prose", *Journal of Open Humanities Data* 6 [2020] https://doi.org/10.5334/johd.13); Harrington Treebanks by M. Harrington (see his *"Perseids Project – Treebanked Commentaries at Tufts University"*, https://perseids-project.github.io/harrington_trees/, last viewed January 18, 2025); Aphtonius by P. YorPapydanova (see her "Treebank of Aphtonius, Progymnasmata" https://github.com/polinayordanova/Treebank-of-Aphtonius-Progymnasmata, last viewed January 18, 2025)]; PapyGreek by M. Vierros and E. Henriksson (see their "PapyGreek Treebanks: A dataset of linguistically annotated Greek documentary papyri" *Journal of Open Humanities Data* 7 [2021] https://doi.org/10.5334/johd.55) annotated. In general, it is probably the Leuven researchers who have experimented the most. See A. Keersmaekers/T. Van Hal, "A Corpus-Based Approach to Conceptual History of Ancient Greek", in G. Kristiansen/K. Franco/S. De Pascale *et al.* (ed.), *Cognitive Sociolinguistics Revisited* (Berlin/Boston: Brill, 2022) 213–25.

[49] See: the case of Gasc (*supra*); A. Vatri/B. McGillivray, "Lemmatization for Ancient Greek. An Experimental Assessment of the State of the Art", *Journal of Greek Linguistics* 20 (2020) 179–196; C. Palladino/K. Farimah/B. Mathiak, "NER on Ancient Greek with Minimal Annotation", *https://dh2020.adho.org/* (last viewed March 2, 2025) and T. Yousef/C. Palladino/S. Jänicke, "Transformer-Based Named Entity Recognition for Ancient Greek", http://dx.doi.org/10.13140/RG.2.2.34846.61761.

[50] See with Buzzetti, "Alle origini dell'informatica umanistica", 99.

in the last decades.[51] In the case of the translation of the Greek Bible, the so-called Septuagint, there are by now decades of research on the importance of the analysis of lexical choices made by the translators to understand a wide range of phenomena, undoubtedly linguistic, but also socio-historical in the original context of production of the corpus, not to mention the inevitable theological aspects connected to interpretation.[52] The approach of systematically studying the lexical choices of the translators of the Bible into Greek is thus rooted in a rich scholarly tradition and is further expanded by a parallel effort that is being placed on the study of the Septuagint translation not only as a relevant source for the textual criticism of the Hebrew Bible, but also as a text that pertains to the history of Greek language, literature, and culture.[53] In this frame, some semantic spheres have been analyzed in relation to how they appear in the Septuagint and to the broader flow of the Greek language's traditions, as testified by both literary and documentary sources.[54]

On the methodological basis of these currents of Septuagint research, the work on semantics characteristic to this project aims at creating a digital tool that may accompany scholars in (among others) semantic quests on the text of the Greek Bible, going beyond word lists of single lemmas considered as static objects. In this paragraph a sample of the initial semantic research on the text of the Septuagint is presented, carried out with traditional methods, yet at the same time with an eye on providing an epistemological bridge between a research based on digital databases with a limited amount of search options (such as the currently available software for the study of the Bible)[55] to the training of new search possibilities offered by a digital tool with a semantic composition. The initial pilot map of semantic spheres has been carried out on the thematic nuclei of "killing" and of "healing", which appear in a pair in Eccl. 3:3. Providing a concrete textual basis as well as sharp epistemological research questions, this preliminary research has served as a starting point to address the necessities of a Biblical scholar, to conceive a semantic search engine that may mirror and at the same time enhance the current debates on Biblical semantics.

---

[51] Biblical semantics has been under investigation with a linguistic approach since the seminal book by James Barr, *The Semantics of Biblical Language* (Oxford: Oxford University Press, 1961). His critique on the methodology of theological Biblical research on semantics have opened the debate on how to make sense of Biblical texts with the tools of modern linguistics; for a history of research in Biblical semantics and a discussion on its open questions, see Moises Silva, *Biblical Words and their Meaning. An Introduction to Lexical Semantics* (Grand Rapids: Zondervan, 1994), 17-34.

[52] For the purposes of this project, the methodological premises of Eberhard Bons (ed.), *The Historical and Theological Lexicon of the Septuagint. Volume I Alpha-Gamma* (Tübingen: Mohr Siebeck, 2020) are especially relevant. For an overview of recent scholarship, see Trevor V. Evans, «The Nature of Septuagint Greek. Language and Lexicography», in Alison G. Salvesen, Timothy Michael Law (eds.), *The Oxford Handbook of the Septuagint* (Oxford: Oxford University Press, 2021), 91-104.

[53] This line of research is now specifically pursued by the recent *Septuagint within the History of Greek's Network*, an international research network originally conceived by James K. Aitken (University of Cambridge) and Trevor V. Evans (Macquarie University) aiming at describing Septuagint language as a witness to the evolution of Greek in postclassical times.

[54] Examplary studies of this kind are Anna Passoni Dall'Acqua, «Notazioni cromatiche dell'Egitto greco-romano. La versione dei LXX e i papiri», *Aegyptus* 78 (1998), 77-115; Anneli Aejmelaeus, «Faith, Hope, and Interpretation: A Lexical and Syntactical Study of the Semantic Field of Hope in the Greek Psalter», in James C. VanderKam, Peter W. Flint, and Emanuel Tov (eds.), Studies in the Hebrew Bible, Qumran, and the Septuagint Presented to Eugene Ulrich (Leiden, Boston, MA: Brill, 2006), 360-376; Dorota Hartman, *Emozioni nella Bibbia. Lessico e passaggi semantici tra Bibbia ebraica e LXX* (Napoli: L'Orientale Università degli Studi – Napoli, 2019); Eberhard Bons, Anna Mambelli, Daniela Scialabba (eds.), *Joy Denied, Joy Rediscovered: From Classical Greek Literature to Byzantine Christianity* (Leiden, Boston, MA: Brill, forthcoming, 2025).

[55] See *BibleWorks. Software for Biblical Exegesis and Research*, 2007; *Thesaurus Linguae Graecae*, University of California, Irvine CA, 2014 (TLG); *Accordance. OakTree Software*, 2017.

To investigate a semantic sphere in the Greek Bible, it is necessary to define on what criteria the research is pursued. Any semantic work is necessarily articulated on several levels, of which the lexical one is only a partial answer, yet the word can be considered as a basic meaning-carrier unit within a text;[56] this map of the semantic spheres of "killing" and "healing" in the Greek Bible thus goes from single words to word groups, and, finally, to groups of groups of words. What can be seen as a simple analysis of the lexicon that pertains to a specific semantic sphere carries different levels of complexity, which may be conceived as proceeding by concentric circles. The image of concentric circles is apt to describe both the methodology of the study, inasmuch as it entails the progressive enlargement of the analysis described above, and of its contents, that is the lemmas, which can be made more complex than their single instances in the text (tokens) by considering the history of their semantic evolution (inter-textual) as well as their textual variants transmitted by the tradition (intra-textual).[57] By means of this simple image of concentric circles, it is possible to reflect on the ways in which the extraordinary corpus of Biblical texts in Greek expresses its senses and thus answer questions of historical linguistics and exegesis.

One of the first research questions of the project has thus been to reconstruct the semantic spheres of "killing" and "healing" within the text of the Septuagint. If the final aim of the project is to define a distance in terms of semantics and the expressions thereof between the Greek Bible and the ancient Christian authors that took inspiration from it, what was once defined the «Biblical coloring[58]» has been addressed first; before going into the linguistic and semantic characteristics of the "resilience" of the Septuagint, a map of the semantic fields under investigation has been traced within the corpus itself. The map gives scholars the opportunity to follow the distribution of a root or word or motive across the evolution of the Greek Scriptures, enlightening the workshop of the translators by examining their lexical choices in light both of the lexical material at their disposal and of the possibilities of word formation and coinage.[59] The methodological steps for the creation of such map follow the relative conceptualizations of our objects of study, and yet the "killing" and "healing" meant in the project title are not to be intended as pure epistemological labels, but rather as ideas that stem directly from the text, following the Greek version of Eccl 3:3: καιρὸς τοῦ ἀποκτεῖναι καὶ καιρὸς τοῦ ἰάσασθαι (a right time to kill, and a right time to heal).[60] What follows describes the research steps that formed and filled the concentric circles.

The research started from a simple recall of the occurrences of the two verbs, ἀποκτείνω (to kill), and ἰάομαι (to heal), in the available softwares for the study of the Bible; their distribution formed the first circle, together with all words derived from the same root.[61] By definition, all occurrences in this first circle have a complete overlap with the researched semantic sphere. A small group of other roots were added to the first circle, based on the frequent co-occurrence with the first ones and on their wide circulation in Greek, as well as the complete overlap of their occurrences

---

[56] «The interpretation of the Greek Bible still depends on a correct understanding of the words it uses. There is no easy way to arrive at this understanding» (E. Bons, *Historical and Theological Lexicon*, xii).

[57] See par. 2 for a brief exposition of the role of variant readings. It is to be noted that the image of concentric circles that progressively enlarge the search possibilities is again mirrored in the "human training" produced by the team and the WP8 *uBIQUity* in the tagging process (cf. article in this volume).

[58] The expression is found in Elias J. Bickerman, «Notes on the Greek Book of Esther», *Proceedings of the Academy for Jewish Research* 20 (1951), 101-133, 114.

[59] A more detailed account of the data will be published on the project website (or elsewhere).

[60] Transl. Peter Gentry, *New English Translation of the Septuagint* (online).

[61] The root is understood as the bearer of the meaning in all its forms, hence the need to consider the root in its entirety from a semantic point of view, but without losing sight of its specific realizations in context; see James K. Aitken, «Lexical Semantics and the Cultural Context of Knowledge in Job 28, Illustrated by the meaning of *ḥāqar*», in Ellen van Wolde (ed.), *Job 28. Cognition in Context* (Leiden, Boston, MA: Brill, 2003), 119-137.

to the analyzed semantic spheres. In the lists all the manifestations of the roots were included, even those that do not occur in the Septuagint, following the standard dictionaries; this allows the contrastive visualization of the available textual material in Greek. The latter is indicated in a different set, to keep it distinct from the strictly Biblical material; this could help in later stages of the project for the definition of the distance between texts (intra-textual search). Even though the recontextualizations implied by the Septuagint concept of resilience may involve a slightly different vocabulary, the imprint of the Biblical text must be recognizable, and the allowed deviation (delta) must be defined as precisely as possible; certainly, malleability at the root level must be allowed, precisely because of the semantic assumptions about the root itself. [62] Those lemmas are not included in this article for reasons of space. The identified roots are as shown in Table 1 below:[63]

Killing (circle 1)	**Healing (circle 1)**
ἀποκτείνω, ἀποκτείνυμι, ἀποκτεινύω (273) σφάζω (85), σφαγή (24), σφαγιάζω (2), σφάγιον, (5), κατασφάζω (11) φονεύω (54), φόνος (24) φονευτής (17), φονοκτονέω (3), φονοκτονία (1) φονώδης (1)	ἰάομαι (66), ἴασις (28) ἰατρεύω (8), ἰατρεία (2), ἰατρεῖον (1), ἰατρός (14) ὑγιάζω (10), ὑγιαίνω (49)

Table 1

The second and third circle contain roots (or, in some cases, simple lemmas) that are co-occurring with those selected in the first circle and are included through an assessment of pertinence, still based on the semantics of the specific passages in which they appear in the Greek Bible. Among those included in the second circle, some roots carry a more sophisticated nuance than those on the first, like verbs indicating *ways* of killing or healing, such as follows:

Killing (circle 2)	**Healing (circle 2)**
ἀνάθεμα (22), ἀναθεματίζω (15) ἀναιρέω (83), ἀναίρεσις (4) ἀπόλλυμι (379) ἀφαιρέω (166) ἀφανίζω (91), ἀφανισμός (58) ἐκκεντέω (10) ἐξολεθρεύω (220), ὀλεθρεύω (22) θανατόω (160), ἀποθνήσκω (602), θάνατος (365), θνητός (9) κρεμάννυμι (37) λιθοβολέω (27) προσαποθνῄσκω (1)	θεραπεύω (26), θεράπων (64), θεράπαινα (8), θεραπεία (5) μοτόω (1)

Table 2

In the second circle, there are some cases of polysemy, which entails the need of disambiguation for a semantic tool; such is the case of ἀναιρέω (to kill, to lift), a verb that is quite frequent in the

[62] Those lemmas are not included in this article for reasons of space.

[63] Each entry is accompanied by the number of occurrences in the corpus, as they appear in TLG.

Greek Bible, but only sometimes in the context under investigation (66 times).[64] The polysemic root of θεραπεύω (to heal, to serve) is another case; only 8 occurrences of the verb are pertinent to the semantic sphere of healing, whereas among the words derived from this root, θεράπων (servant), θεράπαινα (female servant), and θεραπεία (servitude) appear in the corpus, yet none ever refers to the context of healing. Descending the scale of the frequency of pertinence, other lemmas are frequently associated with the ideas of killing and healing, but not necessarily, such as weapons, drugs, and blood. Those are included in the third circle, as follows:

Killing (circle 3)	**Healing (circle 3)**
αἷμα (416)	ἄλγος (6), ἀλγηρός (3)
διασπασμός (1)	ἀορασία (6)
δικαιόω (53)	ἀρρωστία (14), ἀρρώστημα (5)
ἔξοδος (71)	ἀσθενής (24)
θάπτω (178)	ἀφή (69)
θῦμα (15), θυμίαμα (81), θυσία (396), θυσιάζω (41), θυσιαστήριον (437), θύω (137)	δίψα (16)
	ἕλκος (16)
κατασκάπτω (34)	θραῦσμα (17)
καταφθείρω (27)	καθαρίζω (125)
μάχαιρα (191)	κατάκαυμα (10)
ξίφος (16)	κνήφη (1)
ὄλεθρος (24)	λέπρα (39), λεπρός (14)
παραδειγματίζω (6)	λεύκωμα (11)
πατάσσω (434)	μυρεψός (9)
ῥομφαία (259)	μώλωψ (9)
σίδηρος (74)	νόσος (11)
συντέλεια (88)	ὀδύνη (69)
συντρίβω (241)	παραπληξία (1)
τελευτάω (92)	πληγή (96)
τιτρώσκω (20)	συντριβή (25), σύντριμμα (35)
τραύμα (17)	τόξευμα (13)
	φάρμακον (13)
	ψώρα (3)

Table 3

The map is only an initial set of data to follow the distribution of a root, word or motive across the corpus of Greek Scriptures, which might enlighten the workshop of the translators through the examination of their lexical choices in light both of the lexical material at their disposal and of the possibilities of word formation and coinage. The map only gathers the lexical material that can be found by means of reading the textual occurrences in context and it is a static depiction of textual material, which lacks depth in many senses; (a) the words are shown out of context and their relations of co-occurrence are not accessible; (b) they are considered as they appear in the corpus, with no reference to the semantic evolution they may undergo within the corpus and in relation to extra-biblical sources; (c) they are not linked to the textual source in Hebrew and its semantics; (d) they are fixed as they appear in the critical edition, with no indication of variant readings and other

[64] Disambiguation remains a common question of semantic digital tools even for languages with more sophisticated and integrated digital systems; see Marco Passarotti et al., «Interlinking through Lemmas. The Lexical Collection of the LiLa Knowledge Base of Linguistic Resources for Latin», *Studi e Saggi Linguistici* 51/1 (2020), 177-212, esp. 196.

text-historical details. These in-depth considerations are currently possible only through careful and slow analysis of textual data in printed editions. A search engine that may be able to detect different semantic levels (linguistic affinity in terms of root, synonymity, topic, and symbolism) and connect passages based on textual similarity would be extremely helpful in the task of mapping a semantic sphere such as those presented here. Such tool would provide the ground not only for lexical and broader semantic research of a single text, but also for a possible expansion on the intertextual level, both within the corpus and in relation to a textual elsewhere.

4.2. The Role of Variant Readings

The occurrences listed above are the results of a TLG search, that only provides the evidence from one single eclectic edition (deprived of its apparatus), that is no longer the reference edition for many books of the Greek Bible.[65] Each printed critical edition of any text has a precise scope in reconstructing a phase of the text, usually the oldest recoverable. Regardless of the breadth of apparatus details, critical editions give a static impression of the status of one text at a given time.[66] What would be a *desideratum* in the study of intertextuality though is access to the history of the tradition in all the scraps that have survived, which normally constitute the bulk of the apparatus material. One of the aims of the project "Resilient Septuagint" is to consider the text of the Septuagint as a deep and granular entity; each word of the text is not only how it looks like in the printed edition, but rather contains its variant readings, which represent textual material additional to the hypotheses of the edited text and concern different stages and distributions across time and space. In research on semantics, the role of variant readings is to concretely represent the "explosion" of the tradition in textual possibilities, which may have been known to ancient readers of the Greek Bible and thus provide us with information on its reuse. The editors of the Göttingen Septuagint critical editions are aware of these theoretical principles, yet they engage with a huge amount of textual material, of which the necessities of the printed page can account for only a selected part. In the case of the Septuagint, the tradition is further complicated by ancient revisors of the translation, who created parallel versions of the text, contributing to its intrinsic plurality in antiquity.[67] Many of the books edited by the series of the Göttingen Septuagint are accompanied by two distinct apparatuses, (a) of textual variants in the Septuagint tradition; (b) of textual variants from ancient revisions, especially those by Aquila, Symmachus, and Theodotion. This distribution of textual data on the page sees this hexaplaric material as forcibly subjected to the tyranny of the Septuagint tradition, and the tradition subjected to the edited text in its turn, in an abstract hierarchy that does not represent the diffusion of the text in antiquity[68]. In the quest for a complete map of lemmas related to a semantic sphere in the Greek Bible, it would be fair to include the evidence of the apparatuses in the picture. To name one example of how helpful this would be in relation to the semantic sphere of "killing", a passage from Gen 2:17 may suffice. The verse presents

---

[65] The searchable edition in the TLG database is A. Rahlfs, R. Hanhart (eds.), *Septuaginta: Id est Vetus Testamentum graece iuxta LXX interpretes. Editio altera* (Stuttgart: Deutsche Bibelgesellschaft, 2006). 26 volumes have appeared so far in the Göttingen edition: R. Hanhart, J.W. Wevers, J. Ziegler et al. (eds.), *Septuaginta: Vetus Testamentum Graecum auctoritate Academiae Scientiarum Gottingensis editum* (Göttingen: Vandenhoeck & Ruprecht, 1931-).

[66] Ref. to Alpi in the same volume

[67] These materials, in a fragmentary state, are gathered together only in F. Field (ed.), *Origenis Hexaplorum quae supersunt; sive, Veterum Interpretum Graecorum in totum Vetus Testamentum fragmenta*, 2 vols. (Hildesheim: Olms, 1964).

[68] A critical edition of the fragments of Origen's *Hexapla* is in preparation within *The Hexapla Project* (Text and Canon History at Phoenix Seminary). Only the first volume of the series has been published: John Meade, *A Critical Edition of the Hexaplaric Fragments of Job 22–42* (Leuven: Peeters, 2020).

two lemmas that fall within the analyzed vocabulary, ἀποθνήσκω (to die) and θάνατος (death). The second apparatus, however, indicates that among the revisions a different reading is attested, which presents a further lemma, θνητός (mortal). This occurrence of a rare word is not included among the 9 mentioned above, which derive from a search performed only on the edited text (cf. Par 1, Table 2). The dataset called EnGraSept (*Enriched Granular Septuagint*), which is being developed by the team with the materials offered by the apparatuses of the Göttingen edition, appears as a necessary digital tool for semantic research on the text of the Greek Bible, aiming at deepening the dimension of every word in the corpus via the textual material that has been transmitted by tradition and assessed by philological work.

The research questions considered in the project pertain to the interaction between philological, linguistic, and historical research on the Septuagint text in relation to the possibilities opened by digital environments; each research question leads to a constant reframing from the point of view of human epistemology. The different visualization processes made possible by new digital tools represent an exciting new path for philological and exegetical research on Biblical texts and their heritage; a vigilance on the methodological assumptions that precede the training of digital tools for semantic analysis in the field is obviously crucial, as well as it is necessary to validate its results. What is clear is that as scholars we have a chance to participate in this new phase of the tradition of translations and heritage of the Bibles, one in which their texts and legacies are transferred into a new writing environment; how we navigate this transition will probably affect access to these traditions in the future.

4.3. "The New Frontiers of the Nouvelle Vague": (Large) Language Models

Large Language Models (LLMs) and generative AI have significantly transformed the field of computational humanities, offering powerful tools for text analysis, interpretation, and generation. These models, trained on vast *corpora* of texts, rely on deep learning architectures, particularly transformer-based neural networks, to predict and generate coherent sequences of text. Technologies such as OpenAI's GPT series, Google's BERT, and Meta's LLaMA have demonstrated exceptional performance in tasks ranging from question-answering to machine translation, showcasing the power of attention mechanisms and contextual embeddings in understanding language. Recent research shows that the scholarly interest in ML for ancient languages has significantly increased in the last 5 years.[69]

LLMs function by leveraging probabilistic modeling, meaning they predict the most statistically likely continuation of a given prompt based on patterns learned during training. This allows them to generate fluid and contextually relevant text but also introduces fundamental limitations. Since their outputs are inherently stochastic rather than deterministic, they can struggle with precise recall of factual information or exact textual references. Moreover, their effectiveness depends heavily on the availability of high-quality training data. For classical languages such as Greek and Latin, existing *corpora* are relatively limited compared to modern languages, restricting the models' ability to generalize accurately in these domains.

A key component in modern AI-driven text analysis is the use of vector-based semantic embeddings. These embeddings transform words, phrases, or even entire texts into numerical vectors in a high-dimensional space, capturing semantic relationships based on contextual similarities rather than mere lexical overlap. Techniques such as Word2Vec, GloVe, and more advanced transformer-based models like FastText and Sentence-BERT enable nuanced comparisons

[69] See Sommerschield/Assael/Pavlopoulos, "Machine Learning for Ancient Languages".

by mapping semantically related terms closer together in vector space. This capability is particularly useful for identifying implicit textual connections, paraphrases, and thematic correspondences that traditional string-matching algorithms might overlook.

However, vector embeddings come with their own limitations. While they excel at identifying conceptual similarities, they can sometimes introduce false positives due to their reliance on statistical associations rather than direct textual evidence. Additionally, for languages with limited digital resources, such as ancient Greek and Latin, pre-trained embedding models may lack sufficient granularity, necessitating fine-tuning on specialized *corpora* to achieve meaningful results.

For this reason, we advocate for a hybrid research approach that leverages both classical computational techniques and modern AI-driven methodologies. Not only the synsets of Ancient Greek WordNet,[70] but also, more broadly, traditional text alignment algorithms and string-matching techniques are indispensable for identifying explicit quotations and near-verbatim references with certainty. At the same time, vector-based semantic embeddings offer a complementary perspective, enabling the discovery of more nuanced, allusive intertextual relationships that might escape strict textual comparison methods.

This integrated approach finds concrete implementation in tools such as ElasticSearch, which combines robust text indexing and retrieval mechanisms with vector search capabilities. By harnessing the strengths of both methodologies, we aim to build a more reliable and nuanced system for textual analysis — one that balances the rigor of deterministic approaches with the flexibility of probabilistic AI models. In this way, we can maximize both precision and interpretative depth, ensuring that computational tools serve the intricate needs of humanities research rather than merely approximating them.

4.4 Our proposal

In light of the scenario outlined — its potential as well as its limitations — our research group, in collaboration with the uBIQUity team, has developed a proposal for a token-based numerical regression system that, on the one hand, encapsulates the outcomes and spin-offs of humanistic research, and, on the other, lays the groundwork for a workflow which could be followed by a search engine too.[71] This system combines different annotation models based on binary logic that is compatible with both language-aware and language-agnostic approaches typical of traditional algorithmic computational linguistics. Its goal is to account for the distance between strings in a way that can be expressed numerically, starting from the philological complexity inherent in the biblical textual tradition. Above all, we sought to devise a solution that would combine the mathematical certainty provided by traditional algorithms with the resources of vector semantics, pending the development of satisfactory technologies.

While our annotation system is token-based, thanks to ElasticSearch, text reuse searches are easily interoperable with state-of-the-art search methods (*lemma-based*, *n-gram-based*, etc.).[72] We

[70] See G. Scatigno, "Analysis of Anachronistic Lemmas and Semantic Fields in Ancient Greek WordNet", in *13th Global WordNet Conference (GWC2025)*, in press.

[71] The Resilient Septuagint team (see above) cooperates closely with the uBIQUity team lead by Anna Mambelli, with which we share the methodological framework and the interoperable dataset for the study of biblical texts in Ancient Greek and their heritage (for details see the essay by A. Mambelli in this volume).

[72] See F. Peng/D. Schuurmans/W. Keselj/S. Wang, "Language Independent Authorship Attribution with Character Level N-Grams", in *10th Conference of the European Chapter of the Assoociation for Computational Linguistics* (Budapest, Association for Computational Linguistics, 2003), 267-274.

have adopted the biblical verse as our atomic unit — a conventional choice but based on the practice of biblical reuse — with the aim of reconciling the human approach to text, characterized by precise procedures,[73] with that of the machine, facilitated by short character sequences.[74] From this point of view, the INCEpTION platform, developed between 2017 and 2022 by the Technische Universität in Darmstadt,[75] seemed the ideal environment in which to operate. It allowed us to prepare a small patristic *corpus* that is semantically annotated, i.e., metadated in relation to its reuses, with all the information that can be obtained from the critical apparatus.

Depending on the number of annotations collected, it can serve as a training set for the models that our project engineers are currently developing based on the data we provide and in line with our needs or as a benchmarking system. In both cases, it allows us to annotate without sacrificing the complexity of the philological dimension of biblical research, but above all, by ensuring that this very complexity provides the sources or parameters for training models or, in complex systems, improves the efficiency of existing models.

As for datasets, EnGraSept is the solution we designed to associate each token in the edited text of the Göttingen critical edition of the Septuagint with its variants in the apparatus. Through its naming (which refers to an "enriched granularity"), we wanted to credit Roncaglia's insight into the possibility of working with atoms that are structured and articulated within themselves—and it is built using the same methodology as the training set, which brings together published texts and their apparatus. In this case, however, we use ontologies that are already available. The aim, first and foremost, is to make the pages of the Göttingen edition of the LXX parsable by machine so that at a later stage, the software can automatically generate the biblical versions preserved by each witness and, consequently, a dictionary of variants that can be indexed and managed by ElasticSearch.

In this sense, what we are building is not just a technical infrastructure, but a conceptual framework capable of sustaining future research at the intersection of philology and artificial intelligence. The challenge, then, is not how to make machines read like humans, but how to make them see and discover what only humans—and only those equipped with the highest-level knowledge base—have so far been able to see. We believed that the most effective way to pursue this goal was to design a system in which the machine follows the same workflow that underpins humanistic methodologies.

What is ultimately at stake, it seems to us, is the human role itself, which partly consists in validating the machine's outputs and partly in coordinating its operations. But in this case, not merely as the old-style programmer — an all-too-familiar risk for today's digital humanist — but through a new and broader form of collaboration between digital humanists, who design the linguistic tools, and humanists outside the digital domain, whose contributions enrich the very notions of 'semantics' and of 'historical source'.

Bibliography

Aejmelaeus, A., "Faith, Hope, and Interpretation: A Lexical and Syntactical Study of the Semantic Field of Hope in the Greek Psalter", in VanderKam, J.C./Flint, P.W./Tov, E. (ed.), *Studies in the Hebrew Bible, Qumran, and the Septuagint Presented to Eugene Ulrich* (Leiden/Boston: Brill, 2006) 360–376.

[73] See A. Compagnon, *La seconde main ou le travail de la citation* (Paris: Édition du Seuil, 1979), 15-45.

[74] See Roncaglia, *L'architetto e l'oracolo,* 107-108.

[75] As for the INCEpTION platform, see the contribution by A. Mambelli in this volume.

Agirre, E./López de Lacalle, O./Soroa, A., “Random Walks for Knowledge-based Word Sense Disambiguation”, *Computational Linguistics* 40 (2014) 57–84.

Aitken, J.K., “Lexical Semantics and the Cultural Context of Knowledge in Job 28, Illustrated by the Meaning of ḥāqar”, in E. van Wolde (ed.), *Job 28. Cognition in Context* (Leiden/Boston: Brill, 2003) 119–137.

Andrews, T.L./Macé, C. (eds.), *Analysis of Ancient and Medieval Texts and Manuscripts: Digital Approaches* (Turnhout: Brepols, 2014) 183–244.

Archer, D., “Digital Humanities 2006: When Two Became Many”, *LLC* (2007) 103–108.

Bader, W. (ed.), *Und die Wahrheit wurde hinweggefegt: Daniel 8 linguistisch interpretiert* (Tübingen: Franke, 1994).

Bamman, D./Crane, G., “The Logic and Discovery of Textual Allusion,” in *Proceedings of the Second Workshop on Language Technology for Cultural Heritage Data (LaTeCH 2008)*, http://www.lrec-conf.org/proceedings/lrec2008/workshops/W22_Proceedings.pdf#page=31.

Bamman, D./Mambrini, F./Crane, G., “An Ownership Model of Annotation: The Ancient Greek Dependency Treebank”, in M. Passarotti/A. Przepiórkowski/S. Raynaud et al. (ed.), *Proceedings of the Eighth International Workshop on Treebanks and Linguistic Theories* (Milan: EDUCatt, 2009) 5–16.

Barnard, D.T./Fraser, C.H./Logan, G.M., “Generalized Markup for Literary Texts”, *LLC* 3 (1988) 26–31.

Barr, J., *The Semantics of Biblical Language* (Oxford: Oxford University Press, 1961).

Bennett, W.S., “The Place of Semantics in MT Systems,” *LLC* 4 (1989) 200–202.

Bickerman, E.J., “Notes on the Greek Book of Esther”, *Proceedings of the Academy for Jewish Research* 20 (1951) 101–133.

Blackburn, P./Bos, J., *Representation and Inference for Natural Language. A First Course in Computational Semantics* (Stanford: CSLI, 2005).

Blei, D.M./Ng, A.Y./Jordan, M.I., “Latent Dirichlet Allocation”, *Journal of Machine Learning Research* 3 (2003) 993–1022.

Bons, E. (ed.), *The Historical and Theological Lexicon of the Septuagint. Volume I Alpha–Gamma* (Tübingen: Mohr Siebeck, 2020).

Bons, E./Mambelli, A./Scialabba, D. (ed.), *Joy Denied, Joy Rediscovered: From Classical Greek Literature to Byzantine Christianity* (Leiden/Boston: Brill, forthcoming 2025).

Büchler, M., *Informationstechnische Aspekte des Historical Text Re-use* (Leipzig: Universität Leipzig, 2013).

Burdick, A./Drucker, J./Lunefeld, P./Presner, T./Schnapp, J., *Digital Humanities* (Cambridge MA/London: The MIT Press, 2012).

Buzzetti, D., “Alle origini dell’informatica umanistica. Humanities Computing e/o Digital Humanities?”, *Rendiconti Lincei – Scienze Morali Storiche e Filologiche* 9–30 (2019) 71–103.

Buzzetti, D., “Prefazione. Oltre il limite istituzionale”, in F. Ciotti (ed.), *Digital Humanities. Metodi, strumenti, saperi* (Roma: Carocci, 2023) 15–18.

Ciotti, F., “Introduzione. La galassia delle Digital Humanities”, in F. Ciotti (ed.), *Digital Humanities. Metodi, strumenti, saperi* (Roma: Carocci, 2023) 19–23.

Clivaz, C., “The Bible in the Digital Age: Multimodal Scriptures in Communities”, in T. Hutchings/C. Clivaz (ed.), *Digital Humanities and Christianity. An Introduction* (Berlin/Boston: De Gruyter, 2021) 21–46.

Compagnon, A., *La seconde main ou le travail de la citation* (Paris: Édition du Seuil, 1979) 15–45.

Dainese, D./Mambelli, A., “Intertestualità tra Bibbie e antichi commentary cristiani: l’esempio di simul nel De Genesi ad litteram di Agostino”, *Lexicum Philosophicum* 11 (2023–2024) 39–65.

Damian, I./de Almeida, M./de Mello, T./Rodrigues, P., “Convergências entre as Humanidades Digitais e a Ciência da Informação”, *Ibersid* 9 (2015) 79–82.

Evans, T.V., “The Nature of Septuagint Greek. Language and Lexicography”, in Salvesen, A.G./Law, T.M. (ed.), *The Oxford Handbook of the Septuagint* (Oxford: Oxford University Press, 2021) 91–104.

Fellbaum, C. (ed.), *WordNet. An Electronic Lexical Database* (London: MIT Press, 1998).

Field, F. (ed.), *Origenis Hexaplorum quae supersunt; sive, Veterum Interpretum Graecorum in totum Vetus Testamentum fragmenta*, 2 vols. (Hildesheim: Olms, 1964).

Franzini, G./Peverelli, A./Ruffolo, P. et al., “Nunc Est Aestimandum. Towards an Evaluation of the Latin WordNet”, in *Proceedings of the Sixth Italian Conference on Computational Linguistics* (Turin: Associazione Italiana di Linguistica Computazionale, 2019), [https://dx.doi.org/10.5281/zenodo.3518774].

Gardin, J., “Points de vue logicisties sur les méthodologies en sciences sociales”, *Sociologie et sociétés* 25 (1993) 11–22.

Gillmayr-Bucher, S., “A Computer-aided Quest for Allusions to Biblical Texts within Lyric Poetry”, *LLC* 1 (1996) 1–8.

Gillmayr-Bucher, S., “Intertextualität. Zwischen Literaturtheorie und Methodik”, *PzB* 8 (1999) 5–20.

Gorman, V.B., “Dependency Treebanks of Ancient Greek Prose”, *Journal of Open Humanities Data* 6 (2020), https://doi.org/10.5334/johd.13.

Hamidovic, D., “An Introduction to Emerging Digital Culture”, in C. Clivaz/P. Dilley/D. Hamidović (ed.), *Ancient Worlds in Digital Culture* (Leiden/Boston: Brill, 2016) 1–16.

Haney, W.S., *Cyberculture, Cyborgs and Science Fiction: Consciousness and the Posthuman* (Amsterdam/New York: Rodopi, 2006).

Hanhart, R./Wevers, J.W./Ziegler, J. et al. (ed.), *Septuaginta: Vetus Testamentum Graecum auctoritate Academiae Scientiarum Gottingensis editum* (Göttingen: Vandenhoeck & Ruprecht, 1931–).

Harrington, M., "Perseids Project – Treebanked Commentaries at Tufts University", https://perseids-project.github.io/harrington_trees/.

Hartman, D., *Emozioni nella Bibbia. Lessico e passaggi semantici tra Bibbia ebraica e LXX* (Napoli: L'Orientale Università degli Studi – Napoli, 2019).

Haug, D.T.T./Jøhndal, M.L., "Creating a Parallel Treebank of the Old Indo-European Bible Translations", in C. Sporleder/K. Ribarov (ed.), *Proceedings of the Second Workshop on Language Technology for Cultural Heritage Data (LaTeCH 2008)*, http://www.lrec-conf.org/proceedings/lrec2008/workshops/W22_Proceedings.pdf#page=31.

Hohli Trillini, R./Quassdorf, S., "A 'Key to All Quotations'? A Corpus-based Parameter Model of Intertextuality", *LLC* 25 (2010) 269–286.

Horstmann, J./Lück, C./Normann, I., "Systems of Intertextuality: Towards a Formalization of Text Relations for Manual Annotation and Automated Reasoning", *DHQ* 17 (2023) 1–74.

Keersmaekers, A., "The GLAUX Corpus: Methodological Issues in Designing a Long-Term", in N. Tahmasebi/A. Jatowt/Y. Xu et al. (ed.), *Proceedings of the 2nd International Workshop on Computational Approaches to Historical Language Change* (New York: Curran, 2021) 39–50.

Keersmaekers, A./Mercelis, W./Swaelens, T./Van Hal, T., "Creating, Enriching and Valorising Treebanks of Ancient Greek: the Ongoing Pedalion-project", in M. Candito/K. Evang/S. Oepen et al. (ed.), *Proceedings of the 18th International Workshop on Treebanks and Linguistic Theories* (Paris: Association for Computational Linguistics, 2019) 109–117.

Keersmaekers, A./Mercelis, W./Van Hal, T., "Word Sense Disambiguation for Ancient Greek: Sourcing a Training Corpus through Translation Alignment", in *Proceedings of the Ancient Language Processing Workshop* (Shoumen: Incoma, 2023) 148–159.

Keersmaekers, A./Van Hal, T., "A Corpus-Based Approach to Conceptual History of Ancient Greek", in G. Kristiansen/K. Franco/S. De Pascale et al. (ed.), *Cognitive Sociolinguistics Revisited* (Berlin/Boston: Brill, 2022) 213–225.

Kirschenbaum, M., "What Is Digital Humanities and What's It Doing In English Departments", in M.K. Gold, *Debates in the Digital Humanities* (Minneapolis/London: University of Minnesota Press, 2012) 3–11.

Koltay, T., "Library and Information Science and the Digital Humanities: Perceived and Real Strengths and Weakness", *Journal of Documentation* 72 (2016) 781–792.

Köntges, T., "Measuring Philosophy in the First Thousand Years of Greek Literature", *Digital Classics Online* 6 (2020), https://doi.org/10.11588/dco.2020.2.73197.

Li, W./Wang, S./Yu, Z., "Deep Learning and Semantic Concept Space Are Used in Query Expansion", *ACCS* 52 (2018) 175–183.

Lu, M./Sun, X./Wang, S. et al., “Query Expansion via WordNet for Effective Code Search”, in *2015 IEEE 22nd International Conference on Software Analysis, Evolution, and Reengineering (SANER)* (Washington DC: IEEE, 2015) 545–549.

McGillivray, B./Hengchen, S./Lähteenoja, V./Palma, M./Vatri, A., “A Computational Approach to Lexical Polysemy in Ancient Greek”, *DSH* 34 (2019) 893–907.

Meade, J., *A Critical Edition of the Hexaplaric Fragments of Job 22–42* (Leuven: Peeters, 2020).

Melacci, S./Globo, A./Rigutini, L., “Enhancing Modern Supervised Word Sense Disambiguation Models by Semantic Lexical Resources”, in N. Calzolari/K. Choukri/C. Cieri et al. (ed.), *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)* (Miyazaki: ELRA, 2018) 1012–1017.

Merkin, R./Busharia, Z./Meir, E., “The Historical Dictionary of the Hebrew Language”, *LLC* 4 (1989) 271–273.

Merriam, T., “Intertextual Distances, Three Authors”, *LLC* 18 (2003) 379–388.

Mitkov, R. (ed.), *The Oxford Handbook of Computational Linguistics* (Oxford: Oxford University Press, 2003).

Ooi, V.B.Y., *Computer Corpus Lexicography. Edinburgh Textbooks in Empirical Linguistics* (Edinburgh: Edinburgh University Press, 1998).

Orlandi, T./Tomasi, F., “Una storia dell’informatica umanistica in Italia”, in F. Ciotti (ed.), *Digital Humanities. Metodi, strumenti, saperi* (Roma: Carocci, 2023) 35–47.

Orlandi, T./Tomasi, F., “Una storia dell’informatica umanistica in Italia”, in F. Ciotti (ed.), *Digital Humanities. Metodi, strumenti, saperi* (Roma: Carocci, 2023) 35–47.

Packard, D.W., “Computer-Assisted Morphological Analysis of Ancient Greek”, in A. Zampolli/N. Calzolari (ed), *Computational and Mathematical Linguistics: Proceedings of the International Conference on Computational Linguistics* (Firenze: L.S. Olschki, 1977–1980) 343–355.

Pal, A.R./Saha, D., “An Approach to Automatic Text Summarization Using WordNet”, in *2014 IEEE International Advance Computing Conference (IACC)* (Washington DC: IEEE, 2014) 1169–1173.

Palladino, C./Farimah, K./Mathiak, B., “NER on Ancient Greek with Minimal Annotation”, https://dh2020.adho.org/.

Parry, D., “The Digital Humanities or a Digital Humanism”, in M.K. Gold (ed.), *Debates in the Digital Humanities* (Minneapolis/London: University of Minnesota Press, 2012) 429–437.

Passarotti, M. et al., “Interlinking through Lemmas. The Lexical Collection of the LiLa Knowledge Base of Linguistic Resources for Latin”, *Studi e Saggi Linguistici* 51/1 (2020) 177–212.

Passoni Dell’Acqua, A., “Notazioni cromatiche dell’Egitto greco-romano. La versione dei LXX e i papiri”, *Aegyptus* 78 (1998) 77–115.

Peng, F./Schuurmans, D./Keselj, W./Wang, S., “Language Independent Authorship Attribution with Character Level N-Grams”, in *10th Conference of the European Chapter of the Association for Computational Linguistics* (Budapest: Association for Computational Linguistics, 2003) 267–274.

Perrone, V./Palma, M./Hengchen, S. et al., “GASC: Genre-Aware Semantic Change for Ancient Greek”, in N. Tahmasebi/L. Borin/A. Jatowt et al. (ed.), *Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change* (Berlin: Language Science Press, 2021) 56–66.

Piotrowski, D. (ed.), *Lexicographie et Informatique: Autour de l’informatisation du Trésor de la Langue Française* (Paris: Didier Érudition, 1998).

Rahlfs, A./Hanhart, R. (ed.), *Septuaginta: Id est Vetus Testamentum graece iuxta LXX interpretes. Editio altera* (Stuttgart: Deutsche Bibelgesellschaft, 2006).

Riemenschneider, F./Frank, A., “Exploring Large Language Models for Classical Philology”, in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics* (New York: Curran, 2023) vol. 1, 15181–15199.

Romanello, M., “Exploring Citation Networks to Study Intertextuality in Classics”, *DHQ* 10 (2016) 1–46.

Roncaglia, G., *L’architetto e l’oracolo. Forme digitali del sapere da Wikipedia a ChatGPT* (Roma/Bari: Laterza, 2023) 5–21.

Roncaglia, G., *L’età della frammentazione. Cultura del libro e scuola digitale* (Roma/Bari: Laterza, 2018).

Ross, C., “Social Media for Digital Humanities and Community Engagement”, in C. Warwick/M. Terras/J. Nyhan (ed.), *Digital Humanities in Practice* (Cambridge: Cambridge University Press, 2012).

Scatigno, G., “Analysis of Anachronistic Lemmas and Semantic Fields in Ancient Greek WordNet”, in *13th Global WordNet Conference (GWC2025)*, in press.

Scheirer, W./Forstall, C./Coffee, N., “The Sense of a Connection: Automatic Tracing of Intertextuality by Meaning”, *DSH* 31 (2017) 204–217.

Schnapp, J./Presner, T., *Digital Humanities Manifesto 2.0* (2009), https://jeffreyschnapp.com/wp-content/uploads/2011/10/Manifesto_V2.pdf.

Silva, M., *Biblical Words and their Meaning. An Introduction to Lexical Semantics* (Grand Rapids: Zondervan, 1994) 17–34.

Sommerschield, T./Assael, Y./Pavlopoulos, J., “Machine Learning for Ancient Languages: A Survey”, *Computational Linguistics* 49 (2023) 703–747.

Spinaci, G./Colavizza, G./Speroni, S., “A Map of Digital Humanities Research Across Bibliographic Data Sources”, *DSH* 37 (2022) 1254–1268.

Spiro, L., “This Is Why We Fight: Defining the Values of the Digital Humanities”, in M.K. Gold (ed.), *Debates in the Digital Humanities* (Minneapolis/London: University of Minnesota Press, 2012) 16–35.

Stanchev, L., “Semantic Document Clustering Using Information from WordNet and DBPedia”, in *2018 IEEE 12th International Conference on Semantic Computing (ICSC)* (Washington DC: IEEE, 2018) 100–107.

Stopponi, S./Pedrazzini, N./Peels, S./McGillivray, B./Nissim, M., “Evaluation of Distributional Semantic Models of Ancient Greek: Preliminary Results and a Road Map for Future Work”, in *Proceedings of the Ancient Language Processing Workshop* (Shoumen: Incoma, 2023) 49–58.

Tomazzoli, G., “Intertextuality in Dante’s Commedia: Hypermedia Dante Network”, *Bibliotheca Dantesca* 5 (2022) 308–311.

Tov, E., *A Computerized Database for Septuagint Studies: The Parallel Aligned Text of the Greek and Hebrew Bible* (Stellenbosch: JNSL, 1986).

Tov, E., *The Greek and Hebrew Bible. Collected Essays on the Septuagint* (Leiden/Boston/Köln: Brill, 1999).

Vatri, A./McGillivray, B., “Lemmatization for Ancient Greek. An Experimental Assessment of the State of the Art”, *Journal of Greek Linguistics* 20 (2020) 179–196.

Vatri, A./McGillivray, B., “The Diorisis Ancient Greek Corpus. Research Data”, *Journal for the Humanities and Social Sciences* 3 (2018) 55–65.

Vierros, M./Henriksson, E., “PapyGreek Treebanks: A dataset of linguistically annotated Greek documentary papyri”, *Journal of Open Humanities Data* 7 (2021), https://doi.org/10.5334/johd.55.

Wang, C./Song, Y./El-Kishky, A. et al., “Incorporating World Knowledge to Document Clustering via Heterogeneous Information Networks”, in *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York: Association for Computing Machinery, 2015) 1215–1224.

Xie, N./Li, S./Ren, H. et al., “Abstractive Summarization Improved by WordNet-Based Extractive Sentences”, in *CCF International Conference on NLP and Chinese Computing* (Cham: Springer, 2018) 404–415.

Yousef, T./Palladino, C./Jänicke, S., “Transformer-Based Named Entity Recognition for Ancient Greek”, http://dx.doi.org/10.13140/RG.2.2.34846.61761.

Zadeh, L.H., “Text-Score Semantics as a Basis for Computational Approach to the Representation of Meaning”, *LLC* 1 (1986) 24–35.