

The Cultural Evolution of Prototypical Paul in the First Five Centuries: A Distributional Semantic Analysis of Greek Christian Texts¹

Nina Nikki

University of Helsinki
nina.nikki@helsinki.fi

Vojtěch Kaše

University of West Bohemia
kase@kfi.zcu.cz

Zdeňka Špiclová

University of West Bohemia
spiclovz@kfi.zcu.cz

DOI: <https://doi.org/10.25784/jeac.v4i0.1012>



Abstract

The article uncovers how the first five centuries of Greek Christian texts reflect a changing understanding of Paul as a prototypical character. The study joins cognitive and social psychological theory of categorization in defining prototype as a highly contextual, fuzzy set of qualities which captures the make-up of the ingroup in comparison to outgroups. Paul as a prototypical character thus illustrates a group-level logic and an attempt at group identity construction by the early Christian authors. Textual source material poses special challenges because, strictly speaking, prototypes exist only in human cognition and because ancient texts do not accurately represent the social reality behind them. The study approaches prototypical depictions of Paul with the help of distributional semantics models that consider language as, likewise, highly context-dependent and, importantly, allow access to a deeper “meaning” of Paul on a cognitive level beyond mere co-occurrence of words in a given text. The analysis shows that while Paul is in the first three centuries linked, among other things, to oral tradition, gentile mission, and Judaism, the 4–5th centuries connect him to church offices, anti-heretical struggle, and doctrinal content. Curiously, in the later period, Paul also becomes distanced from the figure of Jesus/Christ, perhaps because the latter became veiled in high Christological dogma.

¹ Nina Nikki’s work on this article has been funded by the Finnish Academy as a part of the project “Paul’s Ideas and the Ideas of Paul in Cultural Evolution”. Vojtěch Kaše’s contribution to this article has been funded by the Czech Science Foundation (GAČR) as a part of the project “The Cultural Evolution of Moralizing Religions in the Ancient Mediterranean: A Distant Reading Approach” (GA20-01464S). Zdeňka Špiclová has been supported by the Grant System of the University of West Bohemia within the project SGS-2022-019.

1. Previous scholarship on the reception of Paul in the first centuries CE

The Apostle Paul has had a pivotal influence on Christianity since antiquity.² There is no shortage of modern research aimed at understanding and tracing the reception of his character, his letters, and his thoughts in the first centuries CE.³ The different aspects of Pauline influence have been categorized variously. Daniel Marguerat, for instance, suggests that Paul's effect manifests in three main themes: his letters, his mission to the Gentiles, and his role as a teacher and guardian of orthodoxy.⁴ Some scholars have made a distinction between the reception of Paul as a person and the transmission of his theology,⁵ sometimes claiming that Paul's thinking became more important to subsequent theologians than his character.⁶ Others, such as Benjamin White, prefer to treat these aspects together.⁷ Discussing Paul's influence in the second century Christian culture, Joseph R. Dodson advises to inquire about the extent of Paul's influence, the way Paul is utilized and with what goals,⁸ and about the geographical and authorial range of the reception.⁹

Since at least the second century, a controversy has existed over who gets Paul right.¹⁰ This is tied to what White calls “an archaizing argument” (“We ought to be ‘x’ because Christianity in its earliest genius was ‘x.’”).¹¹ In academic research, this idea was already challenged in the 70s/80s narrative of “Pauline fragmentation,” which posited that “Paul's image, as well as the use of Pauline letters, [evolved] in different directions among a variety of often competing Christian communities, none of which had a monopoly on Paul.”¹² While nowadays

² Paul was heavily employed by a variety of late antique Christian influencers such as Marcion, Valentinus, Irenaeus, and Origen. Early modern scholars aware of the impact of Paul include Ferdinand Christian Baur and Adolf von Harnack.

³ For important works on Paul in the second century, see BENJAMIN L. WHITE, *Remembering Paul: Ancient and Modern Contests over the Image of the Apostle* (New York: Oxford University Press, 2014), 10–11; JOSEPH R. DODSON, “Introduction,” in *Paul and the Second Century*, ed. MICHAEL F. BIRD and JOSEPH R. DODSON (London: T&T Clark, 2011), 1–17. A recent important collection of essays on the early reception of Paul is SIMON BUTTICAZ, ANDREAS DETTWILER, and JENS SCHRÖTER, eds., *Receptions of Paul in Early Christianity: The Person of Paul and His Writings through the Eyes of His Early Interpreters* (Berlin: De Gruyter, 2018).

⁴ DANIEL MARGUERAT, “Paul après Paul: une histoire de réception,” *NTS* 54 (2008): 317–37. For the epistolary part, see the database led by Jennifer Strawbridge for quotations of NT letters attributed to Paul in the first four centuries <https://paulandpatristics.web.ox.ac.uk/> (Last access 14.11.2022) and JENNIFER R. STRAWBRIDGE, *The Pauline Effect* (Berlin: De Gruyter, 2015).

⁵ See WHITE, *Remembering Paul*, 17, for a list of conceptualizations of this difference as one between, e. g., “Bild” and “Theologie,” “legendary” and “epistolary,” “monument” and “document”.

⁶ WHITE, *Remembering Paul*, 46, points to how DAVID RENSBERGER concluded in his 1981 Yale dissertation *As the Apostle Teaches: The Development of the Use of Paul's Letters in Second-Century Christianity* that concerns over Paul's persona were of little interest in the second century.

⁷ WHITE, *Remembering Paul*, 81, insists that “[f]or scholars of early Pauline traditions, both aspects—text and image—must be held together and brought into dialogue in order to provide a thick description of Paul as a persona in the second century.”

⁸ This includes asking whether the appeal was pastoral, theological, argumentative, or personal, and if Paul was appealed to in questions concerning the role of women in the church, church hierarchy or taken up as a personal role model (DODSON, “Introduction,” 4–5).

⁹ While Paul was used by many authors, it is also necessary to ask why other works, such as *The Shepherd of Hermas*, *The Epistle of Barnabas*, *2 Clement*, were reticent about him (DODSON, “Introduction,” 4–8).

¹⁰ See already 2 Pet 3:15–16.

¹¹ WHITE, *Remembering Paul*, 9.

¹² WHITE, *Remembering Paul*, 48.

the debate over the “real” Paul interests religious groups more than academics, there is even in scholarship a continued interest in “the historical Paul,” whose portraits are published at a steady rate.¹³ But two obstacles conceal the historical, “real” Paul. The first one is the inevitable subjectivity of the researcher – a hermeneutical reality well known to modern scholars. The second one has to do with the elasticity and ambiguity of Paul’s own thinking as we encounter it in his letters.¹⁴ It was in fact this very polysemy that provided fodder for the earliest varied representations of Paul.¹⁵

The current article joins White in questioning the meaningfulness of emphasizing the “real” Paul.¹⁶ The image of Paul is, of necessity, ever-changing, and each author adapts a selection of Pauline themes that best suits them, blending them with other interesting or topical motifs. While White deals with the question from a purely humanities and interpretative standpoint, the current article brings to play a cultural evolutionary logic focused on the variation and transmission (via social learning) of cultural content. Importantly, cultural evolution posits no value judgment on cultural variation and change and considers them neither improvement nor corruption.¹⁷

Many studies have addressed the reception of Paul by providing close readings and comparisons of relevant texts. White discusses Irenaeus and 3 Corinthians in particular, while Gregory Sterling provides an insightful juxtaposition of Acts and Ephesians.¹⁸ The current article takes a different approach by analyzing large textual corpora by means of computational text analysis methods or “distant reading.” This makes it possible to obtain a global view of the content and structure of the examined texts and to compare text corpora in a controllable and replicable form. Distant reading methods are not designed or equipped to replace close reading, rather they can help support and enrich the researcher’s view achieved

¹³ According to WHITE, *Remembering Paul*, 11, the dependency on the ideology of the “real” Paul is also visible in the way modern scholarship relies on the authentic seven letters of Paul. An example of such an approach is the book *Paulus, der Gründer des Christentums*, where Gerd Lüdemann tries to find the historical Paul and reconstruct what he really thought, what he wanted and what he did. GERD LÜDEMANN, *Paulus, der Gründer des Christentums* (Springer: zu Klampen Verlag, 2014).

¹⁴ On the challenges inherent in studying the historical Paul, see NINA NIKKI, “Challenges in the Study of the Historical Paul,” in *Common Ground and Diversity in Early Christian Thought and Study: Essays in Memory of Heikki Räisänen*, ed. RAIMO HAKOLA, OUTI LEHTIPUU, and NINA NIKKI (Tübingen: Mohr Siebeck, 2022), 185–209.

¹⁵ See DODSON, “Introduction,” 7, for examples. The complexity in Paul’s thinking meant that “he could be idealized by a variety of reputational entrepreneurs.” WHITE, *Remembering Paul*, 105.

¹⁶ Even DODSON, “Introduction,” 10, is partly captive to this line of reasoning: “it is wise to discern whether or not these blended portrayals of Paul *faithfully* reflect his New Testament legacy.” (Emphasis his).

¹⁷ An invaluable introduction to cultural evolutionary theory is ALEX MESOUDI, *Cultural Evolution: How Darwinian Theory Can Explain Human Culture and Synthesize the Social Sciences* (Chicago: University of Chicago Press, 2011).

¹⁸ Sterling concludes that while both present Paul as a pivotal figure in early Christianity Ephesians focuses on Paul’s thinking (*vita contemplativa*) and Acts on Paul’s life (*vita activa*). GREGORY STERLING, “From Apostle to the Gentiles to Apostle of the Church: Images of Paul at the End of the First Century,” *Zeitschrift für die Neutestamentliche Wissenschaft und Kunde der Älteren Kirche* 99/1 (2008): 74–98.

by classical exegetical work.¹⁹ An important benefit is the possibility to correct pre-understandings and subjective views, which cannot be fully avoided in close reading and classical exegesis. Especially researchers who have been dealing with certain texts and topics for a long time can have much knowledge that naturally shapes their intuition on related issues. Computer analyses allow the researcher to look at the problem without these burdens and evaluate the validity of a hypothesis in a replicable form. The analysis also warns the researcher in case some essential aspect of the problem is in danger of being overlooked or one element in the interpretation is overloaded at the expense of another.

2. Paul as prototypical character

We posit that the descriptions of Paul in early Christian literature can be viewed meaningfully from the perspective of *prototypicality*. The concept of *prototype* originates in the study of cognitive *categorization*, that is, the process by which a person groups perceived stimuli in order to process them more efficiently. This grouping is done by *accentuating* differences between groups and *attenuating* differences between members within a group.²⁰ Henri Tajfel and A. L. Wilkes, who ran the first experiments on cognitive classification, already linked the perceptual phenomenon to *social stereotyping*, suggesting that categorization takes place in social environments as well.²¹ The shift from mere grouping of sensory stimuli to social categorization brings into play *ingroup favoritism* and various identity motives such as a need for self-esteem and belonging.²² These, in turn, have an effect on inter-group competition and ingroup cooperation, which have important ethical implications.²³

The so-called *meta-contrast principle* specifies categorization further. Building on the concepts of accentuation and attenuation, the principle predicts that “a collection of individuals tends to be categorized as a group to a degree *inter alia* that the perceived differences between them are less than the perceived differences between them and other people (outgroups) in

¹⁹ ZDENKA ŠPICLOVÁ and VOJTĚCH KAŠE, “Distant Reading of the Gospel of Thomas and the Gospel of John: Reflection of Methodological Aspects of the Use of Digital Technologies in the Research of Biblical Texts,” *Open Theology* 6/1 (2020): 423–39.

²⁰ The pioneering experiments are reported in HENRI TAJFEL and A. L. WILKES, “Classification and Quantitative Judgment,” *The British Journal of Psychology* 54 (1963): 101–14, 113: “An essential feature of stereotyping is that of exaggerating some differences between groups classified in a certain way, and of minimizing the same differences within such groups.”

²¹ TAJFEL and WILKES, “Classification and Quantitative Judgment,” 112–14. A goal of their study was “to show that evidence for the essential unity of judgment phenomena, social or physical, can be slowly accumulated and that, without denying the importance of individual differences, it is possible to attempt an understanding of seemingly varied phenomena in terms of the same general judgment principles.” TAJFEL and WILKES, “Classification and Quantitative Judgment,” 114. Later, the so-called minimal group paradigm experiments provided important additional evidence for this, HENRI TAJFEL et al. “Social Categorization and Intergroup Behaviour,” *European Journal of Social Psychology* 1 (1971): 149–78.

²² ALEXANDER S. HASLAM, *Psychology in Organizations: The Social Identity Approach*. 2nd ed. (London: Sage Publications, 2004), 18–19.

²³ Ethics in the sense of moral philosophy essentially depends on the same “species-typical cognitive capacities and emotional dispositions” that are at work in social, prosocial, and cooperative situations. ILKKA PYYSIÄINEN, “The Evolution of Morality and Religion,” in *Christianity and the Roots of Morality: Philosophical, Early Christian, and Empirical Perspectives*, ed. PETRI LUOMANEN, ANNE B. PESSI, and ILKKA PYYSIÄINEN (Leiden: Brill, 2017), 44–64, esp. 53.

the comparative context.”²⁴ Here, the concept of *prototypicality* comes into play: a highly prototypical group member represents a given category in relation to other categories. The higher the meta-contrast ratio embodied by the group member, the more prototypical she or he is.²⁵ Although in this article we limit our examination to sources that view Paul favorably and consider him a role model, the term prototype itself can be used of outgroups as well.²⁶ Importantly, prototypicality is relative and highly contextual: group members vary in their degree of prototypicality not only when compared to each other but also depending on contextual factors.²⁷

Many social psychologists as well as biblical scholars applying social psychology have habitually taken a *prototype* in a social situation to denote a *person*,²⁸ but this is not how the concept is understood in cognitive psychology. More accurately, a prototype is “a fuzzy set of attributes (perceptions, attitudes, feelings, and behaviors) that are related to one another in a meaningful way, and that capture similarities within the group and differences between the group and other groups.”²⁹ Fuzziness means that “there is no definite point at which one can say the category ends.”³⁰ For the current study, the idea of a prototype as an indistinct mix of interrelated qualities from different areas of life offers an opportunity to view Paul without artificial classifications such as person versus theology (see above). Concerning the interest of the current journal in ethical questions, it can be pointed out that while prototypes certainly prescribe “appropriate attitudes and behaviors”³¹ for ingroup members, explicitly ethical content is only one, contingent part of what distinguishes one group from another. In fact, the prototypical approach enables detecting the actual role of strictly ethical issues for group identity at a given time as it is expressed through the prototypical character of Paul.³²

²⁴ JOHN C. TURNER and KATHERINE J. REYNOLDS, “Self-Categorization Theory,” in *Handbook of Theories of Social Psychology*, vol. 2, ed. PAUL A. M. LANGE, ARIE W. KRUGLANSKI, and E. TORY HIGGINS (London: Sage, 2012), 399–417.

²⁵ As John Turner and Katherine Reynolds state, a person will be viewed as prototypical “to the degree that the perceived differences between that person and other ingroup members are less than the perceived differences between that person and outgroup members.” TURNER and REYNOLDS, “Self-Categorization Theory,” 404.

²⁶ Although the term stereotype is often reserved for outgroups. PETRI LUOMANEN, “Social Identity, Prototypes, and Exemplars in Gospel Narratives: Methodological Considerations,” in *From Text to Persuasion: Festschrift in Honour of Professor Lauri Thurén on the Occasion of his 60th Birthday*, ed. ANSSI VOITILA et al- (Helsinki: The Finnish Exegetical Society, 2021), 85–110, 89.

²⁷ As Luomanen notes, “when the comparative context changes, the prototype also changes,” which means that “each definition of a prototype should be accompanied by an ample description of all the relevant comparative contexts in that particular situation.” LUOMANEN, “Social Identity, Prototypes, and Exemplars,” 98.

²⁸ Philip Esler, pioneer in combining biblical studies and the SIA, defines a prototype as “a representation of a person thought to typify the group.” PHILIP F. ESLER, *Conflict and Identity in Romans: The Social Setting of Paul’s Letter* (Minneapolis: Fortress Press, 2003), 172–73. Similarly, the social psychologist Alexander Haslam: “an image of an ideal person who embodies [the group’s] character”. HASLAM, *Psychology in Organizations*, 45.

²⁹ MICHAEL HOGG, “Social Identity Theory,” in *Contemporary Social Psychological Theories*, ed. PETER J. BURKE (Stanford: Stanford Social Sciences, 2006), 111–36, 118.

³⁰ MARK HALLAHAN, “Prototypes,” in *Encyclopedia of Social Psychology*, 714–16, 715.

³¹ KARIANNE KARLSHOVEN and DEANNE N. DEN HARTOG, “Ethical Leader Behavior and Leader Effectiveness: The Role of Prototypicality and Trust,” *International Journal of Leadership Studies* 5/2 (2009): 102–20, esp. 102.

³² Ethical concerns have been addressed especially in the study of prototypicality in leadership, see ALEXANDER S. HASLAM, MICHAEL PLATOW, and STEPHEN REICHER, *The New Psychology of Leadership: Identity, Influence, and Power* (Hove: Psychology Press, 2011), 148–50, and KARLSHOVEN and DEN HARTOG, “Ethical Leader Behavior,” 102–20.

Prototypes are thus abstractions whereas the term *exemplar* should be reserved for real group members. Eliot Smith and Michael Zarate, for example, insist that both exemplars – understood as “rank and file members” of the group – and abstract prototypes are necessary for cognitive categorization and that experiences of real-life exemplars inform the formation of prototypes.³³ Neuroimaging studies support this distinction as it has been shown that humans process exemplars and prototypes in different regions of the brain, although sometimes simultaneously.³⁴ There is, however, again a discrepancy in the use of the term *exemplar* between cognitive and social psychology: the latter speaks of *exemplary members* in the sense of real-life individuals who optimally embody prototypical characteristics and are thus *representative* of the group.³⁵

Many caveats and unknowns necessarily remain when attempting to capture how Paul appears as a prototypical character in early Christianity. It must be kept in mind that the terms *exemplar* and *prototype* were “originally developed to describe on-line cognitive processes in the human mind, not textual phenomena.”³⁶ Strictly speaking, prototypes only exist in the mind, where they are constantly being adjusted according to the surrounding circumstances. The Paul of the text is an *exemplar* – not a historical figure but an individual textual encounter, as it were. On the one hand, the textual descriptions of Paul only capture approximations of the ideas in the author’s mind; on the other hand, they feed into the prototypes in the readers’ minds, who, in addition, evaluate their own likeness to them (self-categorization).³⁷ Our knowledge, however, is necessarily limited to the text as we cannot reach the minds of the authors nor the audience.

Inasmuch as the prototypes in the individual minds become consensual and shared, it is possible to speak of *group prototypes*.³⁸ Here again, we must admit a lack of knowledge of the historical comparative situations where the prototypes served to signal a high meta-contrast ratio (for example, the importance of the image of Paul as a martyr in times of persecution etc.).³⁹ It is also important to note that we cannot assume actual clear-cut communities behind early Christian texts, as many scholars have recently pointed out. Stanley Stowers, for example, insists that authors “under specific social conditions produce and interpret myths, not communities” and that the interests of these “mythmakers” cannot be conflated with “the collective minds and wills of communities and peoples.”⁴⁰ Rather than reflect existing group

³³ ELIOT R. SMITH and MICHAEL A. ZARATE, “Exemplar and Prototype Use in Social Categorization,” *Social Cognition* 8 (1990): 243–62, 248.

³⁴ LUOMANEN, “Social Identity, Prototypes, and Exemplars,” 96–97.

³⁵ LUOMANEN, “Social Identity, Prototypes, and Exemplars,” 94.

³⁶ LUOMANEN, “Social Identity, Prototypes, and Exemplars,” 86.

³⁷ LUOMANEN, “Social Identity, Prototypes, and Exemplars,” 101–104.

³⁸ LUOMANEN, “Social Identity, Prototypes, and Exemplars,” 101.

³⁹ LUOMANEN, “Social Identity, Prototypes, and Exemplars,” 102.

⁴⁰ STANLEY K. STOWERS, “Kinds of Myth, Meals, and Power: Paul and the Corinthians,” in *Redescribing Paul and the Corinthians*, ed. RON CAMERON and MERRIL P. MILLER (Atlanta: Society of Biblical Literature, 2011), 105–49, 110. Similarly, LUOMANEN, “Social Identity, Prototypes, and Exemplars,” 103: “The ideas that the authors propagate may sometimes become too easily identified with consensual group norms.” Stowers points out that while Paul, for example, wanted the Corinthians to form a group and “held a theory saying that God had miraculously made them into a community ‘in Christ’”, the Corinthians “never did sociologically form a community.” STOWERS, “Kinds of Myth, Meals, and Power,” 109.

situations, the authors are also more likely to act as entrepreneurs of identity,⁴¹ projecting unto Paul their hopes about what an ideal community would look like and which characteristics would optimally differentiate it from other groups. This means that Paul is not merely an exemplar in the general sense, but *an exemplary and thus prototypical representative of the group* – real, imagined, or intended.

The concept of a literary, prototypical Paul connects to the theory of cultural evolution and cognitive science. We treat Paul as an idea or mental representation within a changing web of other mental representations which are subjects of cultural transmission. Individual features of the mental representations differ from individual to individual, since cultural transmission is always a transformative process.⁴² We assume that this transmission can be described by the term *convergent transformation*, where “one item causes the production of another item whose form tends to deviate from that of the original item in a nonrandom way.”⁴³ This nonrandomness is related to the fact that convergent transformation is understood as a stochastic process⁴⁴ meaning that in convergent transformation some variations are more probable than others. In contrast to biological evolution, where variations are random and are produced blindly, cultural content is quite often created to a certain extent intentionally and purposefully.⁴⁵ This process is transformative by definition, since it is a chain of mental representations being transformed into public representations by the authors of the texts and then again into mental representations by individual members of the audience.⁴⁶ In this case, it can be assumed that the nonrandomness in the diversity of Paul’s portrayals in early Christian texts lies in them being influenced by the interests of individual authors and the audience for which they wrote – whether that audience was a real community or an imagined one.

3. Detecting prototypes with distributional semantics

While fully aware of the limitations concerning access to the authors’ or audiences’ minds, the ever-changing social contexts, and groups and communities behind the text, we approach

⁴¹ LUOMANEN, “Social Identity, Prototypes, and Exemplars,” 103; STOWERS, “Kinds of Myth, Meals, and Power,” 110.

⁴² DAN SPERBER, *Explaining Culture: A naturalistic approach* (Oxford: Blackwell Publishers, 1996), 101.

⁴³ ALBERTO ACERBI et al., “Culture without Copying or Selection,” *Evolutionary Human Sciences* 3 (2021): 1–17, 2. Acerbi et al. in their article convincingly show that “the emergence and long-term persistence of cultural traditions does not necessarily require any processes of transmission whose proper (evolved) function is high-fidelity transmission of cultural information. High-fidelity copying is therefore but one of several factors that can ensure intergenerational stability in an evolutionary system. Cultural traditions can also emerge and remain stable as a consequence of any social process—of which there are many—that produces convergent transformation.” ACERBI et al., “Culture without Copying or Selection,” 12. High-fidelity copying can also be at stake in the transmission of Pauline tradition within early Christian texts, as high-fidelity copying and convergent transformation need not be understood as contradictory forces, but complementary.

⁴⁴ ACERBI et al., “Culture without Copying or Selection,” 4.

⁴⁵ But we do not mean “guided variation” in the strong sense, see ACERBI et al., “Culture without Copying or Selection,” 4.

⁴⁶ SPERBER, *Explaining Culture*. Czachesz provides a good overview of this debate from the perspective of biblical studies. ISTVÁN CZACHESZ, *Cognitive Science and the New Testament: A New Approach to Early Christian Research* (Oxford: Oxford University Press, 2017), 62–85.

the texts by means of computational text analysis methods of distributional semantics,⁴⁷ which, in fact, has much in common with the concept of prototypicality. Distributional semantics advocates a non-essentialist approach to language and denies that words have a general, all-applicable meaning. Similarly, the idea of prototypicality is based on criticism of classic theories which define groups by a set of external criteria.⁴⁸ The basic hypothesis of distributional semantics is that words, which are used in similar context – i. e., have similar distribution – tend to have similar meanings. Prototypes are likewise considered as highly sensitive to perceptual and social contexts.⁴⁹ This long-term context-dependent change is, in fact, our focus when analyzing texts from a time period covering five hundred years. The visualization method of word-embeddings in particular is very suitable for the analysis of prototypicality, as it can capture the contextual change on the level of meanings in the human cognition and not only in terms of words that occur together in the text.

A prototype is an abstraction that represents central aspects of a category. We can therefore conceptualize it as a cultural schema, “the organization of cognitive elements into an abstract mental object capable of being held in working memory with default values or open slots which can be variously filled in with appropriate specifics.”⁵⁰ We can imagine these schemas as cognitive networks or patterns of related words that one creates and strengthens through sharing and that are culturally given and changeable. We therefore assume that the prototype as a cultural schema is reflected in the texts, because the author of the text wants to share and further disseminate this schema, i. e., his or her idea of prototypicality. Thanks to the methods from the field of distributional semantics, we can make these networks of related words and terms in the human mind visible because the author most likely embodied them in the text to some extent.⁵¹ Word-embedding allows us, unlike simple word co-occurrence, to show and

⁴⁷ Distributional semantics have a strong response in cognitive science as a new approach to test cognitive theories using large textual corpora. See “Part II: LSA in Cognitive Theory” in THOMAS K. LANDAUER et al., eds., *Handbook of Latent Semantics* (New York; London: Routledge, 2007), 89–206.

⁴⁸ For the difference between a classical, Aristotelian view of categories and a natural view of categories, see HALLAHAN, “Prototypes,” 715–16.

⁴⁹ Word meanings are fluid and flexible. To model word meanings there must be used a method that is dynamic, not static, and is able to capture the semantic affinity of words, which cannot be inferred simply from the fact that they often appear together in one sentence. Similarly, prototypicality as a fuzzy set of attributes (perceptions, attitudes, feelings, and behaviors) also cannot be extracted simply by co-occurrence of terms in the text. Paul, for example, is initially in biblical and early Christian texts closely linked to Jesus, to Christ and to God. Semantically, these characters function as important content of Paul’s message and its authoritative guarantors. However, their closeness to Paul is also in line with Paul’s imitational strategy, where he depicts himself (and is later depicted by others) as being Christ- or God-like—in following Christ’s suffering, for instance. But these texts never actually say that Paul is Christ, it is included in the text more implicitly. Paul (and pseudo-pauline authors) actually depict Paul in a Christ-like manner, while not explicitly revealing to us that they are doing this. Therefore, this cannot be detected using the word co-occurrence method. But word-embedding is able to recognize this semantic similarity and make it visible, because it is able to point out that although the terms Paul, Jesus, Christ, and God often do not appear together in one sentence, they do appear in sentences that express similar ideas.

⁵⁰ ROY G. D’ANDRADE, *The Development of Cognitive Anthropology* (Cambridge: Cambridge University Press, 1995), 179.

⁵¹ Similarly, Czachesz considers the construction of semantic networks to be a simulation of mental representations in the human mind: “We can view the networks described in this study as models of the network of associations that the text generates in the readers’ or listeners’ minds.” ISTVÁN CZACHESZ, “Network Analysis of Biblical Texts,” *Journal of Cognitive Historiography* 3/1–2 (2016): 43–67, 45. Luomanen also assumes that the authors incorporate their idea of prototypicality into the text: “The writers of the New Testament gospels had a message

reveal the connection between words that are semantically related – without this connection being visible at first sight (e. g., at the sentence level) – and thus contribute to the discovery of what constituted the image of Paul as a prototypical figure of early Christianity and how this image changed in the first five centuries.

4. Methods of the analysis

The application of computational analysis methods we adopt in this study can be understood in the broader research context of digital humanities. The research in digital humanities is usually conceptualized through three basic elements, which are content, tools and methods. Digitally processed content (e. g. digitized text corpora) is the primary source for research; digital tools (e. g. programming languages such as Python or R) enable the analysis, data mining, visualization and modeling of this content; and knowledge of scientific methods⁵² then gives the researcher a framework for interpreting the data obtained. The use of digital content, tools and methods transforms research in humanities not only by establishing broad interdisciplinary collaborations across disciplines, but also by enabling research that would not be possible without digital tools and methods, as it generates new research questions and formulates hypotheses driven by insights that can only be achieved with the help of new technologies.⁵³

In the case of this article, the digitally processed content is the LAGT corpus (Lemmatized Ancient Greek Texts).⁵⁴ In total, LAGT consists of 1,457 ancient Greek texts, 2,891,346 sentences and 31,248,866 words. It has been built by merging textual data from two open-access corpora: the Canonical Greek Literature dataset from the Perseus Digital Library⁵⁵ and the First Thousand Years of Greek dataset of the Open Greek & Latin project.⁵⁶ As the title of the corpus suggests, the texts within LAGT are lemmatized. That means that each word is represented by its lemma, i. e. by its dictionary-like form. Thus, all nouns are in the nominative singular. In addition to that, the words are also filtered on the basis of part-of-speech categories: only nouns, adjectives and verbs are included. This is because the LAGT corpus is designed to facilitate semantically oriented computational text analysis research, for which these part-of-speech categories are the most important. Thus, the first verse of Paul's letter to the Philippians (Παῦλος καὶ Τιμόθεος δοῦλοι Χριστοῦ Ἰησοῦ πᾶσιν τοῖς ἁγίοις ἐν Χριστῷ Ἰησοῦ τοῖς οὖσιν ἐν Φιλίπποις σὺν ἐπισκόποις καὶ διακόνοις.) is represented as the following

to convey, the Good News. Therefore, it is safe to hypothesize that as far as the texts signal normative perceptions, attitudes, feelings, and behaviors—by showing these to be desirable through the narrative—*these likely reflect prototypicalities and prototypes in the minds of their authors.*” (Emphasis his). LUOMANEN, “Social Identity, Prototypes, and Exemplars,” 103.

⁵² For a list of methods see e. g. AHDS Taxonomy of Computational Methods in the Arts and Humanities; Oxford University Digital Humanities Program or Taxonomy of Digital Research Activities in the Humanities.

⁵³ LORNA HUGHES, PANOS CONSTANTOPOULOS, and COSTIS DALLAS, “Digital Methods in the Humanities: Understanding and Describing their Use across the Disciplines,” in *A New Companion to Digital Humanities*, ed. SUSAN SCHREIBMAN, RAY SIEMENS, and JOHN UNSWORTH (Malden: Wiley-Blackwell, 2016), 150–70, 152.

⁵⁴ VOJTĚCH KAŠE, LAGT (v1.0). *Zenodo*. 2021, <https://doi.org/10.5281/zenodo.4971946> (Last access 14.11.2022).

⁵⁵ LISA CERRATO et al., PerseusDL/canonical-greekLit 0.0.2711 (Version 0.0.2711). *Zenodo*, <http://doi.org/10.5281/zenodo.4067170> (Last access 14.11.2022).

⁵⁶ GREGORY R. CRANE et al., First1kGreek (Version 1.1.5070). *Zenodo*, <http://doi.org/10.5281/zenodo.4091475> (Last access 14.11.2022).

list of words: Παῦλος Τιμόθεος δοῦλος Χριστός Ἰησοῦς ἅγιος Χριστός Ἰησοῦς Φίλιπποι ἐπίσκοπος διάκονος. The lemmatization and filtering have been produced by means of automatic algorithms. Therefore, we should not be surprised by occasional mistakes like improperly lemmatized words. However, as our previous research employing this corpus indicates, these occasional mistakes do not prevent us from obtaining highly relevant insights on the basis of the data.⁵⁷

From the LAGT corpus we extract a subset of early Christian texts from the 1st to the 5th century CE. This subset consists of 3,566,128 words dispersed over 147 works associated with 47 individual authors or authorship communities.⁵⁸ We subsequently analyze these data on two levels. First, we focus on individual authors and authorship communities and compare how they refer to Paul by quantitatively analyzing the content of the sentences containing his name. For that purpose, we employ the TFIDF algorithm. TFIDF stands for *term frequency (TF) – inverse document frequency (IDF)*. This algorithm weights the frequency of a term appearing within sentences containing the name of Paul by frequency of the same term in the rest of the text. Adopting this method, for each author or authorship community in the subset of early Christian texts, we detect the words with the highest TFIDF within the sentences containing Paul's name. These words we consider as being very informative concerning how each individual author or authorship community understands Paul.

From the semantic point of view, the TFIDF method has one substantial limitation: It detects only words frequently appearing together with Paul's name. But, for instance, it is completely unsuitable to capture synonyms, as synonyms usually appear independently. Thus, we can hypothesize that the word apostle is repeatedly used by some early Christian texts as a synonym to Paul. However, the TFIDF method focusing on sentences containing Paul's name is not able to detect this.

This brings us to the second, more advanced, computational text analysis, which will be based on distributional (or vector) semantics. Distributional semantics is a quickly evolving research paradigm, employing a broad palette of methods from the computer science field of natural language processing and computational linguistics.⁵⁹ The underlying idea of distributional semantics has been formulated already in the 1950s by Zellig S. Harris as the so-called distributional hypothesis,⁶⁰ claiming that the meaning of a word can be grasped as a function

⁵⁷ VOJTĚCH LINKA and VOJTĚCH KAŠE, "Pain and the Body in Corpus Hippocraticum: A Distributional Semantic Analysis," *Digital Classics Online* (2021): 54–71.

⁵⁸ An example of an authorship community might be Johannine literature, including the Gospel of John and the three epistles, as we assume that the Johannine literature "is the product of a school, shaped by the vocabulary, idiom, point of view, and worldview of a single seminal teacher." JOHN PAINTER, "Johannine Literature: The Gospel and Letters of John," in *The Blackwell Companion to the New Testament*, ed. DAVID E. AUNE (Oxford: Wiley-Blackwell, 2009), 344–72, 345.

⁵⁹ For an overview, see ALESSANDRO LENCI, "Distributional Models of Word Meaning," *Annual Review of Linguistics* 4 (2018): 151–71 and DANIEL JURAFSKY and JAMES H. MARTIN, "Vector Semantics and Embeddings," in *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, ed. DANIEL JURAFSKY and JAMES H. MARTIN, 3rd ed. Draft (2021), see <https://web.stanford.edu/~jurafsky/slp3/>.

⁶⁰ ZELIG S. HARRIS, "Distributional Structure," *Word & World* 10 (1954): 146–62; cf. MAGNUS SAHLGREN, "The Distributional Hypothesis," *Rivista Di Linguistica* 20/1 (2008): 33–53.

of its distribution.⁶¹ By distribution, Harris means a sum of all language environments in which a word appears, where an environment is a particular language context of co-occurring words (i.e. a sentence, a paragraph etc.).⁶² According to Harris, if we take any two words A and B, “the amount of meaning difference correspond[s] roughly to the amount of difference in their environments” and “If A and B have almost identical environments . . . we say they are synonyms.”⁶³ Thus, we see that this approach explicitly focuses on the type of semantic relatedness we were not able to capture by means of the TFIDF method.

Harris himself did not offer any formal model to mathematically evaluate the difference in meaning between any two words. Such models started to be developed in the subsequent decades, together with the growth of large corpora of digitized textual data and with increasing computational power of computers. Such models build on mathematical apparatus of linear algebra, representing and analyzing language data in terms of vectors and matrices. In that context, the meaning of a word is defined by a vector determining its position in a multidimensional space.⁶⁴ Such projections of language data into a N-dimensional space are known as word-embeddings.

Perhaps the most important feature of the vector representations of words and of the word embeddings is that individual vectors can be subjected to comparison. Thus, we can expect that in a functional distributional semantics model two synonymous words (e. g., eye-doctor and oculist, to use Harris’s example) will have highly similar vectors. This similarity will also cause that they will be projected in a close proximity to each other in the word embedding representation of the corpus.

The distributional semantics models are usually created by means of being trained on digitized corpora of texts. There are dozens of algorithms commonly used for such training.⁶⁵ While one, older, family of models is based on algorithms producing the vectors in a deterministic way by means of a series of algebraic transformations of word co-occurrence data (count-based models),⁶⁶ another family is based on machine learning algorithms that directly create the vectors by iteratively learning to optimally predict the contexts of a target word (prediction-based models).⁶⁷ Because of the recent advance of machine learning techniques based on deep neural networks, the prediction-based models have become increasingly popular over the last decade, in some contexts completely replacing the older count-based alternatives. However, some scholars argue that the count-based models still have their applicability. The reason is that the performance of individual algorithms used to train the models is

⁶¹ HARRIS, “Distributional Structure,” 155–58.

⁶² HARRIS, “Distributional Structure,” 146.

⁶³ HARRIS, “Distributional Structure,” 157.

⁶⁴ The mathematical apparatus of linear algebra is not constrained by the 3-dimensional space we are used to, but might be employed to vectors having hundreds of dimensions. See GILBERT STRANG, *Introduction to Linear Algebra* (Wellesley: Wellesley-Cambridge Press, 2016).

⁶⁵ See LENCI, “Distributional Models of Word Meaning,” 151–71.

⁶⁶ Perhaps the best-known example of such a model is the Latent Semantic Analysis. THOMAS K. LANDAUER, PETER W. FOLTZ, and DARRELL LAHAM, “An Introduction to Latent Semantic Analysis,” *Discourse Process* 25/2–3 (1998): 259–84.

⁶⁷ The most widespread model from this family is the word2vec algorithm. TOMAS MIKOLOV et al., “Efficient Estimation of Word Representations in Vector Space,” *ArXiv [Cs.CL]* (2013): 1–12.

dependent on the size of the input data. It appears that when the input data are small, the older count-based models still outperform the prediction-based models.⁶⁸

The dependence of appropriate models on the size of the input textual data is of special relevance for our analysis based on a subset of texts from the LAGT corpus. As we already mentioned, the subset of early Christian texts consists of approx. 3.5 million words. Even when treated as a whole, such amount of textual data is not sufficient to properly train some prediction-based models, which require at least 10 million words to produce comparable results with some count-based models.⁶⁹ However, even the count-based models require to be trained on textual data counting at least hundreds of thousands of words. This effectively prevents us from building its own distributional model for each individual author or authorship community within the corpus. Instead of this, we divide the subset of early Christian texts into two sub-corpora of comparable size: (1) the subcorpus of early Christian texts from the 1st to the 3rd century CE (the 1–3CE subcorpus), (2) the subcorpus of early Christian texts from the 4th to the 5th century CE (the 4–5CE subcorpus).

In this article, we build the two distributional semantic models by means of count-based algorithms including the PPMI preprocessing of word co-occurrence data. In both cases, the procedure is the same. First, we create a word-word co-occurrence matrix of the 2,000 most frequent words within the subcorpus. The word-word co-occurrence matrix allows us to inspect how often any of the 2,000 words appears together within a sentence with any other from those 2,000 words. But this is only a starting point here. Second, the word-word co-occurrence matrix is transformed into a PPMI co-occurrence matrix, in which the raw co-occurrences are modified by means of Mutual Information.⁷⁰ Mutual Information weights the probability of observing two words together (i. e. the frequency of their co-occurrence) by the probabilities of observing them independently (i. e. their overall frequencies). Several variants of Mutual Information calculations are described within the literature. Here we employ the so-called PPMI², which modifies the measure by squaring the joint probability.⁷¹ The resulting matrix can be treated as a list of vectors, with each vector having 2,000 dimensions. Since such vectors typically contain a large number of zeros (there are many words which do not appear together at all, especially rare words), a statistical comparison of these vectors can result in false correlations. Therefore, in the next step, the dimensions of the vectors are reduced from the 2,000 dimensions to 150 dimensions by means of Singular Value Decomposition.⁷² As a result, each of the 2,000 words is associated with a 150-dimensional vector.

⁶⁸ MAGNUS SAHLGREN and ALESSANDRO LENCI, “The Effects of Data Size and Frequency Range on Distributional Semantic Models,” *ArXiv [Cs.CL]* (2006): 1–6.

⁶⁹ The performance of the models is typically evaluated by reference to their ability to automatically detect semantic relationships between words previously annotated by human coders. For instance, the performance of distributional semantics models trained on English corpora is commonly evaluated by their ability to detect synonyms used in the TOEFL tests. See SAHLGREN and LENCI, “The Effects of Data Size and Frequency Range,” 1–6.

⁷⁰ KENNETH W. CHURCH and PATRICK HANKS, “Word Association Norms, Mutual Information, and Lexicography,” *Computational Linguistics* 16/1 (1990): 22–29.

⁷¹ See OMER LEVY, YOAV GOLDBERG, and IDO DAGAN, “Improving Distributional Similarity with Lessons Learned from Word Embeddings,” *Transactions of the Association for Computational Linguistics* 3 (2015): 211–25. In another study (LINKA and KAŠE, “Pain and the Body in Corpus Hippocraticum”), we used another variant of the algorithm, PPMI³, since it worked better with the small corpora we had there.

⁷² STRANG, *Introduction to Linear Algebra*, 364–400.

These 150-dimensional vectors are finally employed to calculate semantic relatedness between any two words, which is a basis for both identifying the nearest neighbors of a word and for visualizing the semantic similarities between words by means of word-embedding plots.

5. Results: From a missionary to the gentiles to a symbol of orthodoxy

5.1. Word co-occurrence and TFIDF

Word co-occurrence and TFIDF allow us to compare individual authors and look at the textual context in which Paul appeared on a micro-level. On Table 1 we offer the 10 terms with the highest TFIDF score within sentences containing Paul's name (Παῦλος) in a subselection of early Christian authors. The 10 terms with the highest TFIDF score within Luke-Acts reflect the volume's focus on missionary activity: Paul works closely together with Barnabas (Βαρναβᾶς), moving from city to city (πόλις). Importantly, the analysis also highlights Paul's complex relationship with Judaism.⁷³ Ἰουδαῖος (Judean, Jew) is, in fact, the most frequent term in sentences with Paul's name. For Luke, Paul himself is a Jew (e. g., Acts 21:24; 23:6) and his mission strategy is to preach first to Jews (e. g., Acts 17:1–3, 10), but he is also challenged by Jews on several occasions (e. g., Acts 17:5; 23:12). As will become evident below, later writers tend to lose this strong connection between Paul and Judaism.

author	TFIDF terms
pseudo-Pauline literature	Ἰησοῦς, ἀπόστολος, Χριστός, θεός, Τιμόθεος, ἅγιος, ἔθνος, γνωρίζω, μυστήριον, θέλημα
Luke-Acts	Ἰουδαῖος, λέγω, ἀνὴρ, Βαρναβᾶς, μέλλω, πολὺς, ἡμέρα, ἀδελφός, πόλις, φημί
Clement of Rome	δύσις, ἀναλαμβάνω, κόσμος, βραβεῖον, ὑποδείκνυμι, φορέσας, φυγαδευθεῖς, λιθασθεῖς, κήρυξ, τέρμα
Ignatius of Antioch	πάροδος, ἀναιρέω, συμμύσται, ἡγιασμένου, διατάσσω, θεός
Epiphanius	ἀπόστολος, λέγω, Πέτρος, ἅγιος, εἰμί, οὗτος, Χριστός, φημί, ἐπιστολή, ἐπίσκοπος
Athanasius of Alexandria	λέγω, γράφω, θεός, μακάριος, χάρις, πολὺς, οἶδα, Χριστός, ἐπιστολή, οὐσία
Socrates of Constantinople	ἐπίσκοπος, ἐσπέριος, ἐκκλησία, γίγνομαι, βασιλεύς, ναυατιανῶν, ψιλός, μακεδόνιον, ἀθανασίω, ἀθανάσιον
Theodoret	ἐπίσκοπος, ζευγματέων, ἐπίσκοπος, ἀθανάσιον, λέγω, οὗτος, Χριστός, ἀληθινός, θεός, εἰμί
Philostorgius	σπανὸν, ὑπογράφας, τελοῦντα, ἐμπικρανάμενον, γάλλω, παραδίδωμι, πῦρ, βασιλεύς, πολὺς
Sozomen	ἀπόστολος, κωνσταντινουπόλεως, ἐκκλησία, γίγνομαι, θρόνος, Μακεδών, ἐπίσκοπος, Ἀντιόχεια, σαμοσατέως, κωνσταντινουπόλει

⁷³ On Paul's relationship to his Jewishness and to the Jewish faith, see MARKUS TIWALD, *Hebräer von Hebräern. Paulus auf dem Hintergrund frühjüdischer Argumentation und biblischer Interpretation* (Freiburg: Herder, 2008), where the author argues that Paul did not part with Judaism, but only promoted a new way of understanding the Jewish faith under the influence of his interpretation of Jesus's death on the cross.

Table 1: 10 words with the highest TFIDF score in sentences containing Paul's name (a subselection of early Christian authors and authorship communities).⁷⁴

In pseudo-Pauline canonical literature, Paul is, first and foremost, known as an apostle (ἀπόστολος).⁷⁵ He is also closely linked to “Jesus,” “Christ,” and “God” (Ἰησοῦς, Χριστός, θεός) (see discussion below). Furthermore, Paul becomes known as a revealer of God's mystery (γνωρίζω, μυστήριον) and he is emphatically the apostle to the gentiles (ἔθνος). Of the Apostolic Fathers, Clement of Rome applies geographical and cosmological terminology to depict Paul as a worldwide missionary (δύσις, κόσμος, τέρμα, κῆρυξ, see 1 Clem. 5:5), while Ignatius of Antioch opts to emphasize Paul's martyrdom (ἀναιρέω), and to prime a connection to him via travel-related metaphors (πάροδος, Ign. Eph. 12:2).

Later Christian authors testify to the versatile and highly contextual reception of Paul's character. While they commonly keep linking the title of apostle to Paul, especially later authors prefer to associate Paul with the stabilizing church offices such as bishop (ἐπίσκοπος). In Socrates of Constantinople and Theodoret, “bishop” is, in fact, the term most frequently seen in sentences with “Paul” (see also Epiphanius and Sozomen). Letters and the process of writing also become variously connected to Paul's character (Athanasius: γράφω, ἐπιστολή; Philostorgius: ὑπογραφέας; Epiphanius: ἐπιστολή).

5.2. Nearest neighbors of Παῦλος within the distributional semantics models

The TFIDF algorithm focuses on word co-occurrences. It allows us to compare usage of certain words between individual authors, but it does not capture deeper semantic relatedness aka synonymity. To obtain some insights on this deeper level, we have to employ different methods. In this respect, we will “zoom-out” from the level of individual authors and look at the texts as two subcorpora: a subcorpus for 1–3 centuries CE and a subcorpus for 4–5 centuries CE. On Table 2 we see the terms revealing the highest similarity with the term Παῦλος in the subcorpora. The similarity is measured as a cosine similarity between two multidimensional vectors representing the words within the distributional semantics model. Thus, for instance, within the distributional model created on the basis of the earlier subcorpus, the term ἀπόστολος has the highest similarity with Παῦλος (0.9061). The term can be considered nearest neighbor of the term Παῦλος. It is important to emphasize once again that this type of neighborhood between any two words is not based on their co-occurrences (i. e. joint appearances within sentences), but rather on similarity of total language contexts in which they appear (i. e. similar distribution of co-occurrences with all remaining words within the corpus). Thus, the method attempts to capture the extent of synonymity between any two words; this does not presuppose their frequent co-appearance, since synonyms tend to substitute each other.

⁷⁴ For a full list of authors and authorship communities with additional metadata, see https://github.com/kasev/paul/blob/master/data/authors_table.csv (Last access 14.11.2022).

⁷⁵ Cf. Luke's virtual denial of this title to Paul, Acts 14:14 forming an exception.

1–3 CE Παῦλος NNs	1–3 CE Παῦλος CoS.	4–5 CE Παῦλος NNs	4–5 CE Παῦλος CoS.
ἀπόστολος	0.9061	αἵρεσις	0.8592
εὐαγγέλιον	0.8773	ἀπόστολος	0.8537
ἐκκλησία	0.874	δέχομαι	0.8454
ἀκούω	0.8728	διδάσκω	0.8435
πιστεύω	0.8718	ἀδελφός	0.8427
πίστις	0.8711	ὑστερος	0.842
ἔθνος	0.8663	ἐπίσκοπος	0.841
ἅγιος	0.8661	πίστις	0.8409
Χριστός	0.8659	μανθάνω	0.8402
διδάσκω	0.8644	δόξα	0.8401
Ἰησοῦς	0.8637	ἐπιστολή	0.8371
τόπος	0.8625	πρέσβυς	0.8353
δίδωμι	0.8621	ὄνομα	0.8338
Ἰουδαῖος	0.8612	γιγνώσκω	0.8335
οἶδα	0.8607	ἐκκλησία	0.833
σωτήρ	0.86	παραδίδωμι	0.8329
ἴδιος	0.8595	Πέτρος	0.8324
ἐθέλω	0.8595	γράμμα	0.8294
προφήτης	0.8595	υἱός	0.8275
πνεῦμα	0.8566	ἴδιος	0.8275

Table 2: Nearest neighbors of the term Παῦλος within the two subcorpora. 20 terms with the highest cosine similarity with Παῦλος in the distributional semantics vectors.

From the 1–3 centuries CE to the 4–5 centuries, there is a significant change in the terminology relating to “Paul” – only five out of 20 terms remain the same (ἀπόστολος, ἴδιος, διδάσκω, πίστις, ἐκκλησία). Paul is consistently known as ἀπόστολος in both periods. If we were to include the numerous instances where Paul is known simply as “the Apostle,”⁷⁶ the connection would be even stronger.

The earlier period paints Paul as a bearer of oral tradition (ἀκούω, δίδωμι), gifted with spirit (πνεῦμα), missionary among the gentiles (ἔθνος), one who spreads the good message (εὐαγγέλιον), and teaches (διδάσκω, οἶδα). Paul is still associated with his Jewish origin (Ἰουδαῖος, προφήτης) and exhibits a strong connection with Jesus (Ἰησοῦς, Χριστός, σωτήρ). In the 4–5 centuries, the term αἵρεσις (heretical sect, faction) climbs to the very top of the list,

⁷⁶ See WHITE, *Remembering Paul*, 145.

although it is entirely absent from the 20 most similar terms in the 1–3 centuries. While prototypical characters always mirror, comment, and shape the group they represent, here the interest in group boundaries becomes explicit and heightened. It is also combined with an emphatically negative disposition towards outsiders. In this process, Paul becomes a token of orthodoxy and a doctrinal figure associated with teaching (διδάσκω), learning (μανθάνω), knowing (γινώσκω), tradition (παραδίδωμι) and other terms linked to the content and reception of beliefs (δόξα, δέχομαι).

The development is in line with another trend, which associates Paul increasingly with letters and written material (ἐπιστολή, γράμμα). The previous strong connection between Paul and the gospel (εὐαγγέλιον) evaporates in the 4–5 centuries – perhaps because the term in Paul’s usage referred to an oral message (note also the disappearance of ἀκούω) and soon became used of the literary canonical gospels. The change may reflect a growing interest in the biblical canon and the emergence of a class of learned Christian experts and offices: note the proximity of Paul to ἐπίσκοπος, πρέσβυς, and ἐκκλησία during this period. Simply put, we see a shift from mission and orality to institutions and literacy. Another, more expected development is the disappearance of Paul’s complex relationship to Judaism (Ἰουδαϊσμός) with the accompanying distinction between Jews and gentiles (ἔθνος), and Paul specifically as the apostle to the latter. As the overwhelming majority of Christians at this later time consisted of non-Jewish believers, the distinction became meaningless.

The texts from 1–3 centuries reveal a strong similarity between Paul and Jesus as Christ/Savior (Χριστός, Ἰησοῦς, σωτήρ). Interestingly, this connection was no longer made in the 4–5 centuries. The term υἱός, which gained a religious sense due to Christianity,⁷⁷ replaces Χριστός, Ἰησοῦς, σωτήρ as a connective feature between Christ and Paul. It may be that the development of high, dogmatic Christology eventually separated Christ and Paul as the former became more and more supernatural (note the appearance of dogmatic Christological terms such as ὁμοούσιον, ὑπόστασις, and πρωτότοκος in the later period). Paul, while highly esteemed, remained a fully human and realistic role model to members of the church. Simultaneously, Paul approaches another exemplar, Peter (Πέτρος). This supports the general trend visible from Acts onwards according to which Peter’s historical role as Paul’s rival (Gal 2:11) gives way to a myth of a unified early church (Acts 15).

5.3. Word-embeddings

On two subplots in Figure 1, we plot data obtained using a scatter plot, which is the standard way to visualize word-embeddings. Based on the relationship between prototypicality and distributional semantics, as postulated above (ch. 3), we suggest that this visualization informs us about the changes in the idea of a prototypical Paul in the two monitored periods. For better clarity and a stronger informative value of our visualizations, we created five categories of terms that are typically associated with Paul in early Christian texts and picked up

⁷⁷ MARTINA A. RODDA, MARCO S.G. SENALDI, and ALESSANDRO LENCI, “Panta rei: Tracking Semantic Change with Distributional Semantics in Ancient Greek,” *Italian Journal of Computational Linguistics* 3/1 (2017): 11–24, esp. 17.

in Pauline research (see Table 3). Here we are thus open to insight gained by means of traditional close reading. These are his Jewish background, his connection with suffering and martyrdom, christological and mariological discourse, terms associated with eschatology and, of course, words referring to the formation of the Church as an institution, i. e. words associated with ecclesiology. Based on the extraction of a table of 2,000 most frequent terms from both subcorpora (counter-balancing the close reading), we then selected 15 words for each category as seen in the table below. Our criterion was that these have to be terms that have the highest possible informative value for a given category and at the same time are so specific that their field of meaning really typically overlaps with the given category. Therefore, we have left ambiguous words, such as σῶμα, λαός, λόγος, φύσις, οὐσία, etc.

jewishness (green)	martyrdom (red)	christology & mariology (yellow)	eschatology (blue)	ecclesiology (purple)
Ἰουδαῖος	μαρτυρία	υἱός	κρίσις	ἐκκλησία
συνέδριον	διώκω	κύριος	ἀπώλεια	διάκονος
Ἑβραῖος	πάθη	μόνος	ἀποκάλυψις	ἀνάθεμα
Ἱεροσόλυμα	μάστιξ	Χριστός	ἀποκαλύπτω	ἐπίσκοπος
ἱερεύς	δεσμωτήριον	Ἰησοῦς	παρουσία	πρεσβύτες
θυσία	φεύγω	γεννάω	θρόνος	κοινωνέω
νομός	ἀποθνήσκω	γωνία	ἄγγελος	κοινωνία
συναγωγή	μάχη	ἀγέννητος	στοιχεῖον	αἵρεσις
διαθήκη	αἰχμαλωσία	σώζω	ὀργή	τάξις
πάσχα	θάνατος	ὑπόστασις	σημεῖον	χάρισμα
σάββατον	πάσχω	πρωτότοκος	διάβολος	ἐκκλησιάζω
νομοθετέω	δεσμός	ὁμοούσιον	δαίμων	προσευχή
ἱερωσύνη	φυλακή	σταυρόω	σκότος	γερουσία
Ἰουδαία	πάσχω	σωτήρ	αἰώνιος	γέρον
μουσέως	ἀποκτείνω	παρθένος	ἄστρον	ἁρμονία

Table 3: Term categories.

The following two figures visualize the semantic closeness of the term Paul with selected words, color-coded according to categories:

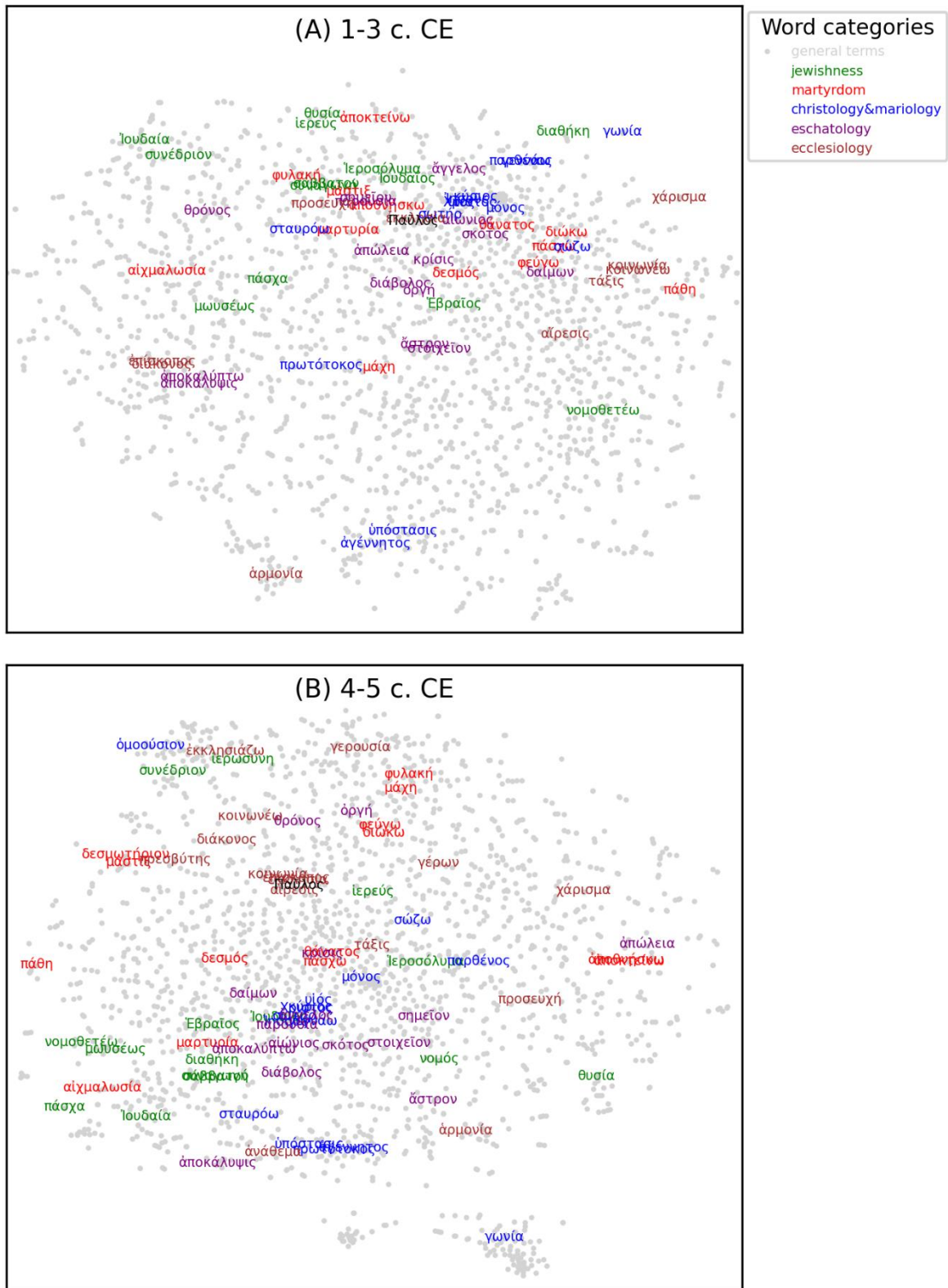


Figure 1: Word-embeddings subplots with the 2,000 most frequent words.

Based on the relationship between prototypicality and distributional semantics, which we have postulated above, we suggest that this visualization informs us about the change of the idea of a prototypical Paul in the two monitored periods. Since the authors we examine are all positively disposed to Paul, we presume they are motivated to mold him into a highly prototypical character. The way Paul appears in a given text is thus much more than an attempt to describe the historical apostle: Paul, and whichever characteristics he is associated with, represents a statement about the ingroup. This conception the authors wanted to spread among their readers – whether members of real, imagined, or intended communities. We can therefore assume that cognitive schemas in the form of mental representations are projected into their texts. These schemas we can imagine as networks of terms associated with a certain concept. One of these concepts is the idea of Paul as a prototype. We therefore assume that the word-embedding method shows us, at least to some extent, what such a cultural schema – created and disseminated by the authors of early Christian texts around the figure of Paul – looked like from the large-scale view and how it changed in the two periods under scrutiny. The word-embeddings data serve here as an approximation of mental representations of the minds of the writers, readers and listeners to the texts. In terms of the cognitive discussion, the texts can be understood as shared cultural schemas, i. e. shared public representations, which both reflect the mental representations in the minds of their producers and cause modifications of mental representations in the minds of their readers or listeners, and this process is to some extent imprinted in these schemas.

The scatter plot images confirm many of the general observations made with the help of our other methods. In the first period (1–3 centuries), Paul is still very strongly associated with his Jewish background. Paul is also connected with martyrdom and is depicted as a moral ideal worth following, a kind of “second Jesus.” It is therefore not surprising that the terms associated with the emerging Christology are also very close to him. The emphasis on martyrdom and suffering relates to ingroup cohesion, moral obligations, and reliable cooperation via the mechanism of commitment or costly or hard-to-fake signaling. Through the prototypical suffering and dying Paul signals that the ingroup members should likewise be willing to sacrifice their lives as the ultimate token of their loyalty to the group.⁷⁸

Eschatological discourse is also more prominent in the first period and takes precedence over terms related to ecclesiology. In contrast, in the second period (4–5 centuries) we observe the exact opposite trend. The awareness of Paul’s connection with ecclesiastical institutions is very clear, while terms referring to Jewishness and martyrdom are more scattered and no longer play a prominent role. On the level of group identity maintenance, we may carefully suggest that costly signals in the form of suffering and martyrdom become replaced with

⁷⁸ On commitment/hard-to-fake/costly signaling, see JOSEPH BULBULIA and RICHARD SOSIS, “Signalling Theory and the Evolutionary Study of Religions,” *Religion* 41/3 (2011): 363–88. On the cultural evolutionary success of the suffering and dying Paul, see NINA NIKKI and ANTTI VANHOJA, “Righteous Sufferer, Scheming Apostate: Traditions of Paul from a Cultural Evolutionary Perspective,” in *Religious Identities in Antiquity and the Early Middle Ages: Walking Together and Parting Ways*, ed. ILKKA LINDSTEDT, NINA NIKKI, and RIIKKA TUORI (Leiden: Brill, 2021), 115–44.

institutional and doctrinal group discipline.⁷⁹ Overall, we can see a shift from Paul as a missionary, speaker and teacher to a doctrinal character and symbol of orthodox church. Importantly, the plot allows us to see differences that are not visible in the table of the nearest neighbors (see Table 2). In that table the distance between the name Paul and the term πίστις (faith) look quite similar for both time periods. However, when we zoom into the neighborhood of πίστις in the word-embeddings plots (see Figure 2), we see that in the 1–3 centuries it resembles terms such as grace (χάρις), promise (ἐπαγγελία), and hope (ἐλπίς), but in the 4–5 centuries it is surrounded by terms denoting churchly doctrine and offices (ἐκκλησία, ἐπίσκοπος, κοινωνία, δόγμα). Further differences, thus, emerge between the terms which initially seemed to behave similarly.

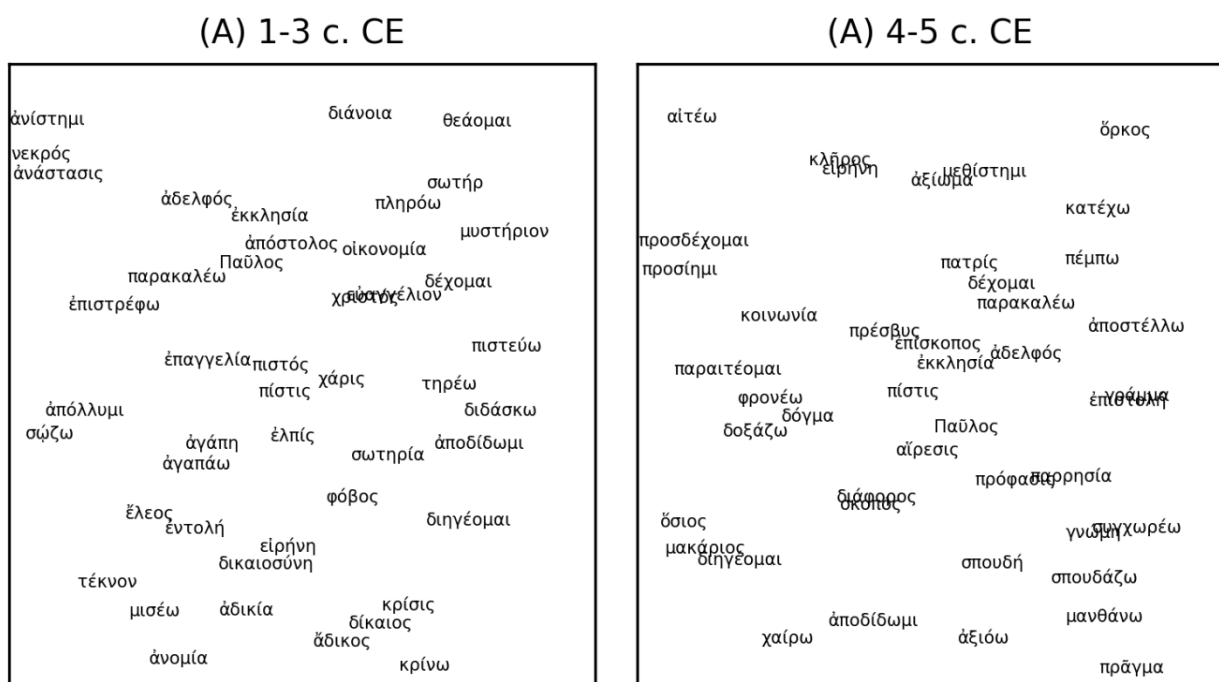


Figure 2: Word-embeddings. πίστις in detail.

6. Conclusions

This article set out to look at how Paul is depicted as a prototypical character in the first five hundred years of Christian literature in the Greek language. The concept of prototypicality,

⁷⁹ In this context, it would be interesting to take a closer look to what extent we can see here a shift from an imagistic mode of religiosity to a doctrinal mode of religiosity, postulated by HARVEY WHITEHOUSE, *Modes of Religiosity: A Cognitive Theory of Religious Transmission* (Walnut Creek, CA: AltaMira Press, 2004). LUTHER H. MARTIN, who encounters aspects of the imagistic and doctrinal modalities of religion among the early Christians, thinks in a similar vein, when he writes that Paul “articulated a principle of authority and transmission that later developed into the doctrinal modality of Christian orthodoxy”. LUTHER H. MARTIN, “The Promise of Cognitive Science for the Study of Early Christianity,” in *Explaining Christian Origins and Early Judaism: Contributions from Cognitive and Social Science*, ed. PETRI LUOMANEN, ILKKA PYYSIÄINEN, and RISTO URO (Leiden: Brill, 2007), 37–56, 50.

as understood and defined in cognitive and social psychology, denotes a fuzzy set of attributes considered to ideally depict the ingroup in relation to outgroups. Since prototypicality is understood to be highly contextual and constantly changing, it was the interest of this article to map these changes in large corpora of texts in the hope of uncovering trends that can be studied together with and compared to what has earlier been suggested by close readings in traditional scholarship.

The article advanced from the perspective of cultural evolution as a general framework encompassing and legitimizing both the individual methods and the general outlook on early Christian tradition. The perspective of cultural evolution denies, for example, that changes in culture denote either development or decline from an early, “original” version of the tradition. Accordingly, the depictions of Paul in early Christian literature were not, in the article, compared to an allegedly historical, “real,” or normative Paul. Rather they were treated as adaptations in their given social and other, such as literary, circumstances. The cultural evolutionary perspective also calls attention to changes in culture over long periods of time and quantitative modeling of these changes. In this article we used distributional semantic modeling to uncover major trends in the changing imagery of Paul.

Our analyses have shown that there is a significant change in the image of Paul in the two observed periods, especially with regard to his relationship to Jesus, his Jewish origin and his martyrdom. At the same time, the shift from the mission and formation of Jesus’s followers towards the institutionalized structures of the early church also had a very significant effect on the prototypical portrayal of the Apostle Paul. This process, in connection with Paul, manifested itself on several levels – from mission to institution, from orality to written tradition, from “good news” to doctrine and orthodoxy and from Paul as a martyr to Paul as the leader of the church.

Finally, we want to suggest some avenues for further investigation into the topic of this article. First, we encourage the reader to examine the supplementary figures we produced for this article, which are available via Github.⁸⁰ There is a wealth of observations to be made which we were not able to fit into this article. As for the five terminological categories we chose for the scatter plot image made on the basis of word embeddings, these were chosen with conscious reflection of relevant literature and scholarly close readings but are by no means comprehensive. Several other categories remain worth examining, such as Paul and Greco-Roman religion, Paul’s relationship to women or rituals, Paul as a letter writer, and especially Paul as a fighter of heresy. The last one raises the issue of the various works that treat Paul as the enemy, which we did not include in our investigation. This tradition includes important texts all the way from mentions in the canonical Revelation of John, Letter of James, and Second Epistle of Peter to the third century Pseudo-Clementine Recognitions. Furthermore, many early authors who were influenced by Paul, wrote in Coptic (the Nag Hammadi Library) or Latin (influential Western church fathers, such as Tertullian). A full investigation into the evolution of the prototypical Paul should include these texts as well.

⁸⁰ See https://github.com/kasev/paul/blob/master/figures/embeddings_large_1-3.png (Last access 14.11.2022) and https://github.com/kasev/paul/blob/master/figures/embeddings_large_4-5.png (Last access 14.11.2022). The Github repository also contains scripts which were used for all the analyses presented in this article.

Bibliography

- ACERBI, ALBERTO, MATHIEU CHARBONNEAU, HELENA MITON, and THOM SCOTT-PHILLIPS. "Culture without Copying or Selection." *Evolutionary Human Sciences* 3 (2021): 1–17.
- BULBULIA, JOSEPH, and RICHARD SOSIS. "Signalling Theory and the Evolutionary Study of Religions." *Religion* 41/3 (2011): 363–88.
- BUTTICAZ, SIMON, ANDREAS DETTWILER, and JENS SCHRÖTER, eds. *Receptions of Paul in Early Christianity: The Person of Paul and His Writings through the Eyes of His Early Interpreters*. Berlin: De Gruyter, 2018.
- CHURCH, KENNETH W., and PATRICK HANKS. "Word Association Norms, Mutual Information, and Lexicography." *Computational Linguistics* 16/1 (1990): 22–29.
- CZACHESZ, ISTVÁN. "Network Analysis of Biblical Texts." *Journal of Cognitive Historiography* 3/1–2 (2016): 43–67.
- CZACHESZ, ISTVÁN. *Cognitive Science and the New Testament: A New Approach to Early Christian Research*. Oxford: Oxford University Press, 2017.
- D'ANDRADE, ROY G. *The Development of Cognitive Anthropology*. Cambridge: Cambridge University Press, 1995.
- DODSON, JOSEPH R. "Introduction." Pages 1–17 in *Paul and the Second Century*. Edited by MICHAEL F. BIRD and JOSEPH R. DODSON. London: T&T Clark, 2011.
- ESLER, PHILIP F. *Conflict and Identity in Romans: The Social Setting of Paul's Letter*. Minneapolis: Fortress Press, 2003.
- HALLAHAN, MARK. "Prototypes." *Encyclopedia of Social Psychology*, 714–16.
- HARRIS, ZELIG S. "Distributional Structure." *Word & World* 10 (1954): 146–62.
- HASLAM, ALEXANDR S. *Psychology in Organizations: The Social Identity Approach*. 2nd ed. London: Sage Publications, 2004.
- HASLAM, ALEXANDR S., MICHAEL PLATOW, and STEPHEN REICHER. *The New Psychology of Leadership: Identity, Influence, and Power*. Hove: Psychology Press, 2011.
- HOGG, MICHAEL. "Social Identity Theory." Pages 111–36 in *Contemporary Social Psychological Theories*. Edited by PETER J. BURKE. Stanford: Stanford Social Sciences, 2006.
- HUGHES, LORNA, PANOS CONSTANTOPOULOS, and COSTIS DALLAS. "Digital Methods in the Humanities: Understanding and Describing their Use across the Disciplines." Pages 150–70 in *A New Companion to Digital Humanities*. Edited by SUSAN SCHREIBMAN, RAY SIEMENS, and JOHN UNSWORTH. Malden: Wiley-Blackwell, 2016.
- JURAFSKY, DANIEL, and JAMES H. MARTIN. "Vector Semantics and Embeddings." Pages 270–85 in *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Edited by DANIEL JURAFSKY and JAMES H. MARTIN. 3rd ed. Draft. 2021, <https://web.stanford.edu/~jurafsky/slp3/>.
- KARLSHOVEN, KARIANNE, and DEANNE N. DEN HARTOG. "Ethical Leader Behavior and Leader Effectiveness: The Role of Prototypicality and Trust." *International Journal of Leadership Studies* 5/2 (2009): 102–20.

- LANDAUER, THOMAS K., PETER W FOLTZ, and DARRELL LAHAM. "An Introduction to Latent Semantic Analysis." *Discourse Process*. 25/2–3 (1998): 259–84.
- LANDAUER, THOMAS K., DANIELLE S. MCNAMARA, SIMON DENNIS, and WALTER KINTSCH. *Handbook of Latent Semantics*. New York; London: Routledge, 2007.
- LENCI, ALESSANDRO. "Distributional Models of Word Meaning." *Annual Revue of Linguistics* 4 (2018): 151–71.
- LEVY, OMER, YOAV GOLDBERG, and IDO DAGAN. "Improving Distributional Similarity with Lessons Learned from Word Embeddings." *Transactions of the Association for Computational Linguistics* 3 (2015): 211–25.
- LINKA, VOJTĚCH, and VOJTĚCH KAŠE. "Pain and the Body in Corpus Hippocraticum: A Distributional Semantic Analysis." *Digital Classics Online* (2021): 54–71.
- LÜDEMANN, GERD. *Paulus, der Gründer des Christentums*. Springe: zu Klampen Verlag, 2014.
- LUOMANEN, PETRI. "Social Identity, Prototypes, and Exemplars in Gospel Narratives: Methodological Considerations." Pages 85–110 in *From Text to Persuasion: Festschrift in Honour of Professor Lauri Thurén on the Occasion of his 60th Birthday*. Edited by ANSSI VOITILA, NIIL O LAHTI, MIKAEL SUNDQVIST, and LOTTA VALVE. Helsinki: The Finnish Exegetical Society, 2021.
- MARGUERAT, DANIEL. "Paul après Paul: une histoire de réception." *New Testament Studies* 54 (2008): 317–37.
- MARTIN, LUTHER H. "The Promise of Cognitive Science for the Study of Early Christianity." Pages 37–56 in *Explaining Christian Origins and Early Judaism: Contributions from Cognitive and Social Science*. Edited by PETRI LUOMANEN, ILKKA PYYSIÄINEN, and RISTO URO. Leiden: Brill, 2007.
- MESOUDI, ALEX. *Cultural Evolution: How Darwinian Theory Can Explain Human Culture and Synthesize the Social Sciences*. Chicago: University of Chicago Press, 2011.
- MIKOLOV, TOMAS, KAI CHEN, GREG CORRADO, and JEFFREY DEAN. "Efficient Estimation of Word Representations in Vector Space." *ArXiv [Cs.CL]* (2013): 1–12.
- NIKKI, NINA. "Challenges in the Study of the Historical Paul." Pages 185–209 in *Common Ground and Diversity in Early Christian Thought and Study: Essays in Memory of Heikki Räisänen*. Edited by RAIMO HAKOLA, OUTI LEHTIPUU, and NINA NIKKI. Tübingen: Mohr Siebeck, in 2022.
- NIKKI, NINA, and ANTTI VANHOJA. "Righteous Sufferer, Scheming Apostate: Traditions of Paul from a Cultural Evolutionary Perspective." Pages 115–44 in *Religious Identities in Antiquity and the Early Middle Ages: Walking Together and Parting Ways*. Edited by ILKKA LINDSTEDT, NINA NIKKI, and RIIKKA TUORI. Leiden: Brill, 2021.
- PAINTER, JOHN. "Johannine Literature: The Gospel and Letters of John." Pages 344–72 in *The Blackwell Companion to the New Testament*. Edited by DAVID E. AUNE. Oxford: Wiley-Blackwell, 2009.

- PYYSIÄINEN, ILKKA. "The Evolution of Morality and Religion." Pages 44–64 in *Christianity and the Roots of Morality: Philosophical, Early Christian, and Empirical Perspectives*. Edited by PETRI LUOMANEN, ANNE B. PESSI, and ILKKA PYYSIÄINEN. Leiden: Brill, 2017.
- RENSBERGER, DAVID: *As the Apostle Teaches: The Development of the Use of Paul's Letters in Second-Century Christianity*. Ph.D. Dissertation, Yale, 1981.
- RODDA, MARTINA A., MARCO S. G. SENALDI, and ALESSANDRO LENCI. "Panta rei: Tracking Semantic Change with Distributional Semantics in Ancient Greek." *Italian Journal of Computational Linguistics* 3/1 (2017): 11–24.
- SAHLGREN, MAGNUS. "The Distributional Hypothesis." *Rivista Di Linguistica* 20/1 (2008): 33–53.
- SAHLGREN, MAGNUS, and ALESSANDRO LENCI. "The Effects of Data Size and Frequency Range on Distributional Semantic Models." *ArXiv [Cs.CL]* (2006): 1–6.
- SMITH, ELIOT R., and MICHAEL A. ZARATE. "Exemplar and Prototype Use in Social Categorization." *Social Cognition* 8 (1990): 243–62.
- SPERBER, DAN. *Explaining Culture: A Naturalistic Approach*. Oxford: Blackwell Publishers, 1996.
- STERLING, GREGORY. "From Apostle to the Gentiles to Apostle of the Church: Images of Paul at the End of the First Century." *Zeitschrift für die Neutestamentliche Wissenschaft und Kunde der Älteren Kirche* 99/1 (2008): 74–98.
- STOWERS, STANLEY K. "Kinds of Myth, Meals, and Power: Paul and the Corinthians." Pages 105–49 in *Redescribing Paul and the Corinthians*. Edited by RON CAMERON, and MERRIL P. MILLER Atlanta: Society of Biblical Literature, 2011.
- STRANG, GILBERT. *Introduction to Linear Algebra*. Wellesley: Wellesley-Cambridge Press, 2016.
- STRAWBRIDGE, JENNIFER R. *The Pauline Effect*. Berlin: De Gruyter, 2015.
- ŠPICLOVÁ, ZDEŇKA, and VOJTĚCH KAŠE. "Distant Reading of the Gospel of Thomas and the Gospel of John: Reflection of Methodological Aspects of the Use of Digital Technologies in the Research of Biblical Texts" *Open Theology* 6/1 (2020): 423–39.
- TAJFEL, HENRI, and A. L. WILKES. "Classification and Quantitative Judgment." *The British Journal of Psychology* 54 (1963): 101–14.
- TAJFEL, HENRI, M. G. BILLIG, R. P. BUNDY, and CLAUDE FLAMENT. "Social Categorization and Intergroup Behaviour," *European Journal of Social Psychology* 1 (1971): 149–78.
- TIWALD, MARKUS. *Hebräer von Hebräern. Paulus auf dem Hintergrund frühjüdischer Argumentation und biblischer Interpretation*. Freiburg: Herder, 2008.
- TURNER, JOHN C., and KATHERINE J. REYNOLDS. "Self-Categorization Theory." Pages 399–417 in *Handbook of Theories of Social Psychology*. Vol. 2. Edited by PAUL A. M. LANGE, ARIE W. KRUGLANSKI, and E. TORY HIGGINS. London: Sage, 2012.

WHITE, BENJAMIN L. *Remembering Paul: Ancient and Modern Contests over the Image of the Apostle*. New York: Oxford University Press, 2014.

WHITEHOUSE, HARVEY. *Modes of Religiosity: A Cognitive Theory of Religious Transmission*. Walnut Creek, CA: AltaMira Press, 2004.