Cambridge Semitic Languages and Cultures

Linguistic Theory and the Biblical Text

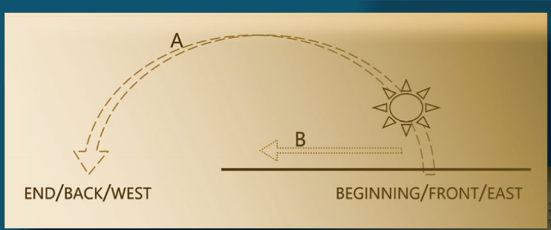
Edited by William A. Ross and Elizabeth Robar

Cognitive Linguistic Theory

Functional Grammar

Historical Linguistics

οὗτος ἦν ἐν ἀρχῇ πρὸς τὸν θεόν. πάντα δι' αὐτοῦ ἐγένετο, καὶ χωρὶς αὐτοῦ ἐγένετο οὐδὲ ἕν ὃ γέγονεν. ἐν αὐτῷ ζωὴ ἦν, καὶ ἡ ζωὴ ἦν τὸ φῶς τῶν ἀνθρώπων· καὶ τὸ φῶς ἐν τῇ σκοτίᾳ φαίνει, καὶ ἡ σκοτία αὐτὸ οὐ κατέλαβεν.



Complexity Theory

Generative Linguistics

This was in the beginning with God. All things through him came to be, and outside of him came to be nothing that came to be. In him was life and the life was the light of people, and the light shines in the darkness and the darkness did not comprehend it (Jn. 1:2–5)

Pragmatics of Information Structure

Computational Linguistic Analysis

UNIVERSITY OF CAMBRIDGE
Faculty of Asian and Middle Eastern Studies

https://www.openbookpublishers.com

William A. Ross and Elizabeth Robar (eds), *Linguistic Theory and the Biblical Text*. Cambridge, UK: Open Book Publishers, 2023, https://doi.org/10.11647/OBP.0358

Any digital material and resources associated with this volume will be available at https://doi.org/10.11647/OBP.0358#resources

Semitic Languages and Cultures 20.

ISSN (print): 2632-6906
ISSN (digital): 2632-6914
ISBN Paperback: 978-1-80511-108-5
ISBN Hardback: 978-1-80511-109-2
ISBN Digital (PDF): 978-1-80511-110-8
DOI: 10.11647/OBP.0358

Cover image: A section of Cisneros' original complutensian polyglot Bible, https://en.wikipedia.org/wiki/File:Cisneros%27_original_complutensian_polyglot_Bible_-2.jpg; additional text and diagrams created by authors.

Cover design by Jeevanjot Kaur Nagpal

The main fonts used in this volume are SIL Charis, SBL Hebrew, and SBL Greek.

HISTORICAL LINGUISTICS AND THE BIBLICAL LANGUAGES

Kaspars Ozoliņš

The student of Biblical Hebrew can observe with relative ease that the language of the Old Testament not only differs from genre to genre and author to author, but also that it exhibits variation across its diachronically diverse writings.[1] Likewise, the classicist engaging with the Greek of the New Testament will note that it differs in important ways from the Greek of earlier time periods. Language change is a fact of life, and the biblical text is certainly no exception to this. Yet the rigorous application of the findings and methodology of historical linguistics to biblical studies has been less prevalent and less thorough than might be expected. This chapter provides a broad introduction to historical linguistics—the study of language change—especially as applied to biblical studies.

[1] This point, like almost anything in biblical scholarship, is disputed. Yet even those who would argue that the Old Testament is entirely the product of a single era still claim that certain authors employed an *archaising* style that reflects a diachronically earlier state of the language. See, e.g., Young, Rezetko, and Ehrensvärd (2008).

 https://doi.org/10.11647/OBP.0358.04

1.0. The History of Historical Linguistics

Although it has important historical precursors, modern linguistics developed in the wake of the Scientific Revolution of the sixteenth and seventeenth centuries in Europe. Historical linguistics in particular plays a special role in the *historical* development of linguistic inquiry, since the first systematic investigations into the nature of language were undertaken by scholars specifically interested in language change and the relationships among languages.[2]

Initial linguistic investigation, especially in the nineteenth century, was focused on the classification and comparison of the older languages of Europe and southwest Asia (the language family that came to be known as Indo-European). The growing impetus for these investigations occurred towards the end of the eighteenth century, as Europeans gradually became familiar with the languages and cultures of ancient India. In this connection, a famous statement from a speech delivered by Sir William Jones to the Asiatic Society in 1786 has been frequently quoted (Fortson 2010, 9):

> The *Sanscrit* language, whatever be its antiquity, is of a wonderful structure; more perfect than the *Greek*, more copious than the *Latin*, and more exquisitely refined than

[2] There are also other important terms that are used to describe the field. One is *comparative* linguistics, which emphasises the comparative method as a means for reconstructing linguistic history (see §2.2.3). Another is *diachronic* linguistics, a term that is perhaps less ambiguous than *historical linguistics*, since the latter could possibly be interpreted as the study of the history of the field of linguistics *in toto*.

> either, yet bearing to both of them a stronger affinity, both in the roots of verbs and the forms of grammar, than could possibly have been produced by accident; so strong indeed, that no philologer could examine them all three, without believing them to have sprung from some common source, which, perhaps, no longer exists; there is a similar reason, though not quite so forcible, for supposing that both the *Gothick* and the *Celtick*, though blended with a very different idiom, had the same origin with the *Sanscrit*; and the old *Persian* might be added to the same family.

Jones' somewhat hyperbolic and subjective descriptions of these languages might strike the modern reader as being rather quaint. Nevertheless, a fundamental axiom of historical linguistics posits that descendent languages are related to one another via a now-lost parent language. Even geographically and chronologically diverse languages can be unexpectedly related. Table 1 compares cognates in English, ancient Greek, Latvian (a Baltic language spoken in Northern Europe), classical Latin, and Hittite (an extinct language originally spoken in present-day Turkey).

Table 1: Cognates in select Indo-European languages

English	**Greek**	**Latvian**	**Latin**	**Hittite**
water	ὑδώρ	*ūdens*	Umbr. *utur*	*wāt-ar, -en-*
brother	φράτηρ	*brālis*	*frāter*	*negnaš*
three	τρεῖς	*trīs*	*trēs*	*tereš*
cow	βοῦς	*govs*	*bōs*	GUD
(I) am	εἰμί	*esmu*	*sum*	*ēšmi*
night	νύξ	*nakts*	*nox*	*nekuz*

These languages exhibit similar basic vocabulary even though they range from the second millennium BC to the present day and from Turkey (Hittite) to Northern Europe (Latvian). The close resemblances demonstrate a genetic link between these

geographically diverse languages (two modern, three ancient) and rule out accidental resemblance or borrowing. Following Jones, nineteenth-century scholars were preoccupied with the relationships of these similar languages and they endeavoured to reconstruct their parent language, which came to be called Proto-Indo-European.

However, after the appearance of Ferdinand de Saussure's classic *Cours de linguistique générale* in 1916, the diachronic dimension of language study became more of a minor player in linguistics and was replaced by a primary focus in the twentieth century on the *synchronic* study of language.[3] Such a neat bifurcation is of course an oversimplification, for diachrony and synchrony are always in relation with one another in language. In fact, a renewed appreciation for the centrality of diachrony in linguistics has grown in recent decades (see, for example, Bybee 2010; Hartmann 2021).

An overview of the history of the discipline will ideally focus on the study of the biblical languages, which are naturally of primary interest to biblical scholars. The languages of the Old and New Testaments (Hebrew, Aramaic, and Greek) belong to two linguistic families: Semitic and Indo-European. As such, it is appropriate to approach the field of historical linguistics from the perspective of these two language families. As it happens, most work in historical linguistics has thus far been done within these same two language families.

---

[3] Ironically, some of Saussure's greatest early work was in the area of historical linguistics (see §1.1. below).

1.1. Indo-European Studies

Koine Greek, the language of the New Testament era, belongs to the Greek (or Hellenic) branch of the Indo-European family of languages. The designation *Koine* (κοινή meaning 'common') refers to the *lingua franca* status of the Attic-Ionic dialect that spread throughout the Greek empire, launched by the conquests of Alexander the Great. Two earlier major periods in the written history of Greek may be identified: Classical/Archaic Greek and Mycenaean Greek (late second millennium BCE). The latter is of interest to Indo-Europeanists and other scholars because of its early attestation and its peculiar writing system, deciphered in the 1950s by Michael Ventris (see Ventris and Chadwick 1973).

The Indo-European family of languages contains several hundred languages (depending on classification) and is the largest such linguistic grouping in the world, by number of speakers. The designation *Indo-European* is intended to (roughly) encompass most of the languages spoken in the lands of Europe and Southwest Asia.[4] The major subfamilies are (in order of earliest written attestation): Anatolian, Indic, Iranian, Greek (or

[4] The German term *indogermanisch* is somewhat more precise, since the westernmost territory (namely, Scandinavia and Iceland) is inhabited by speakers of Germanic languages. (The equivalent term *Indo-Germanic* is now outdated in English.) The extinct Tocharian languages are a geographical outlier, having been spoken in the modern Chinese province of Xinjiang (in central Asia). A few languages in Europe are non-Indo-European, notably the Finno-Ugric languages Estonian, Finnish, and Hungarian. Basque (Spain and France) is a famous language isolate, while Maltese is a Semitic language.

Hellenic), Italic, Celtic, Germanic, Armenian, Tocharian, Slavic, Baltic, Albanian.[5] Several Indo-European languages enjoy a lengthy period of written attestation (rivalled only by some Semitic languages). The earliest attested written records are from the extinct Anatolian family (spoken in present-day Turkey), in particular, Hittite, dating back roughly to the Late Bronze Age.[6]

Early scholarly efforts at reconstructing Proto-Indo-European were influenced by the prestige of Sanskrit, the classical language of India (equivalent in status to that of Latin in Europe). This led scholars to reconstruct Proto-Indo-European with broadly similar features, especially in its vowel system. Subsequent study revealed that Greek, with its inherited five-vowel distinction, was in fact more archaic than the three-vowel distinction of Sanskrit.[7]

---

[5] Several larger subgroupings are typically made by scholars. The most clear is Indo-Iranian (Indic together with Iranian). Most scholars would also group together Baltic and Slavic (Balto-Slavic). Much more controversial is Italo-Celtic. Additionally, certain shared isoglosses lead many scholars to group together Balto-Slavic and Germanic in a northern dialect group, as well as Greek, Indo-Iranian, Armenian, and the poorly attested Phrygian into a southern dialect group.

[6] The earliest attested record of any Indo-European material is found in the form of Hittite loanwords in Old Assyrian texts from Kaniš, dating to the Middle Bronze Age (18th century BCE). See Kloekhorst (2019).

[7] Indo-Iranian (the sub-branch in which Sanskrit is located) distinguishes three vowel qualities (*ĭ̯, ă, ŭ)*, with **ĕ* and **ŏ* having merged to *ă*. Note that the asterisk symbol is universally used in historical linguistics to denote a linguistic reconstruction that is prehistoric (i.e., prior to written records).

A major milestone in Indo-European studies was the reconstruction of a class of consonants called *laryngeals*, first hypothesised by Ferdinand de Saussure (1879). His rather brilliant reasoning is akin to the method now known as internal reconstruction.[8] The empirical existence of these consonants was confirmed by the discovery of Hittite (of the Anatolian branch) in the early twentieth century. The laryngeals were partly preserved intact in Anatolian, while in other branches they were lost, although they left various important traces (such as adjacent vowel colouring, compensatory lengthening, and other effects). Most scholars today reconstruct three laryngeals, conventionally designated *h_1, *h_2, *h_3. Their actual phonetic values are still unknown, though many scholars hypothesise that they were fricatives of some type, perhaps glottal and pharyngeal.[9]

---

[8] In retrospect, one can see a kinship between Saussure's later influential ideas on structuralism and the reasoning he employed here.

[9] An illustration of the crucial role that laryngeals play in Indo-European linguistics is seen, for example, in the interplay of morphological vowel alternations (called *ablaut*) and root structure in the singular and plural perfect stems of verbs like λείπω 'leave' (e.g., λέ-λοιπ-α and λε-λίπ-ομεν). Notice that the singular features an *o*-grade vowel in the root, whereas the plural lacks this. In another pair of forms, such as δέ-δω-κα (1 sg.) and δέ-δο-μεν (1 pl.), the difference observed is a vowel length distinction. However, with the postulation of a laryngeal for this second verbal root, the morphological symmetry is restored: *de-doh_3- (sg.) ~ *de-dh_3- (pl.). The loss of the laryngeal was not without effects: compensatory lengthening in the singular stem (*de-doh_3- > δεδω-) and vowel epenthesis in the plural stem (*de-dh_3- > δεδο-). Compare the reconstructed stems for λέλοιπα ~ λελίπομεν: *le-$loik^w$- (sg.) ~ *le-lik^w- (pl.).

A landmark multi-volume work appearing towards the end of the nineteenth century (1886–1900) is Brugmann and Delbrück's *Grundriß der vergleichenden Grammatik der indogermanischen Sprachen.* This represented the state of current knowledge at the time, prior to the discovery of Anatolian and Tocharian early in the following century. As it turns out, a considerable amount of time would pass before the insights of Tocharian, and especially Anatolian, were to alter the picture of Indo-European.

In recent decades it has become clear that Anatolian was the first to break off as a separate speech community (followed next by Tocharian). Some of the morphologically rich categories common to (Vedic) Sanskrit and Greek are now thought to have been later innovations shared by the core inner-IE branches after the breaking off of Anatolian (which appears to have lacked them).

A major ongoing area of study is the history and status of the so-called *-ḫi* and *-mi* conjugations in Anatolian and Indo-European. In core Indo-European, only the familiar *-mi* conjugation (cf. Gk δίδωμι 'give') is attested, though, interestingly, the long 1 sg. ending -ω of thematic verbs is standardly viewed as going back to *-*o*-h_2 (with the laryngeal ending ultimately related to the *-ḫi* conjugation).[10] Scholarship has demonstrated that the perfect tense-stem and the middle voice are clearly historically linked to the Anatolian *-ḫi* conjugation. The exact relationship is

[10] The so-called thematic verbs (much more common in Greek) belong to the same class as *-mi* verbs; compare the endings of Gk δίδωμι with the thematic verb *bhar-a-* 'bear; carry' in Sanskrit: 1 sg. *bhár-ā-mi* (from earlier **b*^{h}*arā*, cognate with Gk φέρω), 2 sg. *bhár-a-si*, 3 sg. *bhár-a-ti*.

controversial and has been much debated. A comprehensive theory of the Indo-European verb, developed by Jay Jasanoff since the 1970s, has been steadily gaining adherents (see Jasanoff 2003).

1.2. Semitic Studies

Hebrew and Aramaic, the languages of the Old Testament, belong to the Semitic family of languages, itself part of a larger grouping called Afro-Asiatic (which links Semitic with a number of languages spoken in northern and central Africa).[11] The immediate sub-branch to which both Hebrew and Aramaic belong is designated Northwest Semitic. Also classified as Northwest Semitic is Ugaritic, an extinct language spoken at Ugarit, a city state on the northern Levantine coast.[12]

Semitic linguistics has been shaped by particular constraints and methodological challenges. Semitic languages are characterised by a distinctive concatenative morphology, which features (mostly) triconsonantal verbal roots (C_1-C_2-C_3) combined with particular vowel patterns. For example, *k-t-b* 'write'

[11] Afro-Asiatic linguistics is still in its infancy, and many of the non-Semitic languages are poorly known and poorly attested. Additionally, the specific linguistic affiliation of the sub-branches is controversial.

[12] Discovered by accident in 1928, the archaeological site of Ugarit (Tell Ras Shamra) has yielded thousands of cuneiform tablets and other significant artefacts. Ugaritic studies over the past century has offered important linguistic, cultural, and theological insights into the Old Testament. For a broad introduction to Ugaritic studies, see Watson and Wyatt (1999).

combined with a particular vowel pattern to form a verbal adjective in Proto-Semitic: **kātib-* (Hebrew *qal* כֹּתֵב, Aramaic *peʿal* כָּתֵב, Ugaritic G-stem *kātibu*, Arabic form I *kātibun*). A different vowel pattern with the same root yields a different grammatical form, e.g., the G-stem infinitive **katāb-*. Unfortunately, the reconstruction of the various vocalic patterns for both Proto-Semitic and intermediate branches is complicated by the fact that many ancient Semitic writing systems tended to leave out vowels. The writings of the Old Testament, for example, were originally written with a (largely) consonantal writing system, which was only later supplemented by the medieval Masoretic notation system of vowels (*niqqud*).

For example, the exact vocalisation of the Ugaritic word *šd* 'field' must be ascertained from other evidence (since the native Ugaritic orthography is largely consonantal). Its reconstructed form in Proto-Semitic featured an intervocalic glide (between the second root vowel and the case vowel): **śadayum*. The outcome of these so-called *triphthong* sequences varied from language to language.[13] In Biblical Hebrew, for example, III-ה nouns (originally derived from III-*y*/*w* roots) feature word-final *segol* in the absolute (e.g., שָׂדֶה 'field'), with regular loss of final **m*

[13] The term *triphthong* (derived from *diphthong*) is somewhat of a misnomer, as these sequences are better described from the standpoint of Semitic phonology as two vowels separated by an intervocalic glide: V_1wV_2 and V_1yV_2.

(originally used in Proto-Semitic to indicate unbound nominals).[14] Fortunately, among the archaeological finds at Ugarit are numerous cuneiform tablets which contain lists of Ugaritic words spelled with the logosyllabic orthography peculiar to Akkadian. One such example is *ša-du-u* 'field' (RS 20.123+), indicating to us that the outcome of this triphthong sequence in Ugaritic was /-ū/.

Another challenge historical linguists face in dealing with Semitic languages has been determining linguistic subgroupings. The most important criterion for establishing a subgroup is shared innovations between languages. Unfortunately, the significant linguistic contact between Semitic languages in the relatively small Middle East region has meant that genuine shared innovations are often hard to distinguish from borrowings. For more detail about the criteria of shared innovations, as well as the two major models of linguistic affiliation (the tree and wave models), see §2.1.4.

Important work by Robert Hetzron in the 1970s led to mainstream adoption of a Central Semitic node which, in the current standard view, includes three main sub-branches: Northwest Semitic, Arabic, and Old South Arabian (Ṣayhadic). The primary shared innovation justifying this node was the replacement of the imperfective *yaqaṭṭal* verbal form with the *yaqṭulu* form (the ancestor of the familiar *yiqṭol* form in Hebrew).

Some scholars have rejected Hetzron's model, pointing out that there appears to be a set of shared features justifying a

[14] The less common biform שְׂדִי 'field' indicates the original III-*y* root shape.

'south' Semitic branch.[15] For example, Arabic, Old South Arabian, and Modern South Arabian, as well as Ethiopian, all exhibit a distinctive $*p > f$ sound change. Furthermore, the phenomenon of so-called 'broken' (or internal) plurals is widespread among these languages, whereas it is less prevalent (and never occurs without an overt plural suffix) in Northwest Semitic. Huehnergard and Rubin (2011) argue persuasively that these (and other) features are in some cases trivial, while in other cases they provide evidence for intensive contact between the 'southern' languages *after* the split of Central Semitic from the rest of West Semitic.[16]

The status of some other sub-groupings in Semitic also continues to be debated. Potentially of interest to biblical scholars is the ongoing discussion about the status of Aramaic, Ugaritic, and the Canaanite dialects (including Hebrew) within Northwest Semitic (see Pat-El and Wilson-Wright 2018).

2.0. Key Theoretical Commitments and Major Concepts

In historical linguistics, several crucial theoretical commitments are a necessary foundation for the investigation of language change in all its dimensions. These are described below, followed by an exploration of some of the field's major concepts.

---

[15] The 'south Semitic' view was standard among earlier generations of Semitists.

[16] Kogan's recent detailed lexical study (2015) of Semitic isoglosses, including between Arabic and Northwest Semitic, finds additional support for a Central Semitic node.

2.1. Theoretical Commitments

Any scientific discipline contains within it various assumptions (sometimes unstated or otherwise unknown to non-specialists) without which its specific concepts and methods could not be coherent or justified. Four such primary commitments are sketched here.

2.1.1. Pervasiveness of Language Change across All Levels

Human language is the object of all linguistic investigation. A *sine qua non* within the subfield of historical linguistics is the assumption that *all* natural human languages are everywhere and always subject to continuous and perpetual language change at all levels. Consider the somewhat idealised Figure 1 below. All subsystems within a given language (whether phonetics, phonology, etc.) are subject to change over time.[17] This process is generally gradual (although it is not always uniform in intensity), such that speakers do not generally perceive the change to be very great within their own lifetimes. Over a larger timespan, however, such as centuries or millennia, the accumulated changes can result in drastic differences.

[17] This neat segmentation is of course conceptual, and in reality all schools of linguistic thought acknowledge that the boundaries between language domains are highly permeable and interconnected (hence terms such as *morphosyntax*, *pragmatico-semantic*, etc.).

Figure 1: Multi-level language change

Time ⟶			
Language states			
Phonetics$_1$	**Phonetics$_2$**	**…**	**Phonetics$_x$**
Phonology$_1$	**Phonology$_2$**	**…**	**Phonology$_x$**
Morphology$_1$	**Morphology$_2$**	**…**	**Morphology$_x$**
Syntax$_1$	**Syntax$_2$**	**…**	**Syntax$_x$**
Semantics$_1$	**Semantics$_2$**	**…**	**Semantics$_x$**
Pragmatics$_1$	**Pragmatics$_2$**	**…**	**Pragmatics$_x$**

Historical linguists assume that any natural language—even if poorly understood or as yet unknown—will be subject to this phenomenon.[18]

2.1.2. The Regularity of Sound Change

The study of sound change lies at the heart of the discipline of historical linguistics. One key reason for this is the *regularity* of sound change in all human languages—a far-reaching theoretical commitment. Early nineteenth-century observers of language change (e.g., Rasmus Rask, August Schleicher, Franz Bopp, etc.)

[18] In fact, one of the distinguishing factors in classifying languages as 'living' (as opposed to 'extinct' or 'dead') is whether they are (or were) in the process of undergoing change. Nevertheless, the classification of a language as 'dead' is somewhat controversial, given phenomena such as the widespread use of Latin in medieval Europe, for example.

noted systematic sound correspondences between sets of cognate vocabulary, leading them to postulate consistent sound laws to account for these differences (see §2.2.3. on the comparative method). The consistency of the sound correspondences was seen to apply across the board in a given language (for example, the same sound change affected verbs, nouns, adjectives, etc., without exception). The regularity of sound change is largely what makes linguistic reconstruction of prehistoric language states possible. In its strongest form, this claim is linked to the nineteenth-century German Neogrammarians (Ger. *Junggrammatiker*), especially expounded in Osthoff and Brugmann (1878).

The actual degree of regularity of sound changes is a topic of some controversy in historical linguistics. Some types of processes (for example, dissimilation and metathesis) appear to be less than fully regular. Closely tied to this issue is the question of the actual locus of sound change. Does it take place mainly during the language acquisition process of children? Or is it rather to be located in fluent adult speakers? Or both? What is the relationship and interplay between 'regular' sound change and other types of change in language (especially analogy)? For our purposes, it is enough to note that most scholars accept a *general* regularity in language sound change (what could be perhaps termed 'qualified exceptionlessness').

2.1.3. Linguistic Reconstruction from Sound Change

Linguistic reconstruction is the systematic recovery of chronologically earlier language states (often thousands of years earlier). This process includes proto-languages, intermediate daughter

branches, and all other intermediate language states. The end goal of linguistic reconstruction is to provide as comprehensive a history as possible for a given language within its language family. This activity is a constant work-in-progress for scholars engaging with each language family, and in some cases, scholarship has been working towards this goal for centuries.

Although all levels of language are subject to change over time as seen above, the investigation of sound change (phonetics and phonology) in language is agreed to be the major *starting* point and the foundation for the entire enterprise of linguistic reconstruction. This is partly due to the regularity of sound change, discussed above, but there are additional reasons. Language subsystems exist in a roughly hierarchical arrangement, hence sound change naturally becomes the foundation for other, higher-level linguistic reconstruction. This is because higher language domains (such as morphosyntax) ultimately terminate at the phonological/phonetic domains.

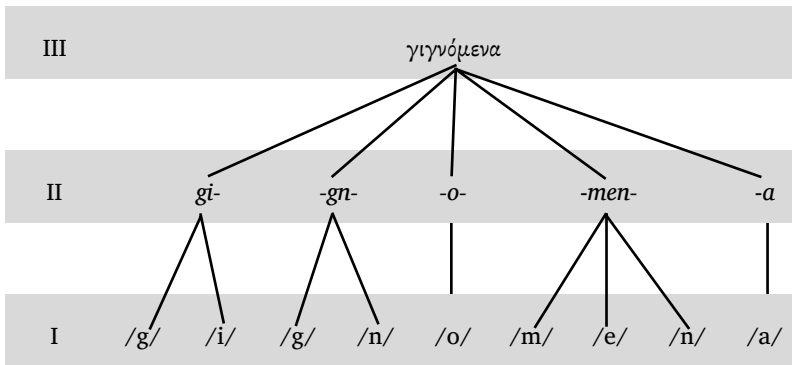
All syntactic constructions, for example, consist of a discrete sequence of morphemes. Each morpheme in turn consists of a discrete sequence of phonemes (contrastive speech sounds within a given language).[19] Finally, each phoneme is an abstract representation of a set of phones (the final, phonetic domain), whose distribution in the language is governed by complex language-specific rules.

---

[19] Naturally, some morphemes in languages can consist of only a single phoneme; but a conceptual distinction between the two is still necessary.

As an example, Figure 2 below illustrates the hierarchical relationships in the Classical Greek present participle γιγνόμενα ‘the things taking place’.[20] Notice that the complete word (level III) consists of five discrete morphemes in the morphological level (II), of which the first, a present-stem reduplicant *gi-*, is especiallly noteworthy.[21] Finally, the five morphemes comprise nine distinctive phonemes in the phonological level (I).

Figure 2: Morphological hierarchy of γιγνόμενα



However, a consequence of this bottom-up schema is that any sound changes will have cascading consequences for higher-level domains. For example, the historic loss of the preconsonantal phoneme /g/ in post-classical Greek (/gignomena/ > /gīnomena/) occasioned compensatory lengthening of /i/ > [i:]. This is evidenced by the consistent spelling of {ει} in this and related

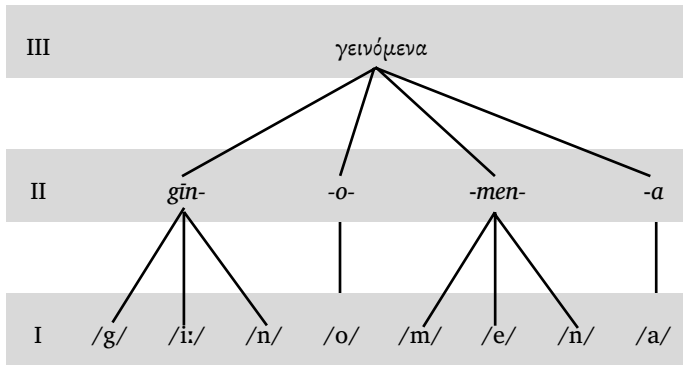
[20] This schema is necessarily simplified, and therefore does not capture the internal hierarchical structures at each of the three levels.

[21] Not to be confused with the perfect-stem reduplicant C_1e- (cf. γέγονα).

present-tense forms in the biblical manuscript tradition, e.g., Luke 9.7: P^{75} א D {γεινομενα}.[22]

This change in the language, although originating at the phonetic/phonological level, has consequences for higher levels of the language. In place of a clear phonological boundary between the present-stem reduplicant *gi-* and the root *-gn-*, one observes a new opaque present-stem *gīn-*. Speakers would now internalise a quasi-suppletive set of verbal stems (e.g., present *gīn-* [γείνεται], aorist *-gen-* [ἐγένετο], perfect *-gon-* [γέγονε]). This new structure is depicted in Figure 3 below.

Figure 3: Revised morphological hierarchy of γιγνόμενα



From this illustration one sees the foundational role that sound change plays in linguistic structures, as well as how it enjoys a certain priority in historical linguistics. Although language change can occur in any language domain, the most secure and

[22] See Williams (2018) for a defence of the view that {ι} ~ {ει} spelling interchanges are not uniformly haphazard, as is usually assumed (part of a larger phenomenon known as 'itacism').

logical foundation for *linguistic reconstruction* is the phonetics/phonology domain (i.e., sound change).[23]

2.1.4. Language Branching and Speech Communities

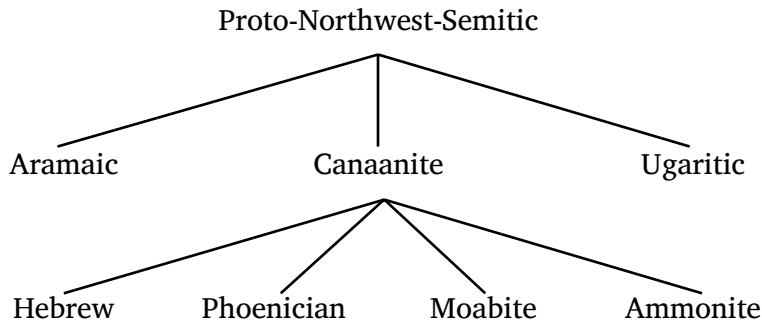
A language or dialect is a more or less unified speech community with a set of shared (though unconsciously accepted) conventions that constitute a particular *grammar*. A grammar may be conceived of as a collection of form-and-meaning pairings.

The language of such a speech community (say, speech community *x*) undergoes innovations over time, such that all of its members (= speakers) essentially hold together linguistically. Suppose, however, that a separation takes place within this speech community, such that in place of *x* there are now two separate speech communities, *y* and *z* (typically separated by geography). An important consequence of this development is that all subsequent language change in speech communities *y* and *z* will be distinctive, unless language contact is reestablished between the communities. This is the central concept known as language branching. See Figure 4 below for a standard model of Northwest Semitic.

---

[23] In practice, this usually means that the most developed linguistic reconstructions tend to be phonological in nature, followed by morpho-syntax and semantics (though these other, less well-developed areas of study in Indo-European historical linguistics have experienced significant advances in recent decades).

Figure 4: Standard model of Northwest Semitic



Over time, the iterative branching process yields a family 'tree', where each 'twig' (or dialect) is linked to a larger series of 'branches' (or distinctive languages), while all branches ultimately converge into a single 'trunk' (the proto-language).[24] While all speech communities are ultimately related to one another within a given family, a hierarchical structure relates each language in a more precise way that captures the individual relationships between any two members.

Once separated, the different changes that two speech communities undergo are generally not predictable beforehand. Within a given language family, when speech communities are no longer in contact with one another, the changes they subsequently undergo will not always be identical.[25] The more time

[24] This specialised field, known as *cladistics* (from Gk κλάδος 'branch'), is also important for several other non-linguistic scientific disciplines, for example, textual criticism.

[25] An active topic of discussion among historical linguists is the degree to which linguistic trajectories can be mapped out, given the common starting point of separated speech communities. In part, this depends on whether language change is best characterised as *prophylactic* or

that passes, the more distinctive each speech community becomes. Here, the imprecise, yet useful, terms *dialect* and *language* apply. Given enough time, separated speech communities sharing a language will develop distinctive dialects. These dialects will grow progressively less mutually intelligible, until, at some point, they are conventionally considered distinctive languages.[26]

In reality, this neat model is complicated by the phenomenon of language contact between distinctive speech communities. Prolonged language contact (whether minor or intensive) can lead to new shared linguistic innovations over time. These innovations (or *isoglosses*) can create challenges in the establishment of linguistic affiliations between the languages of a language family (as even in the example of Northwest Semitic; see §1.2).

To capture this reality, an alternative (but complementary) *wave model* has been used by historical linguists to account for features and changes that spread throughout any speech communities in contact with one another (whether they are linguistically homogeneous or not). The metaphor of a spreading wave (or better: ripples in a pond) nicely illustrates how language change

---

reparative. In other words, does language change primarily occur in order to avoid certain phenomena that are unconsciously disfavoured by speakers, or, rather, does language change occur as a subsequent tool for repairing already existing changes which themselves are disfavoured?

[26] Something of the imprecision in these categories may be observed when considering that the Scandinavian *languages* (Swedish, Norwegian, and Danish) are largely mutually intelligible, yet the numerous *dialects* of Modern Arabic or the varieties of Chinese are not.

is often much more complex than imagined. The wave model is popular with dialectologists, especially in the context of the study of a dialect continuum (or language continuum). Many linguists today do not view the tree and wave models as mutually exclusive, but in fact as complementary, the former accounting for *genetic* relationships, the latter accounting for subsequent external linguistic influences.

2.2. Major Concepts

Building on the previous section, the following is a necessarily brief exploration of some of the methods and tools used by historical linguists in their exploration of language change.

2.2.1. The Role of Philology

As we have seen, linguistics was birthed in the context of the study of ancient languages, which is described in part as philology. Nevertheless, historical linguistics eventually grew to be a separate scientific discipline in its own right, distinct from philology. Consequently, the term *philology* nowadays more properly describes the various tools employed by scholars (not just linguists) in order to study ancient texts. Philology, which has been neatly encapsulated as the art of reading slowly, includes diverse disciplines such as archaeology, paleography, textual criticism, and other historical fields of inquiry.

These tools are still highly relevant for historical linguists, since scholars prize the early written records of a language as a means of conducting linguistic reconstruction. Unfortunately, the

quality and accuracy of such data are uneven for the documentation of a historically prior stage of a spoken language (irrespective of whether that language is currently extinct or living). Knowledge about writing systems (paleography, etc.) is vital for understanding the conventions that native (and non-native) speakers chose in representing a given language. Furthermore, if an ancient text can only be accessed by scholars as the product of a chain of textual transmission (i.e., scribal copying), the discipline of textual criticism must be employed in order to sort out genuine linguistic data from non-linguistic scribal errors or supposed corrections that have arisen in the manuscript tradition.

In fact, any written representation of language, whether ancient or modern, is always incomplete and imprecise due to various factors. To begin with, speakers do not require or depend on a precise phonetic (or even phonological) transcription of language in order to achieve comprehension. An illustration of various levels of transcriptional detail in the Masoretic system and the International Phonetic Alphabet (IPA) is given in the example below.

(1) (a) אפדנו (‘his palace’; Dan. 11.45)
Consonantal representation

(b) אַפַּדְנוֹ
Approximate phonemic representation

(c) /ʔapːadno/
Broad (phonemic) IPA transcription

(d) אַפַּדְנוֹ
Approximate suprasegmental representation

(e) [ʔap.pað.ˈnoː]
Narrow (phonetic) IPA transcription

Varying levels of detail and accuracy are presented for the hapax אַפַּדְנֹו ‘his palace’ (Dan. 11.45) in the example, on a spectrum from (a)–(e). Notice, however, that each non-IPA representation is not fully consistent (hence the designation ‘approximate’). This is an artefact of complex historical factors. For example, the consonantal representation is not fully consonantal, since it includes a final vowel letter (ו–).[27] Additionally, while the Masoretic system itself accounts for vowel quality and generally indicates suprasegmental stress, it does not record vowel length, even though this is generally thought to have been at least partly phonemic in Tiberian Hebrew.[28]

What is particularly interesting about this example, however, is that it is plausibly one of a select few examples of a non-aspirated /p/ in the entire Masoretic reading tradition of the Old Testament. We know from Latin and Greek transcriptions that the normal pronunciation of פ appears to have been not [p], but an aspirated stop [p^{h}], e.g., וְאַרְפַּכְשַׁד (Gen 10:22) // Αρφαξαδ. Geoffrey Khan adduces testimony from the medieval grammarian Saadya Gaon for the hapax legomenon אַפַּדְנֹו ‘his palace’ at Dan. 11.45,

---

[27] The use of vowel letters can be documented from the early first millennium BCE, and it is generally thought that the initial point of contact was the monophthongisation of particular diphthongs in Northwest Semitic. Thus, for example, **mawt* {mwt} > *mōt* {mwt} ‘death’ (cons.), where *w* came to represent *ō*.

[28] There has been something of a revolution in our understanding of the phonology of Tiberian Hebrew in recent decades, largely thanks to the detailed work of Geoffrey Khan. For an excellent brief overview of the history of scholarly interpretation of the Masoretic vocalic system, see Suchard (2018).

which he describes as "between *bet* and *pe* with *dagesh*" (Khan 2020, 215). This appears to be an attempt to capture the hearer's perception of a non-aspirated stop (since *bet* was not aspirated). Khan argues that the reading tradition for this loanword essentially inherited its non-aspiration from its Old Persian source (which lacked aspiration), probably with pharyngealisation, as a phonetic approximation of the loanword in Hebrew. Incredibly, there is even some corroborating evidence for this pronunciation from Latin and Greek transcriptions: e.g., απαδανω (Theodoretus, fifth century AD).[29]

As can be seen from just one example, historical linguists rely heavily on the methods and findings of philology in order to do good linguistic work on language states that are inaccessible with traditional field work methodology. The soundness of their reconstructions and theoretical models in large part depends on sound philological work.

2.2.2. Borrowing[30]

Linguistic borrowing is in fact a pervasive feature of many distinctive speech communities in human societies, since they are frequently in contact (sometimes extensive contact) with other language communities. This contact can take many different

[29] For the full argument, as well as more complete evidence, see Khan (2020, 214–20).

[30] This discussion on borrowing is deliberately placed prior to the sections that follow because discerning and eliminating borrowed linguistic material is an important criterion for establishing legitimate correspondence sets when using the comparative method.

forms. Sometimes it is limited and sporadic; at other times and contexts it can be significant, such that a high level of bilingualism exists between two speech communities.[31] The term *borrowing* generally refers to lexemes and semantic notions which are transferred from one speech community to another. (Note that this excludes other types of linguistic influence, such as syntax, which are no less genuine.)

The actual process of borrowing, like with other diachronic aspects of language, is not instantaneous. A borrowed lexeme begins life as a recognisably foreign element that over time may become progressively integrated into a speech community, such that eventually its users will no longer perceive it as foreign, but perfectly native. When examining lexical material in a speech community, or even in an ancient text (say, that of the Bible), it is important to ascertain where on the scale of adoption a given lexical item is diachronically situated (at the time of its usage). Depending on the analysis, foreign words in a text might be described either as *loanwords* or alternatively as an example of *code-switching*.

A helpful illustration for understanding borrowing is the so-called 'Latinisms' in the Gospel of Mark. As shown in Table 2, these include the following (adapted from Zeichmann 2017):

---

[31] Various complex factors (linguistic and non-linguistic) can play a role in determining the degree of contact, such as the linguistic relatedness of the two speech communities, their relative social ranking, and the degree of societal integration between them.

Table 2: Latin-Greek Borrowing

Latin	Greek	References (GMark)
grabātus ‘cot’	κράβατος	2.4, 9, 11, 12; 6.55
modius ‘a measure’	μόδιον	4.21
legiō ‘legion’	λεγιών	5.9, 15
speculātor ‘scout’	σπεκουλάτωρ	6.27
dēnārius ‘a day’s wage’	δηνάριον	6.37; 12.15; 14.5
pugnus ‘fist’	πυγμή	7.3
sextārius ‘a measure’	ξέστης	7.4
cēnsus ‘census’	κῆνσος	12.14
Caesar ‘Caesar’	καῖσαρ	12.14, 16, 17 [2×]
quadrāns ‘a coin’	κοδράντης	12.42
vae ‘woe’	οὐαί	13.17; 14.21
flagellō ‘to whip’	φραγελλόω	15.15
praetōrium ‘palace’	πραιτώριον	15.16
centuriō ‘centurion’	κεντύριων	15.39, 44, 45

The presence of so many Latin words in Mark’s Gospel has been taken by many commentators as fairly strong evidence for a Roman (i.e., Western) provenance for the writing. Other scholars, however, argue for a Syrian or Palestinian setting. Zeichmann examines the character of these Latinisms and concludes that they are poorly internalised borrowings. Some of the senses Mark uses for the Latin words are quite atypical (for example πυγμή as ‘hand’). The very presence of Latinisms most likely indicates a post-70 CE writing date if in Palestine (when the Roman occupation, the use of Latin, and the prevalence of Roman items, such as coins, increased dramatically). On the other hand, Mark’s rather rough and irregular use of Latin words may indicate only

a rudimentary acquaintance with the language, which is consistent with a Roman provenance.[32]

Another application of studies in loanwords and linguistic borrowing for biblical studies is the use of loanwords to obtain relative dates for the composition of a text. If the historical time period during which cultural and linguistic contact took place between two speech communities is known, in theory approximate dates of composition could be assigned to texts containing particular loanwords. For example, the presence of Greek loanwords in the book of Daniel has frequently been used as an argument in favour of a Hellenistic-era composition. A recent re-examination of the evidence by Benjamin Noonan (2018), however, argues that the phonological makeup of these Greek loanwords (mainly musical terms) indicates that they were borrowed from a non-Attic source (i.e., not the dialect which spread as a consequence of Alexander's Hellenisation). This debate illustrates that linguistic evidence can play a role in assessing the date of a text's composition, but that one should be cautious in relying on such evidence for a conclusive decision (see §3.0 below).[33]

---

[32] Zeichmann himself favours a post-war Syrian or Palestinian provenance, though this is on the assumption that "the author of Mark was [not] a Judaean denizen writing from Rome or a refugee of the Jewish War" (Zeichmann 2017, 47).

[33] A vivid illustration of such overconfidence is the famous assessment of S. R. Driver (1900, lxiii, emphasis original) with regard to the date of Daniel: "The verdict of the language of Daniel is thus clear… the Greek words *demand*, the Hebrew *supports*, and the Aramaic *permits*, a date *after the conquest of Palestine by Alexander the Great*."

2.2.3. The Comparative Method

The comparative method has already been mentioned, but in this section it is explained in more detail, especially as a process. This is one of the most important tools historical linguists make use of when conducting research into a family of languages. As its name suggests, the method involves a *comparison* of assembled lexical items between two or more genetically related languages. It is important to state at the outset that the method is necessarily iterative by nature, as several assumptions required by the method are themselves clarified in the course of implementing the selfsame method. In the initial phase, of course, it may not necessarily have been established that the languages in question are in fact related, as superficial resemblances are not enough to establish such a link (since they could be chance similarities). Hence, there is a necessary element of circularity in the method; an initial provisional hypothesis is usually sufficient to begin. The pervasive presence of a core set of vocabulary that appears to be shared between two languages is a valid starting point for the application of the comparative method.[34]

The first step involves the assembling of a list of lexical items which are plausibly cognate with one another. At this stage, it is critical to exclude any borrowed lexemes in any language, since they would inevitably skew the analysis (see the previous section). This, too, is not always easy to accomplish, as there is a

[34] Sometimes the similarities between two languages have been much obscured by many layers of sound change. For example, it was not immediately obvious to scholars that Old Irish was Indo-European until the meticulous work of Franz Bopp.

degree of circularity in ascertaining which lexemes are borrowed and which ones are inherited.[35] Table 3 below shows a sample correspondence set between Hebrew, Aramaic, and Ugaritic.[36] Although of course these three languages are now known to be genetically related to one another, it is helpful to see exactly how this claim can be established. As one examines the list of correspondences, certain observations emerge. (In the following, only the voiced coronal consonants /d/, /z/, and /ð/ will be examined, though the correspondence sets above also reveal other patterns.) Not only do most of the words look fairly similar to one another, but particular patterns are evident.

For example, sometimes all three witnesses agree in sharing the same consonant /d/: דָּם | דמא | *dm* 'blood'. This is also found in קָדוֹשׁ | קַדִּישׁ | *qdš* 'holy' and צֶמֶד | צמדא | *ṣmd* 'holy', as well as דֶּלֶת | דלתא | *dlt* 'door'.[37] In another set, there is also agreement among the three languages with regard to /z/: זַיִת | זיתא | *zt* 'olive'. However, sometimes the languages disagree: זבח | דבח | *dbḥ* 'sacrifice' and זָהָב | דְּהַב | *ḫrṣ* 'gold'. (Further examination reveals

[35] In some cases, heavy lexical borrowing can eliminate much core vocabulary, which presents a unique set of challenges to scholars. One such example is Armenian, whose vocabulary is much more strongly shaped by Middle Iranian languages than English is by Norman French. It took quite a bit of careful work in the late nineteenth century, for example, to establish that Armenian constitutes a unique branch of Indo-European and that it is not an Iranian language.

[36] This example concerns the reconstruction of Proto-Northwest-Semitic; see Figure 4 in §2.1.4 above.

[37] Note that vowels are included in Table 3 only where there is evidence, but are omitted for verbs.

that in the last set, Ugaritic *ḫrṣ* must be removed from consideration as it is clearly cognate with a poetic equivalent חָרוּץ in Hebrew; the cognate of זָהָב has apparently been lost in Ugaritic.)

Table 3: Example cognate words in Hebrew, Aramaic, and Ugaritic

	BHebrew	**BAramaic**	**Ugaritic**
a.	זָהָב ‘gold’	דְּהַב ‘gold’	*ḫrṣ* ‘gold’
b.	זַיִת ‘olive’	JArm. זיתא ‘olive’	*zt* ‘olive’
c.	זֶה dem. pron.	דִּי rel. pron.	*d/ḏ* rel. pron.
d.	דֶּלֶת ‘door’	JArm. דלתא ‘door’	*dlt* ‘door’
e.	שַׁעַר ‘gate’	תְּרַע ‘door’	*tġr* ‘gate’
f.	דָּם ‘blood’	JArm. דמא ‘blood’	*dm* ‘blood’
g.	זבח ‘sacrifice’	דבח ‘sacrifice’	*dbḥ* ‘sacrifice’
h.	קָדֹשׁ ‘holy’	קַדִּישׁ ‘holy’	*qdš* ‘holy’
i.	שָׁלֹשׁ ‘three’	תְּלָת ‘three’	*ṯlṯ* ‘three’
j.	תַּחַת ‘below’	תְּחוֹת ‘under’	*tḥt* ‘under’
k.	צוּר ‘rock’	טוּר ‘mountain’	*ġr* ‘mountain’
l.	צֶלֶם ‘image’	צֶלֶם ‘image’	*ṣlm* ‘image’
m.	אֶרֶץ ‘earth’	אַרַע/אֶרַק ‘earth’	*ʾarṣ* ‘earth’
n.	צַלְמָוֶת (?)	חֲשׁוֹךְ ‘darkness’	*ġlmt/ẓlmt* ‘darkness’
o.	צֶמֶד ‘yoke’	JArm. צמדא ‘yoke’	*ṣmd* ‘team’
p.	מחץ ‘strike’	מחא ‘strike’	*mḫṣ* ‘fight’

Once this is done, however, there remains a discrepancy between the /d/ of Aramaic and Ugaritic, on the one hand, and the /z/ of Hebrew, on the other. A final correspondence set, the relative/demonstrative pronoun זֶה | דִּי | *ḏ/d* reveals that Ugaritic apparently features two (dialectal) outcomes of the same segment. All these correspondences are then conveniently tabulated on a separate chart (Table 4 below). The next step is the analysis of the sets of segments in order to propose a single reconstructed phoneme per set. The full agreement between the three languages with regard to sets 1 (/z/) and 3 (/d/) allows us to hypothesise

that the immediate ancestor of the languages (Proto-Northwest-Semitic) featured a phonemic inventory with the same two segments: **z* and **d*.

Table 4: Segment correspondence in Hebrew, Aramaic, and Ugaritic

	Hebrew	**Aramaic**	**Ugaritic**
Set 1	/z/ז	/z/ז	z /z/
Set 2	/z/ז	/d/ד	ḏ /ð/ ~ d /d/
Set 3	/d/ד	/d/ד	d /d/

This is the most economical explanation, as it would be highly unlikely that each language independently innovated in the same direction, resulting in such correspondences.

Set 2, however, presents some challenges. Which segment ought we to reconstruct? In the majority of cases, Ugaritic agrees with Aramaic in sharing a /d/ segment (over against the /z/ of Hebrew). The interesting case of the relative pronoun *ḏ* /ð/, however, reveals that in at least some cases, the reflex of this correspondence set in Ugaritic was a voiced interdental fricative. A feature analysis (Table 5) shows that /ð/, although a fricative, is not a sibilant like /z/, and is therefore situated midway between /d/ and /z/. Cross-linguistic patterns further show that it is quite common for interdental fricatives to become sibilants (e.g., *ð > /z/; viz. Hebrew), or else to be 'strengthened' to a stop (e.g., *ð > /d/; viz. Aramaic).

Table 5: Feature analysis of /z/, / ð/, and /d/

	Coronal	**Voiced**	**Sibilant**	**Stop**
/z/	+	+	+	–
/ð/	+	+	–	–
/d/	+	+	–	+

As there is nothing apparent in the environment of these words (from the correspondence sets) to motivate a change in Aramaic or Ugaritic, we must instead hypothesise that an *unconditioned* merger took place in Hebrew between **ð* and **ẓ* (> /z/) that is absent in Aramaic. Correspondingly, a separate unconditioned merger must have taken place in Aramaic between **ð* and **d* (> /d/). The reconstructed segments in each of the correspondence sets are given below in Table 6 in the left column.

Table 6: Reconstructed segments of *z, *ð, and *d

	Hebrew	**Aramaic**	**Ugaritic**
*ẓ	ז/z/	ז/z/	*ẓ* /z/
*ð	ז/z/	ד/d/	*ḏ* /ð/ ~ *d* /d/
*d	ד/d/	ד/d/	*d* /d/

The results of this analysis indicate that Proto-Northwest-Semitic featured a series of three voiced coronal phonemes, which were reduced to two in Hebrew and Aramaic (though in different ways).

2.2.4. A Typology of Sound Change

The previous discussion illustrated the importance of understanding the phonology of the segments in correspondence sets assembled between cognate languages. Additionally, knowledge of common patterns of sound change in other languages provides valuable information for the comparative method. This section provides a brief overview of various types of phonological processes.

A pervasive phenomenon in sound change is *lenition*, or consonantal weakening. In general, consonantal segments frequently tend to change along certain common pathways, potentially leading to segment loss. In table 7 below, three different

places of articulation (bilabial, alveolar, and velar) are arranged in parallel to show common pathways towards ∅ (though other pathways are also possible). For example, a common process in languages is for geminated segments to be degeminated. Voiceless (non-geminated) stops will often become voiced in the environment of surrounding vowels (which are usually voiced in most languages).

Table 7: Pathways towards consonantal weakening

	Bilabial	**Alveolar**	**Velar**
Geminate	/pː/	/tː/	/kː/
Degemination	/p/	/t/	/k/
Voicing	/b/	/d/	/g/
Spirantisation	/β/	/ð/	/ɣ/
Debuccalisation	/h/	/h/	/h/
Segment loss	∅	∅	∅

The final stage prior to segment loss is called *debuccalisation*, in which the oral segment loses its original place of articulation, moving to the glottis (/h/).[38]

Consider an example of lenition from Biblical and Byzantine Greek. The spelling of Σιλουανοῦ (m. gen. sg.) 'Silvanus' at 2 Cor. 1.19, a loanword from Latin *Silvānus* /silwaːnus/, varies somewhat in the manuscript tradition (in which we also find lunate sigma):

(2) (a) ϲιλουανου
P46^{c} א A B C K L P Ψ 1739 *rell*

(b) ϲιλβανου
P46* D F G

[38] A far less common cross-linguistic process than lenition is its reverse, called *fortition*. The change of *ð > /d/ in Aramaic is one such example.

The spelling ϲιλβανου in (2b) indicates that for some later Byzantine scribes, the phoneme represented by {β} had already shifted from a voiced stop /b/ to a fricative /β/, which was perceived as the closest rendering of the Latin proper name (with /w/).[39]

Another common category of sound change is the loss of vocalic segments. Syllables which have lost articulatory prominence in words tend to lead to vowel reduction, and finally loss, which is called *apocope*. The loss of word-internal vowels is called *syncope*. A pervasive example of apocope in Hebrew is the loss of old Semitic case vowels (e.g., nominative **malk-u*, genitive **malk-i*, accusative **malk-a*, all losing their final vowel in מֶלֶךְ 'king').[40]

The juxtaposition of a consonantal segment next to a following sound articulated close to the palate of the mouth (whether a vowel or consonant) leads to a common sound process in many languages called *palatalisation*. For example, in Pre-Greek, the co-occurrence of certain consonants before the palatal glide /j/ (usually spelled as a *y* in English) led to a series of widespread sound changes, including changes in place of articulation (/l/ > /lʲ/, with a palatal co-articulation), changes in manner of articulation (stop > affricate), among others. The examples in (3)

[39] The IPA symbol for a voiced bilabial fricative is a Greek *beta*. Manuscript data taken from Royse (2007, 852). Note that the use of {ου} is itself an attempt to render Latin /w/ in Greek.

[40] An intermediate step was the form **malk*, which subsequently developed an anaptyctic vowel to break up the final consonant cluster *-lk*. This widespread phenomenon in Hebrew is termed *segolisation* (after the name of the final vowel seen in the so-called segolates).

below in Greek are illustrations from various present formations with the Indo-European **-jo-* suffix (contrasted with aorist):

(3) (a) **bal-jo-* > **balʲo-* > **baʎːo-* > βάλλω ‘throw’ (present); **(e-)bal-* > ἔβαλον (aorist)

(b) **klep-jo-* > **klepʲo-* > **kleptʃ͡ːo-* > κλέπτω ‘steal’ (present); **(e-)klep-sa-* > ἔκλεψα (aorist)

(c) **pʰulak-jo-* > **pʰulakʲo-* > **pʰulatʃ͡ːo-* > φυλάσσω ‘guard’ (present); **(e-)pʰulak-sa-* > ἐφύλαξα (aorist)

A very common process in sound change is the assimilation of a segment (or segments) in part or in whole to the features of an adjacent segment. For example, in Biblical Hebrew *hitpaʿʿel* stems, certain features of the root-initial segment spread backwards to the alveolar *t* of the stem prefix. This is referred to as *regressive* assimilation (the reverse is called *progressive* assimilation). For instance, the verbal root דבר formed a *hitpaʿʿel* with regressive voicing assimilation: e.g., **mit-dabir-* > **mid-dabir-* > מִדַּבֵּר (Num. 7.89).

Sometimes this assimilation in *hitpaʿʿel* stems was combined with yet another (often seemingly sporadic) process called metathesis, which involves the exchange of two or more segments, sometimes even when they are non-adjacent. This is a well-known phenomenon in *hitpaʿʿel* stems with roots which begin with a sibilant. For example, the verbal root צדק ‘to be just’ formed a *hitpaʿʿel* in which the emphatic feature of the root-initial sibilant spread to the adjacent dental of the prefix (assimilation). Additionally, the emphatic sibilant was swapped with the dental prefix: **nit-ṣadiq-* > **niṣṭadiq-* > נִצְטַדֶּק (Gen. 44.16; pausal form).

2.2.5. Analogy

In some ways, the two great 'engines' of language change are sound change and analogy (often competing with one another in terms of their effect on a given language). Analogy may be described as the process whereby speakers, in perceiving similarities (potentially of various kinds) between two linguistic structures, proceed to further advance the similarities in other aspects of the same linguistic structures.

For example, some scholars hold that the *weqaṭal* form in Biblical Hebrew is perhaps an analogical development drawing from the *wayyiqṭol*. When final short vowels were lost, the Semitic jussive-preterite form *yaqṭul* merged in most contexts with the imperfective *yaqṭulu*. In Biblical Hebrew, the old (short) *yiqṭol* lives on mainly in the *wayyiqṭol* form. The imperfective (long) *yiqṭol*, by contrast, is used to indicate various modal nuances, as well as imperfective aspect. Later speakers, unaware of this earlier development, could have perceived in the prefixed *waw* of the *wayyiqṭol* as having a 'converting' effect, transforming an imperfective *yiqṭol* into a perfective preterite (hence the *waw*-conversive terminology in some Biblical Hebrew pedagogy). The analogy comes into play when speakers took the suffix conjugation form, prefixed it with *waw*, and 'converted' its meaning from perfective to imperfective (i.e., the inverse of *wayyiqṭol*). Thus the *wayyiqṭol*, on this analysis, represents a direct historical development of the original jussive-preterite *yaqṭul*, while the *weqaṭal* represents an analogical development that was constructed on the model of *wayyiqṭol*.

When engaging in the task of linguistic reconstruction, it is critical to distinguish between genuine historical sound changes and various analogical effects (which sometimes even serve to undo particular sound changes in languages). One common form of analogy is *paradigm levelling*. Speakers over time tend to reduce inflectional distinctions in verbal or nominal paradigms, usually triggered by various kinds of analogy. For example, the 3mp prefix conjugation in Ugaritic features a *t-* element (thus in contrast to the *y-* of Biblical Hebrew). An example may be cited from a letter, KTU 2.63: *ỉlm tġrk tšlmk* 'may the gods (mp) guard you [and] keep you'. Notice that both verbs have a masculine subject (*ỉlm* 'gods'), yet feature a *t*-prefix. Likewise, an Amarna Canaanite letter (*EA* 252) contains an unambiguous example of a 3mp prefix conjugation form with a *t-* element: *ti-ma-ḫa-ṣú-ka* 'they strike you'.

This feature (from second millennium BCE Amarna Canaanite and Ugaritic) is used by Robyn Vern in part to argue against the view that there is such a thing as 'archaic' Biblical Hebrew Poetry. For her, one would need to demonstrate the presence of traces of this feature (among others) in Biblical Hebrew passages that are presumed to be archaic in order to sustain such a claim: "If there is sustainable evidence that there are second millennium remnants of the 3mp *t*-preformative, these poetic texts may be considered of second millennium typology with regard to this particular linguistic feature" (Vern 2011, 203).

Yet a consideration of two alternative models for the reconstructed paradigm of the prefix conjugation in Northwest Semitic indicates that the change of **ya-* → *ta-* is analogically justified

with Reconstruction B, while the reverse is manifestly not under either A or B.[41] It seems that paradigm levelling in Amarna Canaanite and Ugaritic was brought about by the pressure to unify all plural prefixes (apart from 1c) under *ta-*, a scenario favouring Reconstruction B. See Figure 5 below for an illustration of this mechanism.

Figure 5: Reconstructed paradigm of the prefix conjugation in Northwest Semitic

	Reconstruction A	**Reconstruction B**
3m	***ya**-qṭul-ū*	***ya**-qṭul-ū*
3f	**ya*-qṭul-na*	**ta*-qṭul-na*
2m	**ta*-qṭul-ū*	**ta*-qṭul-ū*
2f	**ta*-qṭul-na*	**ta*-qṭul-na*
1c	**na*-qṭul*	**na*-qṭul*

In considering two alternative reconstructions, historical linguists generally favour paradigmatic *heterogeneity* (all things being equal).

A final type of analogy to be discussed is *folk etymology*. A classic example in Biblical Hebrew is the traditional rendering of צַלְמָוֶת ‘shadow of death’, which occurs eighteen times in the Old Testament (most famously in Psalm 23.4). This understanding is based on a morphological segmentation צל ‘shadow’ + מות ‘death’. Such an interpretation is already seen in the ancient translations, e.g., the Septuagint (σκιὰ τοῦ θανάτου ‘shadow of death’) and the Targums (טולא דמותא ‘the shadow of death’). Its

[41] Some scholars make use of an arrow (→) or a double greater-than sign (≫) in order to distinguish analogical change from sound change (>). Note that Reconstruction A is based on the homogeneity (with **ya-*) exhibited by all major, early languages.

parallels in poetry, however, imply that the word is somehow associated with darkness, which is not a necessary interpretation of the phrase ‘shadow of death’. An apparent cognate in Ugaritic, *ġlmt* ‘darkness’ (variant spelling *ẓlmt*), suggests that צלמות should probably be segmented as a derivative in *-ūt* from an unattested stative root צלם* ‘to be dark’ (cf. Akkadian *ṣalāmu* ‘to be dark’). One can grasp by this example how folk etymology works. It occurs when speakers accidentally reinterpret a word or a phrase by morphological resegmentation. The original word צַלְמוּת* ‘darkness’ would have been reanalysed as צל ‘shadow’ + מות ‘death’, leading to sound change (-מוּת > -מָוֶת) that appears to be irregular.

2.2.6. Other theoretical concepts

Space precludes discussion of the many other kinds of language change (including syntactic change and semantic change), as well as additional methods historical linguists use to recover such changes (for example, internal reconstruction). As with many other fields of linguistics, historical linguistics is a broad field, with many sub-disciplines. Scholars continue to make progress in a number of increasingly specialised fields, while at the same time interdisciplinary approaches are growing in popularity, even drawing on fields outside of linguistics (for example, the recent application of Bayesian phylogenetics to language classification).

3.0. Historical Linguistics in Biblical Studies

The methods and tools of historical linguistics have been applied to biblical studies with ever growing intensity over the past few decades. The main area of application by far has been the Old Testament, which by its nature is arguably more amenable to such research. Despite this, improvement in work done in both the Old and New Testaments is a *desideratum*, as some of it can be uninformed at times about a number of generally accepted assumptions of historical linguistics.

Within Old Testament studies, historical linguistics has been applied most significantly (and increasingly controversially) to the question of the dating of its various texts. Fassberg (2016) offers an excellent overview of the history of scholarship in this area. He outlines three major time periods, the first beginning in the nineteenth century with the publication of Gesenius' *Geschichte der hebräischen Sprache und Schrift* (1815) and concluding with the publishing of the first fragments of Ben Sira in 1896 by Solomon Schechter. The second period continues with the incorporation of the insights about the language of Ben Sira, as well as that of the Damascus Document (published in 1910), until the discovery of the Dead Sea Scrolls (DSS) in 1947. The third period continues unto the present day, as the data about the DSS and especially the language contained in its non-biblical texts continues to be analysed. Especially noteworthy are the rigorous methodological principles for dating texts developed by Avi Hurvitz and continued by his students. A significant monograph that represents the cumulation of this research is Hurvitz's lexicon of Late Biblical Hebrew (2014).

Rezetko and Young (2019) argue persuasively that a fourth period may be discerned, which began in the 2000s. Within this current period, the heretofore accepted practice of dating biblical texts by means of linguistic features began to be seriously challenged by a number of scholars, most particularly in a series of works by Ian Young, Robert Rezetko, and Martin Ehrensvärd (Young, Rezetko, and Ehrensvärd 2008; Rezetko and Young 2014). Their work has generated intense debate and fresh research in this area. In the last decade, several major monographs on the question of diachrony in biblical Hebrew have emerged (e.g., Miller-Naudé and Zevit 2012; Moshavi and Notarius 2017). Although a definitive consensus has not yet emerged from all the debate, the discussions have fostered much more methodological caution in deriving actual dates (or even rough time periods) from the results of investigations into the language of particular texts of the Old Testament. To take but a single example, consider the presence of the apparent Aramaism יְתַנּוּ ‘they recount’ (Judg. 5.11) in the linguistically archaic Song of Deborah and Barak.[42] The presence of Aramaisms in Biblical Hebrew texts is typically taken to be evidence of a late date of composition. Yet this is probably an oversimplification, since there would have been early language contact between Aramaic-speaking peoples and those of Canaanite stock (such as Biblical Hebrew speakers). Such forms indicate that care must be taken in assigning dates to the language of biblical texts.

---

[42] The Old Greek rendering δώσουσιν ‘they give’ appears to be an alternative vocalisation reflecting an expected יִתְּנוּ.

An overall assessment of the debate can be briefly offered here. It seems that the confidence exhibited by an earlier generation of scholars has been tempered somewhat. (One thinks of Driver's [1900, lxiii] famously confident assertion about the dating of Daniel over a century ago, also mentioned in §2.2.2 above: "The verdict of the language of Daniel is thus clear… the Greek words *demand*, the Hebrew *supports*, and the Aramaic *permits*, a date *after the conquest of Palestine by Alexander the Great*" [emphasis original].) Even if absolute dates for language states could be established, there is no exact correspondence between language states and ancient biblical texts, which are composite documents received via centuries of textual transmission.

Furthermore, the contributions of scholars such as Gary Rendsburg have helpfully introduced the dialectal dimension to the diachronic debate: the language variation we see in the Old Testament text should not only be attributed to diachronic factors. To this might also be added other sociolinguistic phenomena such as register and even an author's idiolect.

At the same time, the extreme scepticism of Rezetko, Young, and Ehrensvärd seems unwarranted. As Aaron Hornkohl (2017, 75) notes: "We may… reasonably accept that BH has a history and that the general lines of this history may be traced in the Masoretic editions of biblical literature." The textual tradition for the Old Testament is reasonably secure (certainly for the purposes of detecting diachronic change). The Masoretic tradition has been demonstrated to have faithfully preserved minute linguistic nuances (even comparing favourably to many earlier texts

from the Dead Sea, which exhibit a degree of linguistic updating and harmonisation).

The scholarship of Avi Hurvitz and his students has produced fairly rigorous criteria for evaluating potential diachronic changes. For example, the criterion of linguistic opposition establishes evidence for real linguistic change in the complementary distribution of lexemes such as שֵׁשׁ 'linen' (Classical Biblical Hebrew) and בּוּץ 'linen' (Late Biblical Hebrew). Furthermore, although the extra-biblical evidence is meagre, it corroborates the picture we see within the biblical texts.

In short, while the vigorous debate in recent decades has produced positive improvements in methodology and it has helpfully tempered some of the more confident claims of an earlier generation of scholarship, it does not seem to have overturned the basic consensus: the Old Testament is a collection of texts produced over a period of several centuries, and these texts give evidence of diachronic linguistic development.

4.0. Prospects for Further Study and Application

No such equivalent application of the methods and tools of historical linguistics currently exists in New Testament studies. New Testament Greek, however, is a potential growth area and historical linguistics may well provide some tools for better understanding the Greek verb (especially from its Indo-European vantage point). A promising development in this regard is the recent publication of *The Greek Verb Revisited* by editors Steven Runge and Christopher Fresch (2016). Within this collection of essays, one

sees a new focus on the diachronic dimension as a necessary dialogue partner in the serious current scholarly disputes of the nature of the Koine Greek verb (see in particular the essays by Rutger Allen, Peter Gentry, and Amalia Moser).

Historical linguistics as a discipline stands as a welcome tool to scholars of all backgrounds engaged in the study of the biblical languages (as well as other ancient languages). Even the study of a single time period of a language is the study of a language in a state of perpetual (though gradual) diachronic flux and transition. This chapter has hopefully demonstrated that the tools and methods that historical linguists employ can be profitably used by biblical scholars as a means of advancing the state of our exegetical and historical knowledge of the biblical texts.

5.0. Further Reading

5.1. Handbooks, Companions, Etymological Dictionaries

1. Greek morphonology: Rix (1992)
2. Greek etymology: Chantraine (2009)
3. Greek verb: Runge and Fresch (2016)
4. Hebrew morphonology: Blau (2010)
5. Hebrew and Aramaic: Noonan (2020)
6. Non-Semitic Loanwords: Noonan (2019)
7. Hebrew vowels: Suchard (2019)
8. Semitic etymology: Kogan (2015)

5.2. General Introductions

1. Historical linguistics: Campbell (2021)
2. Indo-European: Fortson (2010)
3. Greek: Horrocks (2014)
4. Semitic: Rubin (2010)
5. Hebrew: Sáenz-Badillos (1993)
6. Aramaic: Gzella (2021)

5.3. Foundational Texts

1. Indo-European: Klein, Joseph, and Fritz (2017)
2. Wackernagel's 1918–19 lectures: Langslow (2009)
3. Historical linguistics: Hock (1991)
4. Indo-European: Kuryłowcz and Mayrhofer (1968)
5. Semitic: Huehnergard and Pat-El (2019)

References

Blau, Joshua. 2010. *Phonology and Morphology of Biblical Hebrew*. Linguistic Studies in Ancient West Semitic 2. Winona Lake, IN: Eisenbrauns.

Bybee, Joan. 2007. 'Diachronic Linguistics'. In *The Oxford Handbook of Cognitive Linguistics*, edited by Dirk Geeraerts and Hubert Cuyckens, 945–87. Oxford: Oxford University Press.

———. 2010. *Language, Usage, and Cognition.* Cambridge: Cambridge University Press.

Campbell, Lyle. 2021. *Historical Linguistics: An Introduction*. 4th ed. Cambridge, MA: The MIT Press.

Chantraine, Pierre. 2009. *Dictionaire étymologique de la langue grecque: Histoire des mots.* Librairie Klincksieck: Serie linguistique. Paris: Klincksieck.

de Saussure, Ferdinand. 1916. *Cours de linguistique générale*. Paris: Payot.

———. 1879. *Memoire sur le système primitif des voyelles dans les langues indo-européennes.* Leipzig: Vieweg.

Driver, S. R. 1900. *The Book of Daniel: With Introduction and Notes*. Cambridge: Cambridge University Press.

Fassberg, Steven E. 2016. 'What is Late Biblical Hebrew?'. *Zeitschrift für die alttestamentliche Wissenschaft* 128 (1): 1–15.

Fortson, Benjamin W. 2010. *Indo-European Language and Culture: An Introduction.* Malden, MA: Wiley-Blackwell.

Gesenius, Friedrich H. W. 1815. *Geschichte der hebräischen Sprache und Schrift*. Leipzig: Vogel.

Gzella, Holger. 2021. *Aramaic: A History of the First World Language.* Translated by Benjamin Suchard. Grand Rapids: Eerdmans.

Hartmann, Stefan. 2021. 'Diachronic Cognitive Linguistics: Past, Present, and Future'. *Yearbook of the German Cognitive Linguistics Association* 9 (1): 1–34.

Hock, Hans Henrich. 1991. *Principles of Historical Linguistics*. Trends in Linguistics 34. Berlin: Mouton de Gruyter.

Hornkohl, Aaron. 2017. 'All Is Not Lost: Linguistic Periodization in the Face of Textual and Literary Pluriformity'. In *Advances in Biblical Hebrew Linguistics: Data, Methods, and Analyses*, edited by Adina Moshavi and Tania Notarius, 53–80. Linguistic Studies in Ancient West Semitic 12. Winona Lake, IN: Eisenbrauns.

Horrocks, Geoffrey C. 2014. *Greek: A History of the Language and its Speakers*. Hoboken, NJ: Wiley-Blackwell.

Huehnergard, John, and Aaron Rubin. 2011. 'Phyla and Waves: Models of Classification'. In *The Semitic Languages: An International Handbook*, edited by Stefan Weniger, 259–78. Berlin: De Gruyter.

Huehnergard, John, and Na'ama Pat-El. 2019. *The Semitic Languages.* London: Routledge.

Hurvitz, Avi. 2014. *A Concise Lexicon of Late Biblical Hebrew.* Supplements to Vetus Testamentum 160. Leiden: Brill.

Jasanoff, Jay H. 2003. *Hittite and the Indo-European Verb*. Oxford: Oxford University Press.

Khan, Geoffrey. 2020. *The Tiberian Pronunciation Tradition of Biblical Hebrew.* Vol. 1. Cambridge Semitic Languages and Cultures 1. Cambridge: Open Book Publishers.

Klein, Jared, Brian Joseph, and Matthias Fritz (eds). 2017. *Handbook of Comparative and Historical Indo-European Linguistics*. Handbücher zur Sprach- und Kommunikationswissenschaft 41. Berlin: De Gruyter Mouton.

Kloekhorst, Alwin. 2019. *Kanišite Hittite*. Handbook of Oriental Studies, Section 1: The Near and Middle East 132. Leiden: Brill.

Kogan, Leonid. 2015. *Genealogical Classification of Semitic: The Lexical Isoglosses.* Berlin: De Gruyter.

Kuryłowcz, Jerzy, and Manfred Mayrhofer. 1968. *Indogermanische Grammatik.* Heidelberg: Carl Winter Universitätsverlag.

Langslow, David. 2009. *Jacob Wackernagel, Lectures on Syntax: With Special Reference to Greek, Latin, and Germanic*. Oxford: Oxford University Press.

Metcalf, George J. 1974. 'The Indo-European Hypothesis in the Sixteenth and Seventeenth Centuries'. In *Studies in the History of Linguistics: Traditions and Paradigms*, edited by Dell H. Hymes, 233–57. Bloomington, IN: Indiana University Press.

Miller-Naudé, Cynthia, and Ziony Zevit (eds). 2012. *Diachrony in Biblical Hebrew*. Linguistic Studies in Ancient West Semitic 8. Winona Lake, IN: Eisenbrauns.

Moshavi, Adina, and Tania Notarius (eds). 2017. *Advances in Biblical Hebrew Linguistics: Data, Methods, and Analyses.* Linguistic Studies in Ancient West Semitic 12. Winona Lake, IN: Eisenbrauns.

Noonan, Benjamin. 2018. 'Daniel's Greek Loanwords in Dialectal Perspective'. *Bulletin for Biblical Research* 28 (4): 575–603.

———. 2020. *Advances in the Study of Biblical Hebrew and Aramaic: New Insights for Reading the Old Testament.* Grand Rapids: Zondervan Academic.

Osthoff, Hermann, and Karl Brugmann. 1878. *Morphologische Untersuchungen auf dem Gebeiete der indogermanischen Sprachen.* Vol. 1. Leipzig: Verlag von S. Hirzel.

Pat-El, Na'ama, and Aren Wilson-Wright. 2018. 'Features of Aramaeo-Canaanite'. *Journal of the American Oriental Society* 138 (4): 781–806.

Rezetko, Robert, and Ian Young. 2014. *Historical Linguistics and Biblical Hebrew: Steps Toward an Integrated Approach*. Ancient Near East Monographs 9. Atlanta: SBL Press.

———. 2019. 'Currents in the Historical Linguistics and Dating of the Hebrew Bible: Report on the State of Research as Reflected in Recent Major Publications'. *HIPHIL Novum* 5 (1): 3–95.

Rix, Helmut. 1992. *Historische Grammatik des Griechischen: Laut- und Formenlehre*. Darmstadt: Wissenschaftliche Buchgesellschaft.

Royse, James R. 2007. *Scribal Habits in Early Greek New Testament Papyri*. New Testament Tools, Studies, and Documents 36. Atlanta: SBL Press.

Rubin, Aaron D. 2010. *A Brief Introduction to the Semitic Languages*. Gorgias Handbooks 19. Piscataway, NJ: Gorgias Press.

Runge, Steven E., and Christopher Fresch. 2016. *The Greek Verb Revisited: A Fresh Approach for Biblical Exegesis*. Bellingham, WA: Lexham Press.

Sáenz-Badillos, Angel. 1993. *A History of the Hebrew Language*. Translated by John Elwolde. Cambridge: Cambridge University Press.

Suchard, Benjamin. 2018. 'The Vocalic Phonemes of Tiberian Hebrew'. *Hebrew Studies* 59: 193–207.

Ventris, Michael, and John Chadwick. 1973. *Documents in Mycenaean Greek*. Cambridge: Cambridge University Press.

Vern, Robyn. 2011. *Dating Archaic Biblical Hebrew Poetry: A Critique of the Linguistic Arguments*. Perspectives on Hebrew

Scriptures and its Contexts 10. Piscataway, NJ: Georgias Press.

Watson, Wildred, and Nicolas Wyatt. 1999. *Handbook of Ugaritic Studies*. Handbook of Oriental Studies, Section 1: The Near and Middle East 39. Leiden: Brill.

Williams, Peter J. 2018. 'Semitic Long /i/ Vowels in the Greek of Codex Vaticanus of the New Testament'. In *Studies in Semitic Linguistics and Manuscripts: A Liber Discipulorum in Honour of Professor Geoffrey Khan*, edited by Nadia Vidro, Ronny Vollandt, Esther-Miriam Wagner, and Judith Olszowy-Schlanger, 15–26. Studia Semitica Upsaliensa 201. Uppsala: Acta Universitatis Upsaliensis.

Young, Ian, Robert Rezetko, and Martin Ehrensvärd. 2008. *Linguistic Dating of Biblical Texts*. 2 vols. London: Equinox.

Zeichmann, Christopher B. 2017. 'Loanwords or Code-Switching? Latin Transliteration and the Setting of Mark's Composition'. *Journal of the Jesus Movement in its Jewish Setting* 4: 42–64.