

The Semantics of Word Division in Northwest Semitic Writing Systems

Ugaritic, Phoenician, Hebrew, Moabite and Greek

Robert S. D. Crellin



The Semantics of Word Division in Northwest Semitic Writing Systems

Ugaritic, Phoenician, Hebrew, Moabite and Greek

Robert S. D. Crellin

Published in the United Kingdom in 2022 by
OXBOW BOOKS
The Old Music Hall, 106–108 Cowley Road, Oxford, OX4 1JE

and in the United States by
OXBOW BOOKS
1950 Lawrence Road, Havertown, PA 19083

© Oxbow Books and Robert S. D. Crellin 2022

Hardback edition: ISBN 978-1-78925-677-2
Digital Edition: ISBN 978-1-78925-678-9

A CIP record for this book is available from the British Library

Library of Congress Control Number: 2021949706

An open-access on-line version of this book is available at: http://books.casematepublishing.com/The_Semantics_of_Word_Division.pdf. The online work is licensed under the Creative Commons Attribution 3.0 Unported Licence. To view a copy of this license, visit <http://creativecommons.org/licenses/by/3.0/> or send a letter to Creative Commons, 444 Castro Street, Suite 900, Mountain View, California, 94041, USA. This licence allows for copying any part of the online work for personal and commercial use, providing author attribution is clearly stated.

Some rights reserved. No part of the print edition of the book may be reproduced or transmitted in any form or by any means, electronic or mechanical including photocopying, recording or by any information storage and retrieval system, without permission from the publisher in writing.

Materials provided by third parties remain the copyright of their owners.

Printed in the United Kingdom by Short Run Press

Typeset in India by Lapiz Digital Services, Chennai.

For a complete list of Oxbow titles, please contact:

UNITED KINGDOM
Oxbow Books
Telephone (01865) 241249
Email: oxbow@oxbowbooks.com
www.oxbowbooks.com

UNITED STATES OF AMERICA
Oxbow Books
Telephone (610) 853-9131, Fax (610) 853-9146
Email: queries@casemateacademic.com
www.casemateacademic.com/oxbow

Oxbow Books is part of the Casemate Group

Front cover: The Cloisters Collection, 2018 'Hebrew Bible'



Contents

Acknowledgements.....	vii
Abbreviations.....	viii
1. Introduction.....	1
1.1. What is a word?	1
1.2. Why Northwest Semitic and Greek?	4
1.3. Wordhood in writing systems research.....	7
1.4. Linguistic levels of wordhood	11
1.5. Word division at the syntax-phonology interface.....	28
1.6. Previous scholarship	34
1.7. Method	39
1.8. Outline	51
Part I Phoenician	55
2. Introduction.....	57
2.1. Overview	57
2.2. Literature review	57
2.3. Corpus.....	60
2.4. Linguistic and sociocultural identity of the inscriptions	62
2.5. Proto-alphabetic	63
2.6. Shared characteristics of word division	65
2.7. Divergence in word division practice	65
3. Prosodic words	67
3.1. Introduction	67
3.2. Distribution of word division	67
3.3. Graphematic weight of function words	68
3.4. Morphosyntax of unverbated syntagms	71
3.5. Sandhi assimilation	79
3.6. Comparison of composition and distribution with prosodic words in Tiberian Hebrew.....	81
3.7. Conclusion.....	87
4. Prosodic phrase division.....	88
4.1. Introduction	88
4.2. Syntax of unverbated syntagms	89

4.3. Comparison with prosodic phrases in Tiberian Hebrew 95

4.4. Syntactic vs. prosodic phrase level analysis 102

4.5. Verse form 103

4.6. Conclusion..... 104

Part II Ugaritic alphabetic cuneiform 105

5. Introduction..... 107

5.1. Overview 107

5.2. Literature review 109

5.3. Basic patterns of word division and univerbation 115

5.4. Exceptions to the basic patterns of word division 117

5.5. Line division 118

5.6. Contexts of use..... 118

5.7. Textual issues 120

5.8. Inconsistent nature of univerbation 121

5.9. Hypothesis: Graphematic words represent actual prosodic words 122

6. The Ugaritic ‘Majority’ orthography 124

6.1. Introduction 124

6.2. Syntagms particularly associated with univerbation 124

6.3. Univerbation with nouns 124

6.4. Univerbation with verbs 126

6.5. Univerbation with suffix pronouns..... 127

6.6. Univerbation at clause and phrase boundaries..... 127

6.7. Summary..... 128

7. Quantitative comparison of Ugaritic and Tiberian Hebrew..... 130

7.1. Introduction 130

7.2. Corpus..... 130

7.3. Frequency of occurrence..... 131

7.4. Length of phrase..... 131

7.5. Quantifying the morphosyntactic collocation of linking features 134

7.6. Measuring Association Score B for Ugaritic and Tiberian Hebrew 141

7.7. Visualising morphosyntactic collocation of linking features
with MDS..... 145

7.8. Conclusion..... 150

8. Semantics of word division in the Ugaritic ‘Majority’ orthography:
prosodic word or prosodic phrase 151

8.1. Introduction 151

8.2. Graphematic wordhood in the Ugaritic ‘Majority’ orthography 152

8.3.	Consistency of the representation of ACTUAL PROSODIC WORDHOOD in Ugaritic	155
8.4.	Univerbation at clause boundaries.....	156
8.5.	Adoption of the ‘Majority’ orthography outside of literary contexts.....	159
9.	Separation of prefix clitics.....	162
9.1.	Introduction	162
9.2.	Literary texts.....	163
9.3.	Non-literary texts adopting the ‘Majority’ orthography.....	165
9.4.	Non-literary texts adopting the ‘Minority’ orthography.....	167
9.5.	Conclusion.....	190
Part III Hebrew and Moabite		193
10.	Word division in the consonantal text of the Hebrew Bible	195
10.1.	Introduction	195
10.2.	Morphosyntactic status of graphematic affixes in Tiberian Hebrew	197
10.3.	Morphosyntactic status of graphematic affixes.....	204
10.4.	Graphematic status of graphematic affixes	204
10.5.	Conclusion.....	206
11.	Word division in the consonantal Masoretic Text: Minimal prosodic words.....	207
11.1.	Introduction	207
11.2.	Combining prosody and morphosyntax (Dresher 1994; Dresher 2009).....	225
11.3.	Accounting for graphematic wordhood prosodically	228
11.4.	מה mah ‘What?’	228
11.5.	לֹדֹ'	229
11.6.	Minimal domains for stress assignment and sandhi.....	230
11.7.	Conclusion.....	232
12.	Minimal prosodic words in epigraphic Hebrew and Moabite.....	233
12.1.	Introduction	233
12.2.	Siloam Tunnel inscription.....	235
12.3.	Mesha ^c stelae (KAI 181 and KAI 30).....	236
12.4.	Accounting for word division in the Mesha ^c and Siloam inscriptions.....	240
12.5.	Conclusion.....	241
12.6.	Conclusion to Part III	241

Part IV Epigraphic Greek	243
13. Introduction.....	245
13.1. Overview	245
13.2. Corpus.....	246
13.3. Prosodic wordhood in Ancient Greek	247
13.4. Metre and natural language	249
13.5. Problems with identifying graphematic words with prosodic words	249
13.6. Conclusion.....	261
14. The pitch accent and prosodic words	262
14.1. Introduction	262
14.2. Prosody of postpositives and enclitics.....	262
14.3. Prosody of prepositives and ‘proclitics’.....	264
14.4. Conclusion.....	266
15. Domains of pitch accent and rhythm	268
15.1. Introduction	268
15.2. Challenging the inherited tradition of accentuation	270
15.3. Pitch accentuation and rhythmic prominence have different domains	271
15.4. Rhythmic words are canonically trimoraic or greater	277
15.5. Graphematic words correspond to rhythmic words	279
15.6. Conclusion.....	281
16. Graphematic words with multiple lexicals.....	282
16.1. Introduction	282
16.2. Inconsistency of levels of graphematic representation.....	286
16.3. Prosodic subordination of one lexical to another	287
16.4. Punctuating canonical rhythmic words	291
16.5. Conclusion.....	292
17. Epilogue: The context of word division	294
17.1. Overview	294
17.2. Orality and literacy	295
17.3. Prosodic word level punctuation is a function of the oral performance of texts.....	297
Bibliography.....	301

Acknowledgements

The present study was completed as part of ongoing research under the CREWS project (Contexts of and Relations between Early Writing Systems), funded by the European Research Council under the Horizon 2020 research and innovation programme (grant agreement No 677758), led by Pippa Steele. I would like to record my deep gratitude for the opportunity to work on such a stimulating and interesting topic for the last four years.

The monograph was written in LaTeX-like markup compiled into Word format by VBA written by the author. Tree diagrams were prepared with the *tikz-qtrees* (<https://ctan.org/pkg/tikz-qtrees?lang=en>) and *standalone* (<https://ctan.org/pkg/standalone?lang=en>) LaTeX packages, along with ImageMagick Convert for creating PNG files.

The monograph would not have been possible without the enormous support of a large number of friends, family and colleagues. For their very helpful comments and suggestions, as well as their support in other ways, I would like to express my profound thanks to Ivri Bunis, Jessica Hawxwell, Aaron Hornkohl, Geoffrey Khan, Richard Sproat and Pippa Steele, each of whom read all or part of the monograph in its earlier versions, and offered very helpful corrections or suggestions for improvement. In addition I am very grateful to Chis Golston, Martin Evertz, John Ellison, Natalia Elvira Astoreca, Alice Faber, Ben Kantor and Aaron Koller for sending me copies of their PhD dissertations or papers; to James Diggle, David Goldstein, Torsten Meissner and Nick Zair for stimulating email conversations; to Joaquín Sanmartín and Wilfred Watson for their helpful advice; and to Rupert Thompson for assistance in relation to matters Mycenaean. Needless to say, any remaining errors are my own responsibility.

The book would also not have been possible without the love and support of family: David, Hilary, Steve, Claire, Sara, Rachel, Alex, Isabella, Sophie, Finley, Sarah, Tim, Hannah, Ursula, Philip and Esther. Particular thanks are due to Steve and Claire Jones for going above and beyond the call of duty with help with childcare.

I would like to acknowledge the friendship and support of my colleagues on the CREWS project, and in the Faculties of Classics and Asian and Middle Eastern Studies in Cambridge: Estara Arrant, Natalia Elvira Astoreca, Philip Boyes, James Clackson, Anna Judson, Ben Kantor, Johan Lundberg, Pippa Steele, and Peter Williams.

Finally, I would like to express my deep gratitude to my wife Hannah, for her love, friendship and steadfast support throughout, as well as to our dear son Barnabas, who has been such a great source of happiness and joy.

SDG

Robert Crellin
Lichfield, August 2021

Abbreviations

1Chr	1 Chronicles	Josh	Joshua
2Chr	2 Chronicles	Judg	Judges
1Kgs	1 Kings	L	Low tone (ch. 15)
2Kgs	2 Kings	Lev	Leviticus
1Sam	1 Samuel	Mal	Malachi
2Sam	2 Samuel	MDS	MultiDimensional Scaling
Aesch.	Aeschylus	Nah	Nahum
Adj	Adjective	Neh	Nehemiah
Adv	Adverb	NH	Non-High tone (ch. 15)
BH	Biblical Hebrew	Num	Numbers
BHS	Biblia Hebraica Stuttgartensia	Obad	Obadiah
Conj	Conjunction	ORL	Orthographically Relevant Level
Deut	Deuteronomy	PMH	Prosodic Minimality Hypothesis
DN	Divine Name	PN	Personal Name
DP	Determiner Phrase	POS	Part of Speech
Exod	Exodus	PREF	Prefix conjugation
Ezek	Ezekiel	Prep	Preposition
Gen	Genesis	PP	Prepositional Phrase
GN	Gentilic	Pron	Pronoun
Eur.	Euripides	Prov	Proverbs
Exod	Exodus	Psa	Psalms
H	High tone (ch. 15)	Ptcl	Particle
Hag	Haggai	Qoh	Qohelet (= Ecclesiastes)
Hipp.	Hippolytus	SuffPron	Suffix Pronoun
Isa	Isaiah	TN	Toponym
I.T.	Iphigenia in Tauris	VP	Verb Phrase
Jer	Jeremiah	Zech	Zechariah

Chapter 1

Introduction

1.1. What is a word?

It might at first seem obvious what words are: sequences of letters separated by spaces or punctuation. So in the sentence you have just read, ‘It’, ‘might’ and ‘seem’ would all be ‘words’.

Matters become more complicated when we encounter languages and writing systems that appear not to follow our instincts on what constitutes a ‘word’. A case in point, and a writing system that will feature heavily in the present study, is Hebrew. Here word division follows rather different principles. To illustrate, consider the opening verse of Genesis in the Hebrew Bible (transcription and glossing given immediately below):¹

(1) ← בְּרֵאשִׁית בְּרָא אֱלֹהִים אֶת הַשָּׁמַיִם וְאֶת הָאָרֶץ :

Reading right-to-left we can see that there are sequences of letters interspersed by spaces, and terminating in a mark that looks like punctuation, a colon. So at a first glance, word division in Hebrew appears to be similar to word division in, for example, modern English. But when these sequences are analysed to see what they contain, it is immediately apparent that the principles of word division are different. The most notable difference is that one-letter words are written together with the next word:

(2) Gen 1:1

<i>b=ršyt</i>	<i>brʹ</i>	<i>ʾlhym</i>	<i>ʾt</i>	<i>h-šmym</i>
in=beginning	created	God	OBJ	the-heavens
<i>w='t</i>	<i>h-'rṣ</i>			
and=OBJ	the-earth			

¹ The text of the Hebrew Bible throughout is that of the Westminster Leningrad Codex (<https://tanach.us/>).

‘In the beginning God created the heaven and the earth’ (KJV)²

Units joined by the = sign in the transcription correspond to a single word in the Hebrew text. From this we can see that several small words are written together with the next word:

- Preposition בִּ- *b-* ‘in’: בְּרֵאשִׁית *b=ršyt* ‘in beginning’
- Article הַ- *ha-* ‘the’: הַשָּׁמַיִם *h-šmym* ‘the heavens’
- Conjunction וְ- *w-* ‘and’: וְאֵת *w=t* ‘and OBJ’
- Article הַ- *ha-* ‘the’: הָאָרֶץ *h-rš* ‘the earth’

Adopting the word division orthography of Hebrew for English would give us:

(3) Inthebeginning God created theheavens and theearth.

Genesis 1:1 is by no means unique. In fact, this approach to word division, where small words are written together with the next word is a feature of the writing of many Semitic languages, including Ugaritic, Phoenician and Moabite in the ancient world, and Modern Hebrew and Arabic today. We can see the same thing, for example, in the following excerpt from an early 1st millennium BCE Phoenician inscription from Byblos:

(4) KAI⁵ 1:2

לגלזל. + וּמלך. כ. וּ + ו. ומלמל. וּמלמל. וּמלמל. וּמלמל. וּמלמל. ←

<i>w=l</i> _{<w>}	<i>mlk</i> _{<w>}	<i>b=mlkm</i> _{<w>}	<i>w=skn</i> _{<w>}
and=if	king	among=kings	and=governor
<i>b=sknm</i> _{<w>}	<i>w=tm</i> ² _{<w>}	<i>mḥnt</i> _{<w>}	<i>ly</i> _{<w>}
among=governors	and=commander	camp	rise_up

*gbl*_{<w>}

TN

‘And if a king among kings, or governor among governors, or camp commander should rise up against Byblos’ (trans. with ref. to Donner & Röllig 1968, 2)

As the transcription shows, the conjunction וְ *w-* ‘and’ and the preposition וּ *b-* ‘among’ are written together with the words that follow them.

² Bible translations, if they are not the author’s own, are given from one of three sources: the Authorized Version (KJV), the English Revised Version (ERV) and the Revised Standard Version (RSV). These are listed in the bibliography under their abbreviations. Non-Biblical translations, unless the author’s own, are cited in the normal way.

It is perhaps not so widely known, however, that a subset of Ancient Greek inscriptions from the first half of the 1st millennium BCE adopt a very similar approach to word division. The following is an excerpt of an inscription from the Greek city of Argos, in the Peloponnese, from the 6th century BCE:³

(5) SEG 11:314 1–3 (Argos, 575–500 BCE; text per Probert & Dickey 2015)

→ ΕΠΙΤΟΝΔΕΟΝΕΝ : ΔΑΜΙΙΟΡΓΟΝΤΟΝ : ΤΑΕ[Ν] | ΣΑΘΑΝΑΙΑΝ :
ΕΠΟΙΦΕΣΘΕ : ΤΑΔΕΝ : ΤΑΠΟΙΦΕ | ΜΑΤΑ : ΚΑΙΤΑΧΡΕΜΑΤΑΤΕ

In this inscription words are separated by tripuncts ⟨:⟩ rather than by spaces, which was the method of word division in the Hebrew example given earlier. But in terms of what is separated, there is a remarkable degree of similarity to what we find in Hebrew:

(6) SEG 11:314 1–3 (Argos, 575–500 BCE; text per Probert & Dickey 2015)

epì=tōndeōnēn *damiiorgōntōn* *tà=ens=athanaían* *epoiwésthē*
on=the_following serve_as_damiorgoí the=to=Athena were_made

tadén *tà=poiwémata* *kai=tà=khrémata=te*
these the=works and=the=treasures=both

‘When the following were *damiorgoí*, the following things concerned with Athena were made: the works and the treasures and ...’(trans. Probert & Dickey 2015: 115)

Once again, the = sign is used to denote items that are written together in the original text. We find the same kinds of words written together with the following words as we did in Genesis 1 verse 1:

- Preposition ΕΠΙ *epì* ‘on’: ΕΠΙΤΟΝΔΕΟΝΕΝ *epì=tōndeōnēn* ‘while the following’;
- Article τὰ τὰ ‘the’: ΤΑΠΟΙΦΕΜΑΤΑ *tà=poiwémata* ‘the works’;
- Conjunction ΚΑΙ *kai* ‘and’: ΚΑΙΤΑΧΡΕΜΑΤΑΤΕ *kai=tà=khrémata=te* ‘and the treasures and’.

Unlike Hebrew and other West Semitic languages, this writing convention has not been carried through into modern Greek texts, either in Modern or Ancient Greek. Thus in the recent publication of this inscription by Probert & Dickey (2015), the text is written as follows, with spaces between morphosyntactic words (see also fn 3; Probert & Dickey indicate line division with new lines):

³ The original is written in so-called ‘boustrophedon’, whereby lines alternate in direction between right-to-left and left-to-write. However, since for these purposes we are interested in word division rather than direction of writing, for the sake of this exposition the text is presented as left-to-right only, with | indicating a line break. Probert & Dickey (2015) indicate line division with line breaks.

(7) → Ἐπὶ τῶνδεὼν ἔν : δαμιουργόντων : τὰ ἐ[ν]||ς ἀθαγαίαν : ἐπορφέσθῃ : ταδέν : τὰ ποιφέ||ματα : καὶ τὰ χρέματά τε

Modern editions therefore disguise a fundamental similarity between two sets of writing systems, those of the ancient Northwest Semitic languages Phoenician, Ugaritic, Hebrew and Moabite, on the one hand, and Greek on the other. The primary goal of this study is to establish the principles that govern word division in these writing systems: why did the writers of these texts separate words in the way that they did? Was it conventional only, or can a rationale be discerned? This question occupies the main part of the monograph, Parts I–IV, with one part devoted to each of Phoenician, Ugaritic, Hebrew/Moabite and Greek. I conclude that – with one exception in a subset of Ugaritic texts – that words are divided according to the principles under which units are divided in the spoken language, rather than those that would be implied by a grammatical analysis. In the Epilogue I go on to address what this fact can tell us about the world in which the writers of the inscriptions operated, and in particular, what it might tell us about the relationship between the written and the spoken word in their societies.

The introduction proceeds as follows. First in §1.2 I provide the rationale for the languages and writing systems considered in this study, that is, why I treat Northwest Semitic and Ancient Greek together. Sections 1.3, 1.4 and 1.5 consider the linguistics of word division. After that I outline how the question of word division in Northwest Semitic and Greek has been addressed in previous studies (§1.6). Finally in §1.7 I outline the method used in this study to assess the nature of word division.

1.2. Why Northwest Semitic and Greek?

The present study addresses word division in alphabetic Northwest Semitic and Greek inscriptions up to the mid-1st millennium BCE. However, these languages and their epigraphic practices are rarely studied together, except in the context of Biblical Studies. This is particularly true in the study of the target of word division, where from §1.6 it will be seen that the study of this question has followed quite different paths in the two academic disciplines. Consequently a word of explanation is needed as to why the two are studied together here.

1.2.1. Common origin of the Northwest Semitic and Greek alphabets

Northwest Semitic languages and Greek are generally studied separately from one another because they represent two different language families, viz. Semitic and Indo-European. However, the alphabets used to write these two language sub-branches have a common ancestor: as is well known, the Greek alphabet represents a development of an alphabet used to write a West Semitic language (Naveh 1973a, 1; Waal 2018, 84). The view among Greek scholars has tended to be that this Semitic language was Phoenician, and that the alphabet was adopted by Greek-speakers in

the late 9th or early 8th century BCE (Waal 2020, 110, 121; 2018, 88). Naveh (1973a) challenged the prevailing view of the origin and date of transmission of the Greek alphabet, proposing a date of transmission in the 11th century BCE.⁴ The main arguments for an early transmission date of the Greek alphabet may be summarised as follows (Naveh 1973a; Waal 2018; 2020). First, in the earliest Greek inscriptions the direction of writing is not fixed, varying between left-to-right, right-to-left, and ‘boustrophedon’, *i.e.* where the direction of writing alternates between left-to-right and right-to-left (Waal 2018, 87). This is a property shared with Northwest Semitic inscriptions from the 2nd millennium BCE (Waal 2018, 85). By contrast, in the extant Phoenician inscriptions from the late 2nd/early 1st millennium BCE, the writing direction is fixed, right-to-left (Waal 2018, 85, 93–94). It seems inherently more likely that the Greeks inherited a writing tradition without a fixed direction, than that they inherited a fixed right-to-left tradition and subsequently transformed it back to be more like its more ancient forebear (Naveh 1973a, 2–3; Waal 2018, 93).

Second, in their earliest attestations the Greek alphabets are geographically widespread and show considerable diversity in letter shapes (Waal 2018, 96–100). Despite this, they all share the innovation of the writing of vowel signs, implying a single origin. Given that the first attestations of the Greek alphabet are from the 8th century BCE (Waal 2018, 86), an 8th-century BCE adoption of the alphabet by Greek speakers would entail a high degree of diversification and geographical spread over a very short space of time, which seems implausibly fast (Waal 2020, 110).

Finally, there are striking similarities in the forms of punctuation used in Greek and in Northwest Semitic material from the late 2nd millennium BCE (Waal 2018, 94–96). In Northwest Semitic, the earliest word divider is a short vertical stroke (Naveh 1973b, 206–207). In Ugaritic alphabetic cuneiform this surfaces as the small vertical wedge (Ellison 2002). However, the tripunct ⟨:⟩ is found in the Lachish Ewer from ca. 13th century BCE (Naveh 1973a, 7 n. 27; Waal 2018, 95). In the 1st millennium the short vertical stroke became a dot (Naveh 1973b, 206–207), although the bipunct is found separating words on the Aramaic Tell Fekheriye inscription (Millard & Bordreuil 1982). In Archaic Greek scripts we find the both the vertical stroke and the tri-/bipunct used as word dividers (Waal 2018, 95–96). In Greek scripts where *iota* is a vertical stroke the tripunct is used to separate words, whereas in scripts where *iota* is represented by a vertical stroke, the tri-/bipunct is used (Naveh 1973b, 7 n. 27). The fact that Archaic Greek inscriptions use as word dividers two signs that had passed out of common usage by the 1st millennium BCE points to a transmission date in the 2nd rather than the 1st millennium BCE.

I will return to the significance of word division practices for the history of the transmission of the alphabet in the conclusion. For now, however, it suffices to observe that word-level division by means of the vertical stroke and dots is part-and-parcel of the alphabetic writing system in both Northwest Semitic and Greek. In terms of

⁴ For the suggestion of Aramaic influence in the development of vowel letters, see Woodard (2020, 94–99).

the study of the historical development of alphabetic writing, therefore, it makes a great deal of sense to study the word division practices of the two together, since they are descended from the same original system.

In fact, the net could be cast wider still to include other vowelised alphabets in the ancient Mediterranean, although such is beyond the scope of the present work. It has traditionally been thought that the alphabetic scripts of a number of languages found in the Mediterranean are descended from a Greek prototype, including Phrygian, a number of other Anatolian languages (Carian, Lydian, Lycian, Pamphylian and Sidetic), Etruscan, Italic and Palaeohispanic (Waal 2020, 113–118). The reason for this is that all these scripts have letters for writing vowels, in contradistinction to West Semitic scripts, which lack this feature (Waal 2020, 113–114). However, it has recently been argued (Waal 2020, 118–124) that the Greek alphabet and all other vowelised alphabetic scripts, are in fact descended from another common ancestor which had the innovation of vowel letters. One piece of evidence that points in this direction is the fact that the early Greek alphabets do not have signs for vowels of different lengths. If vowel signs were invented for Greek, we might expect to find the distinction of vowel length to have been made (Alwin Kloekhoest, pers. comm., in Waal 2020, 120–121). If this scenario is correct, it follows that word division in the Greek alphabet is only one representative of the phenomenon among alphabets with vowel letters, and that word division in Phrygian and other vowelised alphabets are independent witnesses of the common ancestor of vowelised alphabets.

Finally, it would also be instructive in future research to bring Latin word division practices into consideration. Classical Latin is distinguished from Greek of the same period by retaining the use of interpuncts to separate words (Wingo 1972, 15), a practice that was abandoned for Greek centuries before. Although Wingo does not include interpuncts in his study of Latin punctuation (Wingo 1972, 14), Latin word division practices share at least some characteristics with Greek and Northwest Semitic, notably the fact that prepositions are only rarely written separately from a following word (Wingo 1972, 16).

1.2.2. Shared environment of the Semitic and Greek speaking worlds

Word division in Greek is not limited to alphabetic writing. In fact, Mycenaean and Cypriot Greek – both written in syllabic scripts unrelated to the alphabet, namely, Linear B and the Cypriot syllabary – attest the phenomenon (Morpurgo Davies 1987, 266). Word-level separation, mostly by vertical strokes, is more reliably found in Linear B than its equivalent in alphabetic texts, and differs in some details from the latter (see Morpurgo Davies 1987, 266–269). Greek written in the Cypriot syllabary also provides evidence of word division, although it is not as frequently found here as it is in Linear B (Morpurgo Davies 1987, 269; Egetmeyer 2010, 528); inscriptions without word division comprise the majority (Egetmeyer 2010, 527). Detailed analysis of word division in syllabic Greek is beyond the scope of the present study. What is significant here is that word division employed along very similar lines to that found

in alphabetic Greek inscriptions is found in ‘genetically’ unrelated writing systems (see further §1.6.2.2 below).

This fact means that very similar principles of word division were either independently developed around the same time in two separate writing communities in the 2nd millennium BCE, or that these principles were in the cultural environment and transcended the barrier between syllabic and alphabetic systems. Given the increasing evidence of multipolar interactions right across the Mediterranean in these periods (Waal 2020, 122), the second possibility seems the more likely. The principles of word division therefore have the potential to shed light on shared attitudes to writing in the 2nd and 1st millennia BCE in the Eastern Mediterranean as a whole. The implications of this study’s findings in this direction are explored in the conclusion.

1.3. Wordhood in writing systems research

1.3.1. Punctuation

Within writing systems research, punctuation has historically held a marginal position. Indeed, the degree to which punctuation might be said to correspond to anything linguistic has been doubted (Neef 2015, 711). Others, however, have advocated its linguistic role (Nunberg 1990). For Nunberg, however, punctuation belongs to the graphical language system only, with no counterpart in the spoken language (Nunberg 1990, 7, 9; as quoted by Krahn 2014, 89–90).

Some modern work on punctuation distinguishes between word division and other punctuation. Thus Wingo, in his study of Latin punctuation in the Classical period (Wingo 1972), interpuncts are not treated. His reason for excluding interpuncts is that ‘word-division was universally used during the period in which we are interested and is therefore to be taken for granted’ (p. 14).

In the present study I present evidence that punctuation in the three writing systems under consideration (Ugaritic alphabetic cuneiform, linear alphabetic Northwest Semitic, alphabetic Greek) is linguistic in denotation. Indeed, I pick up an idea that has a rather long history in European thought, namely, that punctuation is prosodic in denotation. In Renaissance and Early Modern descriptions of the function of punctuation in English, the view that punctuation serves to indicate the manner of oral delivery of a piece of written language – that is, prosody – as opposed to syntax, predominates (Krahn 2014, 63–67, with references). In the 18th and first half of the 19th century, this approach led some to understand punctuation in musical terms (Krahn 2014, 67–68). Syntactic explanations do not predominate until the 19th century (Krahn 2014, 69–74).

1.3.2. Terminology

It is worth separating out four terms that are often used in writing systems research in the same context, frequently with partly or completely overlapping senses, namely

‘script’, ‘writing system’, ‘orthography’ and ‘(natural human) language’ (see Gnanadesikan 2017, 15). With respect to ‘script’ and ‘writing system’, I adopt the following distinctions:

- **Natural human language** A code agreed between two or more human beings for the purpose of communicating information. I take it that both written and spoken language are just as much representations of natural human language as each other. What differs is the level of language that is represented/targeted: spoken language is a phonological representation of human language, while written language is the representation of natural human language by means of visible marks. This is contrary to the view of many linguists, for whom written language is subordinate to spoken language, *e.g.* Bloomfield (1935, 21) (cited by Gnanadesikan 2017, 16): ‘Writing is not language, but merely a way of recording language by means of visible marks.’
- **Writing system** A writing system ‘picks out a certain level – or more often *levels* – of analysis (such as morphemes, syllables and/or segments) and ignores others’ (Gnanadesikan 2017, 16). In a writing system a script is used to represent in written form a particular natural human language (Gnanadesikan 2017, 15).
- **Script**, namely, ‘a set of graphic signs with prototypical forms and prototypical linguistic functions’ (Gnanadesikan 2017, 15; citing Weingarten 2011, 16; cf. Coulmas 1996, 1380). A script is not specific to a language (compare a writing system, which is).
- **Orthography** The set of rules that maps the signs of a script to particular units of the language system, whether semantic, morphological or phonological. The linguistic level of those units is the Orthographically Relevant Level (ORL) (Sproat 2000) of the writing system in question. (The ORTHOGRAPHY here corresponds to the (ORTHO)GRAPHIC CODE in Elvira Astoreca 2020, 55–56.)

The present study is concerned with the linguistic denotation of punctuation, that is, of graphic signs that are used to demarcate suprasegmental units at the levels of the word, the phrase and the clause/sentence. The particular focus is punctuation used to demarcate word-level units. In the terms listed immediately above, I distinguish between the following:

- **Punctuation signs** These belong to the SCRIPT of the writing system. Punctuation signs may take a variety of forms. In this study we will encounter the small vertical wedge ⟨◀⟩ (Ugaritic alphabetic cuneiform), the vertical bar (Linear B and alphabetic Greek) and various interpuncts in linear alphabetic inscriptions in Northwest Semitic languages and Greek: ⟨:⟩, ⟨:⟩, ⟨·⟩. (In including punctuation signs in the script of writing system I follow Elvira Astoreca 2020, 55.)
- **Word dividers** These belong to the orthography of the writing system. Word dividers are punctuation signs used suprasegmentally for the purpose of separating strings of letters from one another. (For the purpose of treating the alphabetic

writing systems under consideration here, a letter is taken to be a grapheme representing a phonological segment, either a consonant or a vowel.) Word dividers are to be distinguished from other suprasegmental markers used to mark out larger sections in a document, such as paragraphs.

1.3.3. Wordhood

As a first approximation, wordhood for a given linguistic domain involves the separating out of minimal units by means of a signal appropriate for that linguistic domain. Wordhood is thereby distinguished from larger unit divisions, such as, for instance, what one might term ‘phrases’ or ‘paragraphs’ (depending on the context), by being the smallest on a scale of unit divisions of a similar nature.

However, while the existence of a linguistic object of this kind, *i.e.* the ‘word’, may appear self-evident, it turns out that the identification of what the ‘word’ actually consists of cross-linguistically is far from straightforward (for discussions of the problem see *e.g.* Horwitz 1971, 6–7; Matthews 1991, 208; Packard 2000, 7–14; Haspelmath 2011). Some linguists have even gone as far as to deny the existence of the word as a linguistic entity altogether (Horwitz 1971, 7, with references). The difficulties that linguists and philologists of Northwest Semitic languages have faced in accounting for wordhood in Northwest Semitic languages can therefore be seen as a species of the broader problem of defining wordhood more generally.

As Haspelmath (2011) points out, wordhood is a concept relative to the particular language(s) under investigation. However, since the languages under investigation here share many structural features, and in several cases are closely related, the problems associated with language-universal notions of wordhood are not fundamental. Indeed, as long as any presupposed notion of wordhood is not held on to too tightly, establishing what kinds of units qualified as words in the minds of ancient writers could help inform discussions of wordhood cross-linguistically.

Key to the problem of ‘wordhood’ in general is that what constitutes a ‘word’ varies according to the linguistic domain under consideration. The present study is concerned primarily with the written – or graphematic – word. In the graphematic domain, in English, ‘words’ are separated by means of spaces to the left and to the right.⁵ Thus the following sequence of characters:

(8) Readersareinthelibrary

can be separated out into the following ‘words’ by the interspersal of spaces:

⁵ In the Northwest Semitic writing systems considered for this study ‘words’ are mostly separated from one another by means of dots – or interpuncts (see further §1.4.5.2).

(9) Readers are in the library

By contrast, in the phonological domain, a sequence of phonological units, *e.g.* phonemes, syllables and feet, participate in a ‘word’ by virtue of sharing a single primary stress or accent, and are bounded by certain junctural phenomena, such as (the lack of) sandhi (see further §1.4.2 below).

1.3.4. Target level of word punctuation

At §1.3.2 I adopted the term ‘writing system’ to denote the system by which written signs are used to represent a particular natural human language. Graphematic words may be separated by various particular signs in the script, or a space without any sign (see further §1.4.5.2 below). What holds these signs together is their function, namely, to demarcate minimal grapheme sequences. As such, the study is not concerned with particular signs in a script, but in terms of a particular function in writing systems, namely, the target of minimal graphematically bounded units in Northwest Semitic writing systems.

A writing system targets in principle a particular level or levels of linguistic analysis (Sproat 2000). Sproat (2000) introduced the term Orthographically Relevant Level (ORL) to describe this linguistic level for a given writing system. The ‘level’ referred to in this term refers to a derivational level (Richard Sproat, pers. comm.), presupposing in principle a derivational linear grammar whereby (morpho-)syntax precedes phonology. Although such a linear grammar is assumed in the present study, particularly in the relationship between prosody and morphosyntax (§1.5), the term ‘ORL’ itself is used here in a broader sense to refer to the linguistic domain relevant for word division, outside of any linear processing. Accordingly, semantics, graphematics, prosody and morphosyntax are all possible target levels for word division in a given writing system (§1.4), and are examined in the following section:

- Semantics (§1.4.1)
- Prosody / phonology (§1.4.2)
- Morphosyntax (§1.4.3)
- Syntax (§1.4.4)
- Graphematics (§1.4.5)

1.3.5. Consistency

In addition to introducing the notion of a writing system’s ORL, Sproat (2000, 16) also makes the following claim:

The ORL for a given writing system (as used for a particular language) represents a consistent level of linguistic representation.

In more expanded terms, the claim of consistency is intended to mean that the ORL of a given writing system ‘is consistent across the entire vocabulary of the language’

(Sproat 2000, 19). The claim is originally conceived of as applying to graphemes representing segmental units, rather than suprasegmental markers, such as word dividers. Nevertheless, it is interesting to consider if the claim of consistency might be said to apply also at the suprasegmental level to the target level of word punctuation. This question becomes of particular interest in the context of the present study, since one of the chief problems associated with word division in Northwest Semitic writing systems is that the word division strategies employed are *inconsistent* (§2.2, §5.2).

1.4. Linguistic levels of wordhood

1.4.1. *Semantic wordhood*

1.4.1.1. *Lexical vs. functional morphemes*

From the perspective of semantics, word division amounts to the breaking up of meaning-bearing units into chunks. A long-recognised distinction between morphemes is that between lexical and functional, based on the nature of their referents, that is, on their semantics (for the distinction, see *e.g.* Sapir 1921, 88–107; Zwicky 1985, 69; Evertz 2018, 139–140):⁶

- **Lexical** (also known as CONTENT WORDS, *e.g.* Golston 1995) include nouns, adjectives and verbs, whose denotation is external referents outside the world of the discourse, whether objects, qualities or events, respectively.
- **Functional** (also known as GRAMMATICAL WORDS, Hayes 1989, 207) including articles, prepositions, relative pronouns, conjunctions, anaphoric pronouns and negatives (*cf.* Selkirk 1986; Golston 1995; Vis 2013), whose purpose is to negotiate the relationships between content words in the linguistic structure.⁷

1.4.1.2. *Integrating the discourse situation*

Missing from the dichotomy of lexical and function morphemes is the existence of a third group of morphemes whose function is to negotiate a relationship between the linguistic framework and the discourse situation. In English these are markers such as, ‘you know’, ‘of course’ etc. Vajda (2005, 404) therefore introduces a three-way distinction in ‘typological primitives’:

- **Referential** ‘denotative meaning, or semantic content that exists independent of the speech act itself’
- **Discourse** ‘connotative, stylistic, pragmatic, or any grammatical feature that mechanically references the speech situation or its participants’
- **Phrasal** ‘grammatical features unrelated to the speech situation itself’

⁶ Cf. Vajda (2005, 403) who distinguishes between ‘syntactic patterns’ and ‘content words’.

⁷ Cf. Vajda (2005, 403): ‘rules capable of expressing meaning when combined with lexemes but lacking intrinsic referential meaning of their own’.

Vajda's distinction between 'phrasal' and 'discourse' morphemes will turn out to be significant in our context, since one of the two types of word division in evidence for Ugaritic alphabetic cuneiform treats functional morphemes differently on this basis. Thus, in one of the two principal word division orthographies in Ugaritic, phrasal morphemes – such as \mathbb{L} b-, \mathbb{M} l-, k- and \mathbb{N} w- – are regularly separated from the surrounding words. By contrast, discourse morphemes are written together with the (usually) prior morpheme (§9.4).

1.4.2. Prosodic / phonological wordhood

1.4.2.1. Prosodic structures in spoken language

The distinction between lexical and function words is relevant not just to semantics. The lexical or functional nature of a morpheme is also broadly correlated with phonological features. As Selkirk (1996, 187) observes, 'Words belonging to functional categories display phonological properties significantly different from those of words belonging to lexical categories.' In particular, they are said to be prosodically 'deficient' in some way, that is, dependent on another morpheme at the phonological level of language.⁸ This is to say that function morphemes are often identified with the class of morphemes known as CLITICS (see Inkelas 1989, 293, and references there).⁹

However, while there is a tendency for function morphemes to be prosodically deficient, that is, CLITICS, there are exceptions to this generalisation. Consider, for example, Ancient Greek enclitics $\varphi\eta\mu\acute{\iota}$ *phēmí* 'say' and $\epsilon\acute{\iota}\mu\acute{\iota}$ *eimí* 'be': these form a single pitch accentual word with a foregoing morpheme despite the fact that they are lexicals (see further §13.5.1.3). In fact, as we shall see, this is an important issue for prosodic and graphematic wordhood in Northwest Semitic, since it is often the case that not only function morphemes, but also lexical items are incorporated into the prosodic (and graphematic) structures of neighbouring morphemes.¹⁰ Furthermore, as Inkelas (1989, ch. 8) shows on the basis of English, not all function words need to be CLITICS. Therefore, while semantics and prosody are related, they are not isomorphic: prosodic features do not follow directly from semantic features. The issue may be resolved through the identification of two types of clitic (Anderson 2005, 13, 23, 31):

- **Phonological (or 'simple') clitics** 'A linguistic element whose phonological form is deficient in that it lacks prosodic structure at the level of the (Prosodic) Word' (Anderson 2005, 23);

⁸ Thus Inkelas (1989, 293) defines CLITICS as 'morphological "words" – with the special property of being prosodically dependent on some other element' (my emphasis).

⁹ In this vein, Hayes (1989, 207) defines a CLITIC GROUP as 'a single content word together with all contiguous grammatical [*i.e.* function] words' (cf. similarly Zec & Inkelas 1990, 368 n. 1).

¹⁰ On the proclitic nature of verb forms in early Indo-European and Hebrew, see Kurylowicz (1959).

- **Morphological (or ‘special’) clitics** ‘a linguistic element whose position with respect to the other elements of the phrase or clause follows a distinct set of principles, separate from those of the independently motivated syntax of free elements of the language’ (Anderson 2005, 31).

Importantly, special clitics may, but need not, also be phonological clitics.

The fact that a morpheme can depend prosodically on another implies the existence of a prosodic structure in which morphemes participate. The first ‘word-level’ prosodic unit might be termed the PROSODIC or PHONOLOGICAL word (cf. Matthews 1991, ch. 11), denoted ω . The prosodic word consists of a prosodically independent morpheme, together with any dependent morphemes. This is the ‘domain in which phonological processes apply’ (Vis 2013 citing Hall 1999; see also DeCaen & Dresher 2020).¹¹

Above the prosodic word, several further levels of prosodic unit have been identified in a hierarchy. Into these prosodic words can be incorporated (see Nespor & Vogel 2007; Selkirk 2011), viz. the PHONOLOGICAL PHRASE (φ), INTONATIONAL PHRASE and UTTERANCE (DeCaen & Dresher 2020) (υ):

$$(10) \omega < \varphi < \iota < \upsilon$$

The present study will be concerned primarily with the lowest ‘word’-level prosodic unit, namely the PROSODIC WORD, although we will occasionally refer to the PROSODIC PHRASE.

1.4.2.2. Characteristics of prosodic words

Across languages, prosodic words have been observed to share the following characteristics:

- A single primary accent/stress;
- Junctural (sandhi) phenomena, that is the sharing of morphological features at morpheme boundaries.

Each of these are now briefly discussed in turn.

ACCENTUATION

One of the consequences of a prosodic word having a single primary accent or stress is that it can incorporate one or more morphemes that carry no stress of their own (Klavans 2019[1995], 129–132). Morphemes with no stress of their own may be in principle of one of two kinds:

¹¹ Nespor & Vogel (2007) differentiate between the CLITIC GROUP and PROSODIC WORD as two different levels of the prosodic hierarchy, devoting a separate chapter to each. For the lack of support for a distinct clitic group level, however, see Hall (1999, 9–10).

- **Unstressable morphemes**, that is, morphemes that may not be stressed or accented under any circumstances;
- **Optionally stressed morphemes**, that is, morphemes that may or may not carry a primary accent depending on the context.

Of the second kind, Klavans (2019[1995], 132, cf. 152) gives the example of object pronouns in English, *e.g.*:¹²

(11) *He sees her.*

Compare the following two prosodic analyses of this sentence:

(12) (*He 'sees her*_ω)

(13) (*He 'sees*_ω) ('*her*_ω)

The reading in (12) involves a single prosodic word, with the primary stress on *sees*. Example (13), by contrast, involves two prosodic words, one with the primary stress on *sees*, the other on *her*. The first one might term the 'unmarked' reading, while the second could be used in a situation where the speaker seeks to contrast the referent of *her* with someone else.

Opposed to optionally stressed morphemes are morphemes that may not be stressed under any circumstances. An example of such a morpheme in English is the indefinite article *a/an*. For the author, a speaker of British English, it is not possible to stress this morpheme, *e.g.*:¹³

(14) *(*I 'want*_ω) ('*an*_ω) ('*apple*_ω)

As we will see, Tiberian Hebrew too has a distinction between optionally stressed morphemes, and those that may not carry the primary accent under any circumstances.

It should be pointed out that the inability to carry a prosodic word's primary stress does not mean that a clitic may not carry an accent or stress of any kind. There are various processes in the world's languages whereby a morpheme may be stressed or accented secondarily (Klavans 2019[1995], 141). In the following example, the sequence φίλος τίς τι *phílos tís ti* carries two accents, but only one primary accent, on φίλος *phílos*. The accent on τίς *tís* is not lexical, but secondarily derived from collocation with enclitic τι *ti*.

¹² Examples adapted from Klavans (2019[1995], 132).

¹³ The one circumstance under which 'a/an' can receive primary stress, and therefore stand as an independent prosodic word is when it is uttered as a citation form. This can happen, for example, when correcting a child or non-native English speaker, *e.g.* 'a apple' corrected to 'an apple' or in discussion concerning the use of 'a'/'an' in phrases such as 'a/an historian'. My thanks to an anonymous reviewer for pointing this out.

(15)

→ φίλος τίς τι εἶπε
 ('philos ,tis ti_ω) ('eipe_ω)
 friend some something said
 'a certain friend said something'

Such a secondary process of accentuation can generate a prosodic word with a primary accent. Consider the following proclitic=enclitic combination in Ancient Greek, where proclitic ἐν *en* carries the retrojected accent from enclitic τινι *tini* (Klavans 2019[1995], 142):

(16) Klavans (2019[1995], 142)

→ ἔν τινι
 'en tini
 in something/someone
 'in something/someone'

We should point out in closing this subsection that Zwicky (1985, 287) states that the accentual test 'should never ... be used as the sole (or even major) criterion for a classification, though it can support a classification established on other criteria'. Zwicky identifies two problems, one 'minor', and the other 'major'. The minor problem is that 'some languages do permit clitics to be accented in certain circumstances'. The major problem is that 'many clearly independent words – e.g. English prepositions, determiners, and auxiliary verbs of English – normally occur without phrasal accent'. The issue that Zwicky is addressing here is the optional nature of the prosodic incorporation of certain morphemes.

Neither of these problems seem to be fundamental. In particular, the 'major' problem, the fact that 'many clear independent words ... normally occur without phrasal accent' is really a problem of definition. On what grounds should these be considered 'clearly independent words'? It seems, rather, that such units can both be considered independent words from a morphosyntactic perspective, and dependent from a prosodic perspective. The major problem, namely, the fact that clitics may be accented under certain circumstances, can be resolved by recognising two categories of accent, one primary, the other secondary.

JUNCTURAL/SANDHI PHENOMENA

Consider the following example sentence in English:¹⁴

¹⁴ On the general validity of sandhi phenomena for discovering prosodic domains, and discussion, see Devine & Stephens (1994, 289–290).

(17) *I have got you.*

This may be split into prosodic words as follows:

(18) ('I have)_ω ('got you)_ω

Under certain circumstances, notably fast speech, sandhi phenomena can be observed to take place within the domain of the prosodic word. In this example, *have* may be reduced to /v/, and the sequence *got you* [gɔt ju:] can be reduced to [gɔtʃa]:

(19)

(I've)_ω (gotcha)_ω
 [(ajv)_ω (gɔtʃa)_ω]

Junctural phenomena can occur at more than one layer of prosodic analysis. This is the case in Tiberian Hebrew, where spirantisation across a morpheme boundary is a phenomenon that occurs at the level of the prosodic phrase, rather than the prosodic word (§1.4.2.6). By contrast, sandhi assimilation of morpheme-final /-n/ in Tiberian Hebrew and Phoenician is more restricted, likely belonging to the level of the prosodic word (§3.5).

1.4.2.3. Construction of prosodic words

All linguistic material that has output at the phonological level must be incorporated into the prosodic structure. This is known as the FULL INTERPRETATION constraint (Goldstein 2016, 48). It means that any prosodically deficient morphemes must be incorporated into prosodic units, minimally, a prosodic word.

For our purposes the most relevant distinction is between internal and affixal clitics, which together with their host project a prosodic word and a recursive prosodic word respectively. An internal clitic is incorporated with its host before any stress assignment, so that the accent is calculated over the host and clitic as a whole. By contrast, an affixal clitic is incorporated after stress assignment on the host; a secondary accent is then projected at the recursive prosodic word level.¹⁵

1.4.2.4. Minimal prosodic words

In the prosodic phonological framework adopted here, prosodic words are composed of prosodic feet (Σ), prosodic feet are composed of prosodic syllables (σ), and prosodic syllables are composed of morae (μ).

¹⁵ For the possible ways in which prosodically deficient morphemes can be incorporated into prosodic words, see Selkirk 1996; Anderson 2005, 46; Goldstein 2016, 48. Note that I follow Goldstein (2016, 45–48) and Anderson (2005) in allowing for the violation of the Strict Layer Hypothesis.

Furthermore, per Figure 1.1 there is a minimality constraint on the prosodic foot, namely FOOT-BINARITY, also known as the Prosodic Minimality Hypothesis (PMH) (for the term, see Blumenfeld 2011). According to the PMH, (prosodic) ‘feet are binary at the moraic or syllabic level of analysis’ (Evertz 2018, 27; see also Prince & Smolensky 2002, 50). Since syllables contain morae, a minimal prosodic foot is bimoraic (Prince & Smolensky 2002, 50) cross-linguistically. In turn, since the prosodic word consists of at least one prosodic foot the minimal prosodic word must also be bimoraic. Although when first proposed the Prosodic Minimality Hypothesis (PMH) was presented as a rule, in the succeeding years evidence has come to light that not all languages necessarily adhere to it. Nevertheless, as Blumenfeld (2011) shows, the hypothesis is not ready to be abandoned, and turns out to be very helpful for the present study.

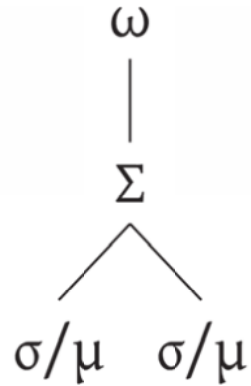


Figure 1.1: Binary structure of the prosodic word

This framework provides a context for understanding the circumstances under which one might expect to find cliticisation of particular morphemes, especially for understanding the difference between morphemes that are always stressless, and those that optionally carry primary stress (cf. §1.4.2.2). This is to say that the crosslinguistic constraint of binarity on the prosodic foot would lead to the expectation that shorter, monomoraic, morphemes should never be capable of carrying primary stress, while morphemes satisfying foot binarity should be capable of doing so.

1.4.2.5. Syllable/foot structure and accentuation in Tiberian Hebrew

Of the languages studied in this monograph, prosodic wordhood *per se* has been studied in both Tiberian Hebrew and Ancient Greek. In this introductory part I illustrate how prosodic words and prosodic phrases manifest themselves in Tiberian Hebrew. The manifestation of prosodic wordhood in Ancient Greek turns out to be more complicated than the generally assumed cross-linguistic picture. This is therefore described in Part IV at §13.3 and §13.5.1.

Vowels in Tiberian Hebrew may be realised phonetically as either short or long. The length of vowels whose length is unspecified at the phonological level may be predicted from its position in its syllable and on its position relative to stress: long vowels occur in stressed syllables (whether open or closed), and in open unstressed syllables, while short vowels occur in unstressed closed syllables (Khan 2020, 268, 279). Phonologically long vowels are realised long. There is, finally, a class of structurally short vowels, that are realised as short even in open syllables. These vowels are marked in pointed texts by *shwa* or *ḥaṭef* (see further Khan 2013, 305–422).

At §1.4.2.4 it was observed that feet are across languages minimally binary, that is, either bimoraic or bisyllabic. We will see that this fact turns out to have important implications for graphematic word division in Tiberian Hebrew (Part III). A canonical phonetic syllable in Tiberian Hebrew is bimoraic, *i.e.* its coda consists of two elements, either a vowel and a consonant, or a long vowel, per the FOOT-BINARITY constraint (§1.4.2.4; Khan 2020, 279, 290). A phonetic syllable's onset consists maximally of one consonant (see Khan 1987, 40). The foot, or phonological syllable, differs from the phonetic syllable in permitting onsets of more than one consonant (see Khan 1987, 40).

The fact that the phonetic and phonological syllables are subject to different constraints means that in mapping from the latter to the former certain adjustments are made. Important for our purposes is the fact that in a phonological syllable of the shape CCVC, the first consonant cluster must be broken up in the transition to the phonetic level (Khan 2020, 349). This is achieved by the insertion of an epenthetic vowel, *i.e.* Cv.CVC (Khan 1987).

As we have seen, a minimal prosodic word is bimoraic at the phonological level. This is to say that it must minimally consist of a bimoraic foot. Accordingly, a monomoraic morpheme at the phonological level, such as one consisting of a consonant and a vowel of unspecified length, does not constitute a prosodic word, and it cannot carry its own primary stress accent, even when realised as bimoraic at the phonetic level.

The rules of accentuation in Tiberian Hebrew can be modelled as taking place after syllabification and phonetic realisation. Accentuation is subject to the following constraints:

- The unit of accentuation in Tiberian Hebrew is the prosodic word. Since a prosodic word must be bimoraic at the phonological level, it follows that a morpheme of the shape CV, where V is a vowel of unspecified length cannot occur as an independent prosodic word carrying its own primary stress;
- Primary stress in Tiberian Hebrew falls in principle on the vowel before the final consonant of the prosodic word, and may therefore occur on either the ultimate or penultimate syllables (Prince 1975, 19; Dresher 2009, 99). In practice, however, the rules for the assignment of primary stress in Tiberian Hebrew are complex (for further details the reader is directed to Prince 1975);
- Adjacent phonetic syllables cannot in principle be accented (for exceptions, see Khan 2020, 496–508);
- The secondary stress in a prosodic word is in principle placed 'on a long vowel in an open syllable that is separated from the main stress by at least one other syllable' (Khan 2020, 458). The calculation is made at the phonetic rather than phonological level, meaning that epenthetic syllables count as intervening syllables (cf. Khan 2020, 460).

1.4.2.6. Prosodic words and prosodic phrases in Tiberian Hebrew

I observed at §1.4.2.2 that prosodic words are most commonly associated in the literature with two phenomena: 1) sharing a single primary accent, and 2) junctural (sandhi) phenomena. In the present study I follow Dresher (1994; 2009) and Khan (2020) in taking *maqfef* to indicate that the units thereby joined share a single main stress (Khan 2020, 509). This is to say that such units constitute a single prosodic word (Dresher 2009, 98). By contrast, prosodic phrases are indicated by strings of prosodic words carrying conjunctive accents (Dresher 1994, 3–4).

For completeness, however, I should point out that not all scholars take this view. Thus Aronoff (1985) implies that prosodic words can consist of elements joined by a combination of *maqfef* and conjunctive accents. Consider, for example, Aronoff's treatment of Isa 10:12 (Aronoff 1985, 44; Aronoff leaves out the initial preposition עַל 'al):

(20) Isa 10:12

← עַל-פְּרִי-גִדְלִי לְבַב מִלְךְ-אַשּׁוּר

ʔ=pry=gdl lbb mlk='šwr

[for=[fruit]=[size [heart [king=Assyria_{NP}]_{NP}]_{NP}]_{PP}]

'for the fruit of the size of the heart of the King of Assyria' (trans. after Aronoff)

Aronoff discusses (20) in relation to the possibility of construct chain recursion: the example consists of a series of noun phrases in construct, as the syntactic analysis shows. The relevance for present purposes is that for Aronoff such series of nested construct chains, consisting, as in this case, of units joined by a combination of *maqfef* and conjunctive accents, constitute single phonological words, just as single two-word construct phrases (Aronoff 1985, 44):

From a phonological point of view, these longer sequences are exactly analogous to simple two-word construct phrases: they form single phonological words.

In Tiberian Hebrew, sandhi phenomena are not limited to sequences joined by *maqfef*, but may extend out to sequences joined by conjunctive accents (see Khan 2020, 536–541, who also discusses exceptions), *e.g.*:¹⁶

¹⁶ Since *paseq* has the effect of blocking sandhi phenomena, for the purposes of this investigation it is treated as if it were a disjunctive accent.

(21) Gen 1:5

וַיְהִי־בֹקֶר ←

w=yhy=bqr

and=become.PST=morning

‘and it was morning’

(22) Gen 19:21

נִשְׂאתִי פָנֶיךָ ←

ns^ʔty pny-k

I_lift.PRF face-your

‘(lit.) I lift your face’

Therefore, while there is a cross-linguistic distinction between internal and external sandhi, with the former pertaining to prosodic words, and the latter to prosodic phrases, the distinction does not appear to hold in Tiberian Hebrew. Accordingly, while *maqfef* sequences are the domain of the primary accent in Tiberian Hebrew, sequences joined by conjunctive accents are the domain of sandhi phenomena. Sandhi *per se* is therefore not an indication of prosodic wordhood in Tiberian Hebrew.

Further complicating the matter is that, for reasons of orthoepy, conjunctive accents were secondarily applied in Tiberian Hebrew to sequences that were unaccented (Khan 2020, 100–101). This was in order ‘to minimize the number of separate orthographic words that had no accent and so were at risk of being slurred over’ (Khan 2020, 100). Furthermore:

The Tiberian tradition, in general, is more orthoepic in this respect than the Babylonian tradition through the Tiberian practice of placing conjunctive accents on orthographic words between disjunctive accents. In the Babylonian tradition, there are only disjunctive accents and the words between these are left without any accent.

As a result, graphematic words whose vocalisation corresponds to their unaccented form, secondarily receive a (conjunctive) accent.

There is, therefore, at least some overlap between *maqfef* sequences and sequences joined by conjunctive accents.

However, the distinction between prosodic words and prosodic phrases in Tiberian Hebrew is still worth making, for the very reason that they are domains in principle of different phenomena, viz. accentuation and sandhi phenomena, and this is the distinction that will be adopted henceforth.

1.4.2.7. Prosodic words in writing systems

Although prosodic words belong, first and foremost, to the prosodic domain, they are highly relevant for graphematic word division, since, especially in the ancient world, prosodic words turn out to be frequent targets of word division in ancient

writing systems. This is especially so in the writing systems to be discussed in the present study, where I argue that the majority of Northwest Semitic writing systems from the late 2nd and early 1st millennia BCE, as well as Archaic and early Classical Greek, use prosodic wordhood as the basis of graphematic word division.

1.4.3. *Morphosyntactic wordhood*

For those brought up in Western European and North American modes of writing, the intuitive notion of ‘wordhood’ corresponds most closely to what one might call the MORPHOSYNTACTIC WORD, corresponding to syntactically free units (Packard 2000, 12).¹⁷ In Bloomfieldian terms, where a word is a ‘minimum free form’ (Bloomfield 1933, 178), word division can be said to separate these minimum free forms. In this framework, ‘free forms’ contrast with ‘bound forms’, that is, elements that are collocationally restricted to particular morphosyntactic classes. Bloomfield (1933, 177) gives the example of the English plural morpheme *-s*, as in ‘*hats, books, cups, etc.*’. This is to say that wordhood is at the boundary of the morphology-syntax interface.

The problem with seeing wordhood in terms of the boundary between morphology and syntax is that it is very difficult to give an account of ‘free forms’ that works cross-linguistically on this basis (Matthews 1991, 208; Haspelmath 2011). Various criteria have been advanced for distinguishing between affixes and clitics, that is, between morphological units and syntactic ‘words’. Wintner (2000, 333–336), for example, lists a number of tests to assess the morphosyntactic status of the Modern Hebrew definite article, including:

- **Binding** Both clitics and affixes may be regarded as bound morphemes, that is, they cannot stand in isolation, and do not bear independent stress.
- **Sandhi** Affixes are more likely than clitics to bring about sandhi phenomena.
- **Movement** Affixes must move with the morpheme(s) to which they are affixed.
- **Semantic idiosyncrasies** It is more common to find semantic idiosyncrasies with affixes than clitics.
- **Selectivity** Affixes have a greater propensity to be selective about their hosts than clitics.
- **Coordination** In coordinate constructions affixes cannot have scope over both elements, and must therefore be repeated on each element. By contrast, clitics have scope over both elements. See Miller (1992b) and Wintner (2000, 335–336).

In addition, Anderson (2005, 33), citing Zwicky & Pullum (1983), observes that ‘[c]litics, but not affixes, can be attached to material already containing clitics’.

These criteria, and others like them, are introduced in order to capture some basic intuitions about the nature of the word, and therefore, the morphology-syntax interface (cf. Matthews 1991, ch. 11):

¹⁷ Packard (2000, 12–13) terms this the ‘syntactic word’. My thanks to Richard Sproat for the reference.

- Words are **internally cohesive** units: other WORDS cannot intervene between the elements of another word. For example, words (and phrases) can be coordinated, *e.g. The man and his dog*, where two phrases (*i.e.*, in this case, two sets of two words) are joined by the conjunction *and*. Thus, a single morphological item cannot have scope over a coordinate construction, so as to modify a phrase consisting of two elements modified by a single morpheme, *e.g. *The man walk- and sing-s*, whereby the present tense suffix *-s* modifies both *walk* and *sing*.
- Words are **minimally free**: Words can stand on their own, and collocate with a wide range of other words, whereas morphological elements depend on other elements and cannot stand on their own. Thus the conjunction *and* can coordinate almost any kind of word, including verbs, nouns, adjectives and other conjunctions, *e.g. If and when I finish this book* Similarly, words with lexical content, such as verbs, nouns and adjectives are freely interchangeable, and, in some languages, at least, such as Latin, can seemingly occur in almost any order. Even in English, such freedom is possible, albeit very marked. Thus *Home will I go* is a possible equivalent of *I will go home*.
- A corollary of the free distribution of words is that words can be coordinated, whereas affixes cannot. For example, **The man walk-ed and -s* is not possible because the morpheme *-s* cannot stand alone, and cannot, therefore, have scope across a conjunction. By contrast, words can have scope across a conjunction, *e.g. in Up and over the hill*, both the prepositions *up* and *over* govern *the hill*.

So far so good. The problem, however, with a hard-and-fast distinction between words and affixes is that it is not hard to find examples that challenge the distinction as laid out immediately above. Thus in English *uni-* and *bi-* in *e.g. unilateral* and *bilateral* look like affixes, in that they cannot be coordinated:

(23) **Both the uni- and bi-lateral discussions went well.*

However, *bi-* and *tri-* are somehow more separable in the following:

(24) ?The bi- and tri-centenary celebrations were both great successes.

(25) The Ugaritic bi- and tri-conanantal prepositions display interesting behaviour.

Similarly, the possessive suffix *-s* looks like an affix in the following:

(26) *The boy's house*

However, the possibility for phrasal scope is shown by the acceptability of the following:

(27) *Justine and Richard's wedding anniversary.*

It would seem then, that the English possessive suffix is something of a hybrid between an affix and a free syntactic unit, a PHRASAL AFFIX, *i.e.* 'morphemes having affixal properties, but which attach to phrases rather than to words' (Miller 1992a, 110).

There is a similar problem with prefixes such as *anti-* in its capacity to have scope across coordination. Miller (1992a, 157) reports the infelicity of the following examples in French:

(28) Miller (1992a, 157)

?? *Une lotion antipuces et poux*
 [= *Une lotion antipuces et (anti)poux*]
 'An antifleas and lice lotion'

(29) Miller (1992a, 157)

?? *Les architectures postromanes et gothiques*
 [= *Les architectures postromanes et (post) gothiques*]
 'Postromanesque and gothic architectures'

By contrast, the following example is acceptable:

(30) Miller (1992a, 157)

C'est un juge anti-dommages et intérêts.

'He is an anti-compensation judge.'

In the above examples, (28) and (29), the prefixes *anti-* and *post-* cannot have scope over both elements of the respective coordinate constructions. By contrast, in (30) the affix *anti-* has scope over both *dommages* and *intérêts*.

English *anti-* demonstrates a similar issue. Compare the following:

(31) ? *The anti-management and discipline proposals.*

(32) *The anti-slavery and trafficking movement.*

(33) *He is anti-food and wine.*

(34) *He is anti-tea and cake.*

Miller (1992a, 157) proposes that the acceptability of (30) is due to the lexicalisation of the phrase *dommages et intérêts*. A similar case might be made to account for the examples (32), (33), (34). It is certainly true that *slavery and trafficking*, *food and wine*

and *tea and cake*, as phrases, are more common than *management and discipline*. Yet they do not seem to behave as monolithic lexical items at least in terms of pluralisation:¹⁸

(35) * *tea and cakes*

(36) * *food and wines*

In the light of these and similar considerations (cf. Matthews 1991, ch. 11; Miller 1992a), for the purposes of the present study I do not assume a binary distinction between morphology and syntax, and therefore, between morphological and syntactic ‘words’. I opt rather to identify MINIMAL MORPHOSYNTACTIC ELEMENTS, or MORPHEMES. These may then be more or less morphological or syntactic: the more dependent a given morpheme is on another (type of) morpheme, that is, the less free it is to collocate with any other morpheme, and the lower the acceptability of other elements intervening, the more affixal it is. Conversely, the less selective and the more phrasal the scope of the morpheme, the more syntactic it is. Morphemes can then be arranged on a cline, from more morphological to more syntactic, per Table 1.1.

Table 1.1: Morphosyntactic cline for English

← morphological	syntactic →
plural -s	prefixes <i>e.g.</i> <i>anti-</i> possessive -’s conjunctions
	nouns
	verbs
	adjectives

1.4.4. Syntactic wordhood

A syntactic word division strategy, in contrast to a morphosyntactic one, would univertbate syntactic phrases. Consider the following sentence:

(37) The man caught sight of his friend, whom he had not seen for a long time

This can be broken up into syntactic phrases along the following lines:

(38) [The man_{SUBJP}] [caught sight of his friend_{VP}] [whom_{REL}] [he_{SUBJP}] [had not seen_{NEGP}] [for a long time_{ADV}]

¹⁸ The phrase *slavery and trafficking* cannot be pluralised because the nouns are abstract in the first place.

Applying a word division strategy that sought to univertate phrases would return something like this:

(39) Theman caughtsightofhisfriend whom he hadnotseen foralongtime

In practice we do not find such word division strategies. We do, however, find something similar in word division strategies that target prosodic phrases. On this see the discussion at §1.5 below.

1.4.5. Graphematic wordhood

1.4.5.1. Graphematic structures in written language

Word division in writing necessarily involves graphematic units. Over the last twenty years there has been a growing appreciation that writing systems have their own internal structure which run parallel to, but may be independent of, those of prosody, morphosyntax and semantics (cf. Evertz 2018, 2–3). A graphematic hierarchy has been proposed by analogy with the prosodic hierarchy in spoken language (Evertz & Primus 2013; Evertz 2018). Thus, the graphematic word $\langle \omega \rangle$ may be said to be made up of graphematic feet $\langle \Sigma \rangle$, which in turn comprise graphematic syllables $\langle \sigma \rangle$ etc., per (40):

(40) $\langle \omega \rangle > \langle \Sigma \rangle > \langle \sigma \rangle$

1.4.5.2. Identifying graphematic words

The present study is primarily concerned with the denotation of the highest element of this hierarchy, the graphematic word. A prerequisite of investigating graphematic words is, of course, to identify them, that is, by specifying the means by which they are bounded.

Most recent discussions of the graphematic word have considered writing systems that separate graphematic words by means of spaces. Thus Evertz (2018, 21) defines the graphematic word as follows (for the same or similar definitions, see Cook 2004, 42; Evertz & Primus 2013, 2; see also for Tiberian Hebrew Drescher 1994):

[A] g[raphematic]-word is a continuous sequence of letters bordered by spaces.

The Northwest Semitic texts considered in the present study do not, for the most part, make use of spaces to separate graphematic words, preferring various kinds of interpuncts, viz. graphemes comprising a number of dots. In the earliest linear alphabetic texts, word division was marked by a short vertical stroke (Naveh 1973b, 206), but dots and spaces are used in later texts (Naveh 1973b, 207; Lehmann 2016, 37–38*). In Ugaritic alphabetic cuneiform, word division is indicated by means of the small vertical wedge (Ellison 2002).

The fact that the same level of unit can in principle be demarcated either by dots or by spaces may be seen by comparison of the Hebrew Bible texts from the Dead Sea Scrolls (DSS), and the Siloam tunnel inscription. In these two sources, the same

word-level unit – as I will argue, the MINIMAL PROSODIC WORD—is demarcated by means of spaces in the DSS, but dots in the Siloam inscription (see Chapter 12).

For the purpose of the present monograph, therefore, I take the graphematic word to be:

Any sequence of characters separated from surrounding characters by spaces or interpuncts.

1.4.5.3. Writing systems without graphematic words

Many texts from the ancient world do not separate word-level units at all. These texts are said to be written in *scriptio continua*. This is especially the case in many Phoenician texts, as well as later Greek inscriptions. Since these texts do not have units corresponding to the graphematic word, they do not fall within the scope of the present study.

It is worth pausing to make two observations. First, it is sometimes claimed that *scriptio continua* is possible only in the context of writing with vowels, and that writing without vowels necessitates word division. Thus Saenger (1997) in his seminal work on word division in the Middle Ages in Europe asserts (p. 9) that:

The uninterrupted writing of ancient *scriptura continua* was possible only in the context of a writing system that had a complete set of signs for the unambiguous transcription of pronounced speech. This occurred for the first time in Indo-European languages when the Greek adapted the Phoenician alphabet by adding symbols for vowels ... Before the introduction of vowels to the Phoenician alphabet, all the ancient languages of the Mediterranean world – syllabic or alphabetical, Semitic or Indo-European – were written with word separation by either space, points, or both in conjunction. After the introduction of vowels, word separation was no longer necessary to eliminate an unacceptable level of ambiguity.

This is not in fact the case: *scriptio continua* is attested well before the arrival of vowel writing in the Greek alphabet. For instance, certain Ugaritic texts from the late 2nd millennium BCE are written without word separation (Tropper 2012, 69). Many Phoenician texts are also written in *scriptio continua* (Steiner 2016, 330).

It is worth adding an additional *caveat*. Lehmann (2005; 2016) has shown that Phoenician and Aramaic inscriptions that had been thought of as being written in *scriptio continua* are, in fact written with spaces separating graphematic words. It is entirely possible that other inscriptions previously thought to have been written in *scriptio continua* are in fact written with words divided by spaces. This is a topic that will no doubt be pursued in future research.

1.4.5.4. Minimal graphematic words

In addition to defining the graphematic word in terms of its boundaries, it may also be defined according to its subcomponents under the graphematic hierarchy (§1.4.5.1):

(41) A graphematic word ‘consists of at least one graphematic foot, which in turn consists of at least one graphematic syllable’ (trans. from Evertz 2016, 392).

Under this definition, a graphematic word must minimally consist of at least one graphematic syllable. What constitutes a graphematic syllable will obviously be writing-system specific. For English and German, Evertz (2018, 47, with references) adopts the following definition:

(42) Every g[raphematic]-syllable has a v[owel]-letter in its peak.

For English and German, therefore, the definition of the minimal graphematic word is dependent on the definition of the vowel letter: if a grapheme is a vowel letter, it will satisfy the definition of graphematic wordhood. Vowel letters can be defined in several ways:

- Phonologically, that is, with reference to correspondences phonological vowels (Evertz 2018, 47);
- Definitionally, simply by listing the set of vowel letters, *e.g.* ⟨a, e, i, o, u⟩ ;
- Distributionally, by comparing which graphemes may be substituted for one another;
- Graphetically, by comparing the features of letter shapes.

The particular definition of the vowel letter adopted for German and English is not important for present purposes. However one chooses to define the vowel letter, the corollary of the definitions (41) and (42) above is that a minimal graphematic word is a vowel letter. This is borne out by the fact that both ⟨a⟩ and ⟨I⟩ (albeit capitalised) are valid graphematic words. Conversely, the formulations make the prediction that a single consonant letter does not constitute a valid graphematic word, a result borne out by the absence of any such words in English or German.¹⁹

Of course, simply meeting the definition of minimal graphematic wordhood does not guarantee existence as a graphematic word: for existence as a graphematic word, a morpheme corresponding to the grapheme sequence must exist in the morpheme inventory of the language represented.

This definition of minimal graphematic wordhood need not apply to writing systems in general. Indeed in English it has been argued (Evertz 2016) that minimal lexical words can be defined in terms of graphematic weight. A writing system could in principle adopt such a definition for all graphematic words. For example, a minimal graphematic word could consist of at least one graphematic syllable.

¹⁹ This excludes, of course, abbreviations marked by a final period. These, of course, cannot stand on their own between spaces, and so are not graphematic words in the same sense that a word such as ⟨a⟩ is.

We will explore a purely graphematic explanation for graphematic wordhood in Tiberian Hebrew at §10.4 below, exploiting the notion of the MINIMAL GRAPHEMATIC WORD, before ultimately rejecting such an explanation in favour of a prosodic one.

1.4.6. Line division

The analysis of line division *per se* is not within the scope of the present study. It does, however, bear upon the question of graphematic word division. This is because multiline texts in the Northwest Semitic writing systems under consideration often do not separate words between lines where we would expect there to be a word division, were the two words to have been written on the same line. It is often the case, therefore, that line division provides *de facto* word division. Whether or not line division functions as word division must be assessed on an inscription-by-inscription basis. The issue is addressed where it arises. For the purpose of transparency, line division in glossed examples is indicated by the subscript symbol_(λ), whereas word division arising by the use of a word divider is indicated by the subscript symbol_(ω).

1.5. Word division at the syntax-phonology interface

1.5.1. Prosodic phrasing and EDGE ALIGNMENT

To this point I have surveyed wordhood at semantic, prosodic, morphosyntactic and graphematic levels. In the chapters that follow, it is particularly the prosodic and morphosyntactic levels that are most relevant, in the interaction known as the ‘syntax-phonology interface’ (cf. Fortson 2008, 9). While these linguistic structures are not isomorphic, there has been shown to be an important relationship between them.²⁰

Cross-linguistically it has been found that the left/right edge of the prosodic phrase aligns with the corresponding edge of the maximal projections of syntactic XPs (Selkirk 1996; Truckenbrodt 2007; for an analysis of Ancient Greek in these terms see Golston 1995). Whether it is the right or left edge of the syntactic structure that aligns with prosodic boundaries is in principle determined by the direction in which recursion occurs in the language in question. English is a right-recursive language, which is to say that if you want to extend an XP, you extend it to the right:

(43) [*The house*_{DP}] [*belongs to me*_{VP}]

(44) [*The house*_{DP}] [*of my friend Joe*_{PP}] [*belongs to me*_{VP}]

(45) [*The house*_{DP}] [*of my friend Joe*_{PP}] [*with the blue lintels*_{PP}] [*belongs to me*_{VP}]

²⁰ On the independence of prosody and syntax see Devine & Stephens (1994, 409). Compare the Copresence Hypothesis, namely, that ‘Prosodic and syntactic structure are autonomous and copresent’ (Inkelas 1989, 14). For the non-isomorphy of prosody and syntax in the Native American language Luhooteed, see also Beck (1999).

Prosodic phrase boundaries, corresponding with pauses in the stream of speech, then align with the right-most edge of the maximal projections of the XPs, DP and PP and VP, respectively:

- (46) (The house_φ) (belongs to me_φ)
 (47) (The house_φ) (of my friend Joe_φ) (belongs to me_φ)
 (48) (The house_φ) (of my friend Joe_φ) (with the blue lintels_φ) (belongs to me_φ)

Other phrasings are possible, depending on the speed of speech (cf. Devine & Stephens 1994, 389). In the case of the last example, a faster rate of speech would result in the subject being grouped as a single prosodic phrase, *i.e.*:

- (49) (The house of my friend Joe with the blue lintels_φ) (belongs to me_φ)

A yet faster rate of speech would result in the whole sentence being grouped as a single prosodic phrase:

- (50) (The house of my friend Joe with the blue lintels belongs to me_φ)

What all these phrasings have in common is that a prosodic phrase boundary, wherever it occurs, is placed at the right edge of a syntactic phrase.

By contrast, it would not in general sound natural to place prosodic phrase boundaries in the middle of syntactic units. It may even make the sentence quite difficult to understand, *e.g.*:

- (51) ?? (The house of my friend_φ) (Joe_φ) (with the blue_φ) (lintels_φ) (belongs to_φ) (me_φ)

There is no such alignment between syntactic XPs and prosodic words, as the following example shows (cf. Shattuck-Hufnagel & Turk 1996, 197–205; Hall 1999):

- (52) *Max wants an apple and three oranges.*

[Max _{SUBJP}]	[wants	[[an apple _{NP}]	and [three
('Max _ω)	('wants _ω)	(an'apple _ω)	(and'three _ω)
		oranges _{NP-NP']_{VP}]}	
		('oranges _ω)	

While the prosodic words *Max*, *an apple* and *oranges* right align with NP constituents, the prosodic words *and three* and *oranges* do not map directly on to syntactic constituents. The issue is that from a syntactic perspective *and*, as a conjunction, is dominated by neither of the immediate NPs *an apple* and *three oranges*, but rather by

the higher level NP', headed by the *v want*. By contrast, from a phonological perspective, *and* and *three* are dominated by one node, headed by *three*.²¹

1.5.2. Syntax and prosodic words

Although prosodic wordhood is not sensitive to prosodic phrasing, from (68) it can be seen that each prosodic word consists of at least one morpheme (cf. Hall 1999, 2).²² This implies that the construction of prosodic words requires the existence of units at the morphosyntactic level, from which prosodic words can be built. This in turn implies a linear grammar whereby morphosyntactic rules are processed before prosodic ones (cf. Wintner 2000; Anderson 2005; Fortson 2008, 9–10). It can therefore be said that the elements of the utterance are arranged according to the rules of morphosyntax; only after morphosyntax is the prosodic status of each morphosyntactic element realised. Through the FULL INTERPRETATION constraint (see above), every morphosyntactic element is incorporated into the prosodic structure, with prosodically deficient items, such as *an* and *and* in (68), incorporated into prosodic word units with neighbouring items capable of heading prosodic words, such as *apple* and *three*.

1.5.3. Syntax and prosodic phrasing in Tiberian Hebrew

Let us demonstrate how this works out in Tiberian Hebrew. As in English, Hebrew has right-branching syntax (Dresher 1994, 18). It is consistent with the general cross-linguistic picture, therefore, that in Tiberian Hebrew conjunctive phrases, at least in principle, align with the right edges of XP^{MAX} (Dresher 1994), and therefore that prosodic phrase boundaries do not bisect syntactic phrases. In the following examples, we see the right edges of the prosodic phrases aligning with the right edges of the PP and VPs respectively:

(53) Gen 3:16

בְּעֵצָב תֵּלְדוּ בָנִים ←		
(b=ʕb _φ)	(tldy	bnym _φ)
[in=pain _{pp}]	[you_shall_bear	children _{vp}]
‘in pain shall you bear children’		

In practice, matters are a bit more complicated. One issue is that it is possible to find VPs apparently bisected by prosodic phrase boundaries at points that do not align with XP^{MAX}. The following sentence is divided into two prosodic phrases, with a phrasal boundary after אֶזְבַּח ‘I will sacrifice’:

²¹ The notion of prosodic HEADEDNESS here is that of Evertz (2018, 96), viz. ‘the hierarchically highest element within a unit’, i.e. ‘the only obligatory element’. The head ‘determines basic properties of other elements within the same unit (sisters in a tree diagram) and of the unit as a whole.’

²² For minimality constraints on the size of the *prosodic* word (as opposed to the *grammatical* word) see Hall (1999, 7–8).

(54) Psa 116:17

לְךָ־אֶזְבַּח זִבְחַ תּוֹדָה ←

(l=k ≡ ʿ-zbh_φ) (zbh twdh_φ)[to=you.SG_{pp}] ≡ [I-sacrifice sacrifice thanks_{v_{pp}}]

‘to you I sacrifice a sacrifice of thanks’

In the next example a prosodic phrase boundary occurs after the subject, resulting in a split VP where the verb belongs to one prosodic phrase, and the direct object to another:

(55) Gen 1:11

וַיִּצְרַח הָאָרֶץ וַיִּשְׂא דְשָׁן ←

(tdšʷ h-ʿrṣ_φ) (dšʷ_φ)[let_sprout_v] [the-earth_{NP}] [sprouting_{NP}]

‘let the earth put forth vegetation’ (RSV)

However, in these examples, the problem may only be apparent. This is because both the verb זָבַח *zbh* ‘sacrifice’ and the verb וַיִּשְׂא *dšʷ* ‘sprout’ can be intransitive. In the case of זָבַח *zbh* this is so in the two other instances where the syntagm *l + x + zbh* occurs, namely, 2Kgs 17:36 and 2Chr 28:23. In the same way, with reference to the other occurrence of the verb וַיִּשְׂא ‘sprout’, at Joel 2:22, the verb is intransitive. Consequently, the division into two phrases at (55) results in a valid VP in the initial phrase, even if its righthand boundary turns out not to be the rightmost boundary of the final VP. Note too that in both examples, the direct object is a noun from the same root as the verb, implying that the NP merely expresses the internal direct object. Accordingly, in both examples, the first prosodic phrase boundary can be viewed as the boundary of a valid XP^{MAX}, even if it is not so in the context of the final sentence as a whole. When the sentence is expanded to the right with an optional object phrase, this is then contained in its own prosodic phrase.

It is perfectly possible, furthermore, to find sentences involving verbs with optional direct objects to include the verb, subject and direct object in a single prosodic phrase. For example, the verb נָדַר *ndr* ‘vow’ may optionally take a direct object (compare Num 6:21 with Num 30:10). The following example involves a transitive case, again with a cognate object. However, unlike the previous two examples, the verb along with its core arguments are included in a single prosodic phrase:

(56) Num 21:2

וַיִּנָּדַר יִשְׂרָאֵל לַיהוָה

(w=ydr ysrʿl ndr_φ) (l=yhwh_φ)and=[vowed_v] [Israel_{NP}] vow_{NP}] [to=DN_{pp}]_{v_{pp}}]

‘And Israel vowed a vow unto the LORD’ (KJV)

It may be relevant in this case that all instances of the *wayyiqṭol* of נדר *ndr* ‘vow’ are of the form *wydr* (+ SUBJ) + *ndr* (see Gen 28:20, Judg 11:30, 1Sam 1:11 and Jonah 1:16). It is possible that, because the object is so fixed a part of the expression, there was a preference for avoiding a pause after the first (subject) NP. Thus, in the only other exactly parallel syntagm, at Judg 11:30, a pause is avoided after the first NP. The reason for avoiding a pause here may well be that, because נדר *ndr* can be analysed as an ‘internal object’, it is at some level of analysis a continuation of the Verb constituent, rather than forming a separate subbranch of the Verb Phrase.

One consequence of the coincidence of prosodic phrases with the edges of syntactic phrases is that construct chains in Tiberian Hebrew are in most cases linked by conjunctive accents (cf. Park 2020, 120–121). A case in point is the conjunctive phrase זבה תודה *zḥw twdh* ‘a sacrifice of thanks’ at (54) above. There are usually good prosodic grounds for cases where this rule is not followed. Park (2020, 121) cites the following instances:

(57) Lev 6:2

← זאת תורת העֹלָה
 (zʔt twrt_φ) (h-ʕlh_φ)
 [this] [law the-burnt_offering_{NP}]
 ‘This [is] the law of the burnt offering’ (KJV)

(58) Lev 6:7

← וְזֹאת תּוֹרַת הַמִּנְחָה
 (w=zʔt twrt_φ) (h-mnḥh_φ)
 and=[this] [law the-grain_offering_{NP}]
 ‘And this is the law of the cereal offering’ (ESV)

In this case the unusual prosodic phrasing can be accounted for with reference to the notion of contrastive focus: a list of laws is being given, and each is introduced with the same formula, namely זאת תורה *zʔt twrt* + *x* ‘This is the law *x*’, where *x* is the name of the law in question. Since the introductory phrase is the same in each case, the purpose of the unusual phrasing is to highlight the part that changes in each case, namely the name of the law.

Other factors are at play in prosodic phrasing beyond the considerations of edge alignment, for details of which the reader is referred to Dresher (1994) and Park (2020). The important point for present purposes is that in general XPs are not bisected by prosodic phrase boundaries. Finally, we should note that clause boundaries coincide in Tiberian Hebrew with boundaries marked by disjunctive accents. Consider the following example of a bicolon in Isaiah:²³

²³ There follows another bicolon, and so arguably this verse could be seen as a tetracolon.

(59) Isa 40:4

← כָּל־נְיָא יִנְשָׂא וְכָל־הַר וְנִבְעָה יִשְׁפָּלוּ

(kl=gy_φ) (yns_φ) (w=kl=hr) w=gb^ch_φ)

[every=valley lift_up.PASS] and=[every=mountain and=hill

(yšplw_φ)

be_low]

‘Every valley shall be exalted, and every mountain and hill shall be made low’

(KJV)

1.5.4. Syntax and prosodic words in Tiberian Hebrew

Now that we have surveyed the syntactic distribution of prosodic phrases in Tiberian Hebrew, let us turn to that of prosodic words. These *may* align with the right edge of XP^{MAX}, thus apparently mirroring the distribution of prosodic phrases. Thus, in the following example, the whole sentence constitutes a single prosodic phrase, and right edge of VP^{MAX} aligns with the right edge of the conjunctive phrase:

(60) Psa 106:11

← וַיִּכְסּוּ־מַיִם צָרֵיהֶם

(w=yksw≡mym sry=hm_φ)**and=cover.PST.3PL≡waters** enemies=their‘**And the waters covered** their enemies’ (KJV)

However, the prosodic phrase is divided into two prosodic words, the first of whose right edge aligns with the right edge of the maximal projection of the subject NP.

The following is parallel, with prosodic word division aligning with the right edges of the PP and of the sentence as a whole:

(61) Job 22:24

← וְשִׂית־עַל־עָפָר בְּצָר

(w=šyt=^cpr_ω) (bšr_ω)and=[set_v]=[on=dust_{PP}] [gold_{NP}]

‘Then shall you lay gold in the dust’ (after KJV)

Unlike prosodic phrases, however, alignment with the right edge of XP^{MAX} is not a general requirement of prosodic words. This can be seen in the fact that an NP can readily be split across two prosodic words. Thus in the following two examples the subject NPs, אַף יַעֲקֹב *y^cqb* ‘anger of Jacob’, and מַלְאֲךְ אֱלֹהִים *ml^k’l^hym* ‘angel of God’, respectively, are split across two prosodic words:

(62) Gen 30:2

וַיִּחַר-אַף יַעֲקֹב בְּרַחֵל ←

(w=yħr ≡³p_ω) (y^ʿqb_ω) (b=rhl_ω)and=[kindle.PST_v] ≡[anger PN_{NP}] [at=PN_{pp}]

‘And Jacob’s anger was kindled against Rachel’ (KJV)

(63) Gen 21:17

וַיִּקְרָא מַלְאֲכֵי אֱלֹהִים | אֶל-הַגֵּר ←

(w=yqr³_ω) (ml^ʿk_ω) (ʾl^hym_ω) (ʾl=hgr_ω)and=[called_v] [angel God_{NP}] [to=PN_{pp}]

‘And the angel of God called to Hagar’ (KJV)

Evidence will be presented, however, that prosody must make reference to the syntax at least twice, once in the construction of prosodic words, where prosody checks morpheme boundaries, and once in the construction of prosodic phrases. This is argued to be the case (§8.2) because prosodic words, at least in Tiberian Hebrew and Ugaritic, can incorporate morphemes across syntactic phrase boundaries. If so, this is evidence that prosodic phrases are sensitive to syntactic phrasing only at the boundaries of prosodic words; if two morphemes either side of a syntactic phrase boundary have already been combined into a single prosodic word, a phrase boundary does not occur there.

1.6. Previous scholarship

1.6.1. Northwest Semitic

In the orthographies of both ancient and modern Northwest Semitic languages there is agreement that word division is not (morpho-)syntactic in the way that it is, for example, in English or German (cf. *e.g.* Ravid 2012, 111–112; Lehmann 2016). This emerges clearly in (2) above from the fact that the prefix forms $\text{-} \dot{\text{ו}}$ *w-*, $\text{-} \text{ב}$ *b-* and $\text{-} \text{ה}$ *ha-* are translated with words with independent orthographic status in the English translation, namely, *and*, *in* and *the*. Word division in Northwest Semitic orthographies must therefore mark out some other kind of unit, larger than the morpheme. However, this is where the area of scholarly consensus comes to an end.

One set of scholars hold that word division morphosyntactic units, albeit not the same morphosyntactic unit targeted by word division in English. Donner & Röllig (1968, 2), for instance, presuppose a syntactic explanation in their discussion of orthographic proclitics in the Phoenician ³Aḥirom inscription (KAI 1):²⁴

²⁴ Original: ‘[D]as Relativum $\dot{\text{ו}}$... ist wegen seiner engen syntaktischen Verbindung mit dem Verbum diesem ohne Worttrenner vorangestellt’.

The relative \uparrow ..., because of its close syntactic connection with the verb, is placed before it without a word divider

In the same vein Millard (2012b, 25) states that Hebrew scribes practiced word division ‘normally with a point after each word, except when they were bound together grammatically.’ The implication, at least from the use of the term ‘grammatical’, is that it is morphosyntactic, rather than prosodic, factors that lead to the obligatory orthographic cliticization of words like $\text{-בִּ} b-$ ‘in’, $\text{-הַ} ha-$ ‘the’ and $\text{-וַ} w-$ ‘and’, and, furthermore, that orthographic wordhood in Hebrew is a function of grammar, rather than prosody. Similar, at least in this respect, is Aronoff (1985), who argues that Masoretic punctuation as a whole has a syntactic basis.

By contrast, Friedrich, Röllig & Amadasi Guzzo in their grammar of Phoenician-Punic (Friedrich, Röllig & Amadasi Guzzo 1999, 146, §219) relate the issue of word division explicitly to the question of accent, that is, prosody: they state that in the oldest phase of the language the governing noun of a genitive construction retains its original vocalisation.²⁵ This is to say that the accent is retained. The implication is that, were the accent to have been lost, we would find genitive constructions written as a single graphematic word.²⁶ As we will see (§3.6), this logic does not in fact follow, at least as far as it depends on comparison with Tiberian Hebrew, since not all such chains form single prosodic words there.

Robertson (1994, 361–363) takes a similarly prosodic approach in her treatment of Ugaritic literary material. Although she does not finally decide exactly what kind of unit is demarcated, Robertson suggests that word division there is ‘based at least in part on a sound length value which may have been related to some aspect of verse structure’ (Robertson 1994, 363).

Other scholars are more ambivalent. Lehmann (2016, 37*) puts the matter as follows:

[T]he signs generally known as word dividers by no means mark lexemic word boundaries in every case. Rather, these often seem to be mere delimitation marks for prosodic breath units or morpho-grammatical and other units.

Lehmann leaves open the precise purpose of word division in West Semitic orthographies. He concludes, in general terms, that the word divider is a ‘low-level supra-segmental graphic delimitation mark’, proposing the term ‘low-level graphic separation mark’ to describe it (Lehmann 2016, 38*). Lehmann mentions two specific

²⁵ Friedrich, Röllig & Amadasi Guzzo (1999, 146, §219): ‘Die alten Inschriften mit Worttrennung schreiben Genitivverbindungen gewöhnlich ungetrennt wie *ein* Wort ... Die Stellung des regierenden Nomens im tonlosen ‘Status constructus’ hat im Phönizisch-Punischen der älteren Zeit den ursprünglichen Vokalismus des Wortes erhalten.’

²⁶ In a similar vein, compare the identification of proclitic prepositions (Friedrich, Röllig & Amadasi Guzzo 1999, 180, §251) and the representation with *shwa* of the conjunctions ⟨w⟩ and ⟨k⟩ (Friedrich, Röllig & Amadasi Guzzo 1999, 185–186, §257).

possibilities, namely, that word dividers serve as delimitation marks for a) ‘prosodic breath units’, or b) ‘morpho-grammatical’ units, thereby identifying the domains in which word division may be operating as either prosody (= phonology) or morpho-syntax.

Finally, prosody and morphosyntax sit side-by-side in Drescher’s analysis of Biblical Hebrew orthography: in Drescher (1994, 9), Drescher distinguishes ‘grammatical’ clitics, on the one hand, which ‘are morphemes that obligatorily cliticize onto their host’ and ‘may never stand as independent words’, from ‘prosodic clitics’, which are ‘potentially independent words which are cliticized in particular situations’.

As the foregoing brief survey shows, the literature on word division in Northwest Semitic is characterised by a lack of consensus, and in some quarters, by a certain vagueness, concerning the target of word division. This the case both on whether graphematic words correspond to (morpho-)syntactic or prosodic units, and on what characterises a prosodic word. To my knowledge, with only one exception, to be discussed immediately below, none of the scholars who have examined the orthographic word in Northwest Semitic writing systems have considered the possibility that it might correspond to the PROSODIC WORD as defined in the phonological linguistic literature (§1.4.2).

The exception just mentioned is Drescher’s work on prosody in the Tiberian Hebrew tradition (esp. Drescher 1994; 2009). For Drescher the prosodic word corresponds not to the unit separated by spaces in the consonantal text, but to that which is joined by *maqqef*. *Maqqef* is a dash-like grapheme ⟨⁻⟩ which is used to join two words of the consonantal text, e.g. Gen. 1:2 **עַל־פְּנֵי הַמַּיִם** *ʿl=ḥny h-mym* ‘over the face of the water’, where the preposition **עַל** *ʿl* ‘on, over’ and the noun **פְּנֵי** *ḥny* ‘face’ are joined by *maqqef*.

Since *maqqef* was introduced in the early medieval scribal tradition, this unit has no counterpart in ancient Northwest Semitic writing. For Drescher, the orthographic word, that is, the word separated by spaces in the pre-medieval manuscript tradition corresponds to the unit separated either by spaces or by *maqqef* in the medieval tradition. For Drescher this unit is rather a POTENTIAL PROSODIC WORD. This insight turns out to be very helpful for Tiberian Hebrew, not least since there we have direct access to a prosodic parsing of the Biblical Hebrew material in the form of the tradition of accents, against which we can test the prosodic status of the orthographic word. However, we lack this information for Ugaritic and the early epigraphic sources for Northwest Semitic writing.

This does not mean, however, that nothing can be said about the linguistics of word division in these purely epigraphic sources. This is because the word division strategy adopted for a given language will have a profile corresponding to the nature of wordhood at that level (§1.4). Section 1.7 outlines the methods that will be used in the course of this study to achieve this. Before that, however, I provide a brief survey of the study of graphematic word division in Ancient Greek and the evidence for the consensus opinion there.

1.6.2. Ancient Greek

1.6.2.1. Alphabetic

For the most part Ancient Greek inscriptions written in the alphabet eschew the use of word dividers (Morpurgo Davies 1987, 270; Golston 1995, 347; Wachter 2010, 53; Steele 2020, 139). However, an important subset – many of which hail from Eastern Ionia (*i.e.* the West coast of modern Turkey) and the area around Argos (Wachter 1999, 365) – do make use of word dividers to mark out word-level units (Morpurgo Davies 1987, 270; Devine & Stephens 1994, 326–330; Wachter 1999, 365–367; Goldstein 2010, 55–56; Threatte 2015[1980], 79–80). In contrast to Northwest Semitic studies, for Ancient Greek the suggestion that a subset of inscriptions separate word-level units on the basis of prosody has been proposed at least as early as Kaiser (1887). Furthermore, a considerable body of recent scholarship has either looked at this question directly (Morpurgo Davies 1987; Devine & Stephens 1994; Wachter 1999) or has relied on it in order to elicit facts about the prosodic word in Ancient Greek (*e.g.* Golston 1995, 347; Luraghi 2013; Vis 2013; Goldstein 2016, 67–68). Units marked out at the level of the word have been identified with ‘accentual units’ in Greek (Wachter 1999, 366). In fact, univertation in the written records of several Old Indo-European languages, not just Greek, is taken to be indicative of accentual dependence (Clackson 2007, 168).

A good Greek example is the ‘ $\phi\eta\delta\iota\acute{\epsilon}\sigma\tau\alpha\varsigma$ ’ inscription from Archaic Argos (Probert & Dickey 2015). The identification of graphematic words with prosodic words suggests itself as a first approximation by the fact that certain function words are univertated with neighbouring morphs and lexical items, while lexical words are generally separated from surrounding morphs by word dividers (§1.1). This is consistent with the cross-linguistic tendency, as we have seen, for function words to be prosodically weak (§1.4.2.1).

It has long been recognised that Greek function words fall into two groups – so-called pre- and post-positives (*cf.* Dover 1960) – and that these items are often subject to certain restrictions on their position in the sentence and/or clause in which they sit. In particular, certain items have a strong tendency, or are obliged, to appear in so-called ‘second’ position, a phenomenon referred to as ‘Wackernagel’s Law’. These issues are addressed in greater detail in Chapter 16. For the time being, however, it is enough to observe that in SEG 11:314 prepositives are univertated with sequences including a following lexical, whilst postpositives are univertated with sequences including a preceding lexical. Thus postpositive and enclitic TE *te* is written together with the preceding XPEMATA *khrẽmatá*, whilst prepositives KAI *kaí* and TA *tá* are written before it (§1.1).

This is paralleled more broadly in Ancient Greek inscriptions with word-level punctuation. In Attic inscriptions, for example, it has been observed that the prepositives the article and the conjunction KAI *kaí* are liable to be written together with the following word without word divider (Morpurgo Davies 1987, 271; Threatte 2015[1980], 80). In her analysis of the *Teiae Dirae* inscriptions from the city of Teos in

Ionia (now the west coast of Turkey), Morpurgo Davies (1987, 271) observes that prepositives, including forms of the article, prepositions and conjunctions are not followed by punctuation marks, whilst the postpositives ΔΕ *dé* and ΑΝ *án* are.

Returning to SEG 11:314, further support for the identification of graphematic words with prosodic words in this inscription comes from the fact that the sequences so demarcated do not make any sense as syntactic units, but are very plausible prosodic units (for this point in general see Devine & Stephens 1994, 287). Determiner phrases are generally unverbated in this inscription, of which the examples immediately above provide several examples.

Despite a number of scholars either proposing or assuming that graphematic words correspond to prosodic words, the issue of inconsistency still raises its head in some quarters of the literature, *e.g.* Gagarin & Perlman (2016, 51):

The use of word-division is neither consistent nor systematic, though it generally does not occur between natural groupings of words, such as article and noun or preposition and noun.

The fact that word division in a subset of Ancient Greek inscriptions is inconsistent will be used to support the claim of this study that this is the target of word division in an important group of Northwest Semitic inscriptions (§1.7.4; Part IV). However, while the prosodic wordhood of graphematic words in these inscriptions is uncontroversial, there are in fact some *prima facie* difficulties with a straightforward identification of graphematic words and prosodic words in the Greek epigraphic material. These difficulties have not, to my knowledge, been addressed directly (§13.5). Attempting to address these, however, has the potential to pay dividends in terms of further refining our understanding of the purpose of word division, and how this practice is interpreted in different linguistic circumstances.

1.6.2.2. Syllabic

We saw at §1.2.2 that word-level demarcation is not the preserve of alphabetic writing: Greek written in both the Linear B and Cypriot syllabaries also provides evidence of word-level demarcation. In Linear B, an important point of overlap with alphabetic word division is that function words, including polysyllabic ones, are frequently unverbated with a neighbouring word (Morpurgo Davies 1987, 267). On this basis, Morpurgo Davies (1987, 267–269) argues that orthographic words for the most part correspond to accentual groups, or, in our terms, prosodic words. On the Cypriot side inscriptions with word-level demarcation often have pre- and post-positives written together with a neighbouring word, although this is not always the case (Egetmeyer 2010, 528); in general what is punctuated in these inscriptions is taken to be the prosodic word (Devine & Stephens 1994, 326).²⁷

²⁷ Or, in Devine & Stephens' terms, the 'appositive group'.

1.7. Method

1.7.1. Introduction

The primary goal of the study is to establish which strategy or strategies – that is, prosodic, (morpho-)syntactic, semantic and graphematic – best account(s) for word division in the Northwest Semitic writing system(s) in the late 2nd and early 1st millennia BCE. A general difficulty to overcome, however, is that different word division strategies can yield the same results. Thus, for English, were *an* and *apple* written together as one orthographic word, viz. *anapple*, this could be for prosodic or semantic reasons, as we have seen. There is a need, therefore, to find a way to disambiguate between possible underlying explanations.

1.7.2. Difficulties

1.7.2.1. Material preservation

A major obstacle is identifying word dividers in the first place. Punctuation marks in our texts are subject to a higher degree of uncertainty compared to other classes of signs (Morpurgo Davies 1987, 270). The reasons for this are various, but include the following:

- Punctuation marks are more vulnerable to damage compared to other signs, owing to the fact that they are epigraphically less prominent. Punctuation marks often consist of a single line, a series of dots in a line, or even a single dot. They are therefore one (or zero) dimensional. By contrast, full letters are usually two dimensional. Therefore, if a punctuation mark suffers the same absolute amount of damage as a letter, the detrimental effect on its legibility will, all things being equal, be greater.
- Given that an overarching rationale for punctuation has, at least in traditional analyses, been hard to find, editors may have sometimes avoided seeing word dividers where they in fact exist. A case in point is that of an inscription from Dreros on Crete, where punctuation marks were interpreted as letters in the first edition (compare Van Effenterre 1946; with Van Effenterre 1961). For other instances where editors have not seen, or not agreed on seeing, punctuation, see *e.g.* Morpurgo Davies (1987, 277 n. 21).

These issues are referred to as ‘epigraphical and editorial “noise”’ by Devine & Stephens (1994, 389). This ‘noise’ is, however, perhaps less of a problem now than it has been in the past. This is mainly because editors are increasingly aware of the potential significance of punctuation (see *e.g.* Probert & Dickey 2015 where considerable attention is paid to this issue). Furthermore, although the presence of individual punctuation marks may be affected, there is enough clarity to be clear on the general trends, and, for these purposes, on the existence of potential difficulties. As high quality digital photographs of the inscriptions become more widely available, the problem will continue to diminish.

1.7.2.2. Direct access to prosody

An ancient language is directly accessed only via the written record of inscriptions and documents which have been preserved from the period when it was spoken. Any attempt to identify the principles of orthographic word division therefore runs the risk of circularity. In particular, it might be supposed that the lack of direct access to (representations of) the prosodic level of language necessarily precludes saying anything about prosody at all, thus, excluding the possibility of assessing a major component of the language system. As will be seen, however, there are good grounds for attempting to derive information about the linguistic level of word division in ancient writing systems, beyond merely the graphematic level (cf. also Fortson 2008, 6–9).

The evidence adduced in this study for the purpose of ascertaining the ORL of graphematic word division falls into two categories, internal and external. These are now discussed in turn.

1.7.3. Internal evidence

Lines of internal evidence include the following:

- Graphematic/phonological weight of unverbated morphemes;
- Correspondence between graphematic wordhood and the scope of sandhi phenomena;
- Correspondence between graphematic word boundaries and morphosyntactic boundaries;
- Correspondence between graphematic word boundaries and syntactic boundaries.

1.7.3.1. Graphematic/phonological weight of unverbated morphemes

The assignment of prosodic wordhood shares certain characteristics across languages (§1.4.2). One point of commonality is that function morphemes have a tendency to be prosodically deficient. Furthermore, prosodically deficient morphemes are often minimal in ways that can be measured without access to prosody *per se*, viz. the fact that they are often short. Compare, for example, the English function morphemes *for* and *before*: *for* can be prosodically deficient, as in the following example:

(64) (I 'went_ω) ('out_ω) (for 'tea_ω) (be'fore_ω) ('dinner_ω)

For and *before* are functionally equivalent in this example in that they are both prepositions. Yet *for* is prosodically weaker than *before*: the former carries no stress accent, whereas the latter may carry either primary or secondary stress depending on the particular point being made.

If in written English words were divided on prosodic grounds, we might obtain the following result, where *before* is separated, but *for* is not:

(65) *Iwent out fortea before dinner.*

The consequence of these observations for present purposes is that it should be possible to infer principles of word division on circumstantial grounds from the kinds of units that are separated. Thus, finding that graphematically dependent words were both phonologically short and semantically functional might constitute circumstantial evidence that word division operated on the basis of prosody, since these are the kinds of morphemes that are prosodically deficient cross-linguistically.

It is also possible to use the phonological/graphematic weight of unverbated morphemes to distinguish between phonological and semantic word division. Thus, on the one hand separating words on semantic and prosodic grounds yields the same result for the representation of *for tea* in (64) in both cases, namely, *fortea*. However, the two underlying word division strategies can be distinguished by comparing the treatment of non-prosodically weak forms. Thus in *before dinner* semantic word division yields a single orthographic word *beforedinner*, while prosodic word division yields two separate words *before dinner*.

1.7.3.2. Sandhi phenomena

We noted at §1.4.2.2 above that prosodic wordhood is cross-linguistically associated with junctural phenomena, that is, the sharing of phonological features across a morpheme boundary within the prosodic word, but not at its boundaries. Assimilation of morpheme-final /-n/, is associated in Tiberian Hebrew and Phoenician with the level of the prosodic word, and, where this coincides with graphematic word boundaries, will be used as evidence for the correspondence between graphematic words and prosodic words.

1.7.3.3. Consistency and the correspondence with morphosyntactic boundaries

An important line of evidence regarding the target of word division in a given writing system is the level of its morphosyntactic consistency, that is, the regularity with which word division corresponds with the boundaries of what one might term 'dictionary words'.

Those familiar with the writing systems of languages with their roots in Western Europe have come to expect word division to be consistent at the level of morphosyntax. In principle, consistent morphosyntactic word division would treat morphemes equally according to their morphosyntactic status rather than phonological proportions. Thus, for example, all prepositions should be treated in the same way, either by writing them independently, or by unverbating them with a neighbouring morpheme. This is because all prepositions play the same morphosyntactic role, regardless of their phonological size.

In modern written English, for example, prepositions are always written as separate graphematic words, regardless of their prosodic status, *e.g.*:

(66) (I 'went_ω) (for 'tea

(67) (I 'went_ω) (be'fore_ω) (the 'judge_ω)

In (66), *for* is proclitic, and forms a single prosodic word with *tea*. By contrast, in (67), *before* has its own primary stress, and constitutes a prosodic word in its own right. From a morphosyntactic perspective, however, *for* and *before* are both prepositions, and both govern their respective NPs *tea* and *the judge* in the same way. This parallel morphosyntactic status is reflected in the system of word division, where both *for* and *before* are written as separate graphematic words.

A long-standing issue in the analysis of early Northwest Semitic inscriptions is the apparent lack of systematic relationship between morphosyntactic boundaries and regular word division of this kind. Note, for instance, that three out of four of the following ‘principles’ provided by Millard (1970, 15) in his summary of Northwest Semitic epigraphic evidence for word division are prefaced by ‘sometimes’:²⁸

Words are separated from each other except for

(1) single-letter proclitic particles (e.g. k, w, b)

(2) sometimes the *nota accusativi*

(3) sometimes bound forms (construct+regnant noun, infinitive absolute+regnant verb)

(4) sometimes the third person plural suffix

The morphosyntactic inconsistency of word division in Northwest Semitic writing systems is usually seen, implicitly at least, as a problem. Thus handbooks and other scholarly literature on word division in alphabetic cuneiform are wont to present word division there as inconsistent, and pay little attention to it. Thus, for example, Sivan (2001, 11) states ‘The Ugaritian scribes were not consistent in dividing words’ (cf. similar remarks in Wansbrough 1983, 222; Huehnergard 2012, 22).²⁹ Tropper, in his monumental grammar (Tropper 2012), which runs to some 1068 numbered pages, affords barely three sides to word division (pp. 68–70), where he simply describes the phenomena.

Claims of inconsistency in word division in early Northwest Semitic inscriptions betray an underlying assumption, namely, that words *should* be separated from one another on the basis of morphosyntax/semantics, for it is only according to these principles that word division is implicitly being measured, e.g. Huehnergard (2012, 22):

²⁸ Cf. Lehmann (2016, 39*), who states: ‘There would appear to be an increasing disdain for dividing devices in epigraphic research, especially spaces, the more when these are only very small and inconsistent gaps. This may be related to a certain methodological helplessness. The minute size and occasional unsteadiness of such spaces seem to undermine all rules of word separation vs. non-separation hitherto known from dotted writing with graphic separation marks (Millard 1970, Naveh 1973).’

²⁹ Wansbrough (1983, 222): ‘The problem there is the random and hence indeterminate functional load of that device [i.e. the word divider].’

Noun phrases ... may appear without a word divider between them, as in bt b^cl [the house of Ba^l]; but there are many other exceptions where it is not clear why the word divider has been omitted. (My emphasis)

In the light of Sproat's claim that writing systems, in principle, at least, target a consistent level of linguistic representation (Sproat 2000; §1.3.5), it is worth considering whether in fact word division in these writing systems is consistent, but merely represents a level of language different from that of morphosyntax. Indeed, the present study sets out to do just this, and identify the ORL of word division in these writing systems.

To embark on such a project is not, of course, to deny that the documents considered in this study were written by humans, or to assume, implicitly or otherwise, that humans are not capable of making mistakes. We will, of course, see some level of inconsistency on the grounds of human error. However, discussions of word division in ancient West Semitic orthographies, particularly when considering the epigraphic material, tend to emphasise the inconsistency at the expense of pointing out the consistencies. The impression one gets from the literature is that they simply were not very good at what they did. This, in my view, does a great disservice to the writers of these documents, many of whom must have been highly skilled at their craft.

My claim of consistency in the separation of word-level units in these writing systems rests on the observation that word division in Northwest Semitic inscriptions is not random. For example, we do not regularly find word dividers dissecting morphemes, such as the lexical root. Thus while Millard's categories 2, 3 and 4 in p. 42 above are evidence of a level of inconsistency, the ubiquity of the orthographic dependency of 'monoconsontal prefixes' is rarely highlighted. Indeed, to my knowledge there is only one early Northwest Semitic orthography where these prefixes are written separately, namely, the Ugaritic orthography (on which see §9.4 below). Furthermore, the remarkable consistency of treatment of these items is at odds with the orthographies of modern European languages, where we do not find equivalent morphemes written as prefixes, and is suggestive that some underlying principle is at work. In seeking to establish the principles underlying word division in West Semitic orthographies, therefore, it seems reasonable to start from the consistencies, and to work out from there to seek to account for the inconsistencies.

Key is the possibility of deriving morphosyntactically 'inconsistent' word division from the application of regular, albeit non-morphosyntactic, principles. This is *par excellence* the case in respect of word division according to prosodic words. Consider the following sentence:

(68) I want an apple and three oranges.

This can be given the following prosodic analysis:

(69) (I 'want ω) (an 'apple ω) (and 'three ω) ('oranges ω)

A word division strategy demarcating prosodic words would therefore yield:

(70) *Iwant anapple andthree oranges.*

There is, however, another possible prosodic output, where *and* occurs in a strong (*i.e.* accented) form:

(71) (I 'want ω) (an 'apple ω) ('and) ('three ω) ('oranges ω)

This prosodic variant emphasises the conjunction *and*, and, in consequence, the addition of the three oranges. Prosodic word division of (71) would obtain the following:

(72) *Iwant anapple and three oranges*

Now let us conduct a thought experiment. Imagine two separate documents representing *I want an apple and three oranges*. For both documents prosodic word division is used, but one document represents the prosodic variant in (69), while the other represents the prosodic variant in (71). The results would look inconsistent, but in actual fact the orthographic principles underlying word division in both cases are consistent, but applied to represent two different prosodic variants.

Note, furthermore, that, at least for the present author, the following prosodic variant, with a strong variant of *an* is not felicitous:³⁰

(73) *(I 'want ω) ('an ω) ('apple ω) (and 'three ω) ('oranges ω)

The net result of the thought experiment, therefore, is that prosodic word division of *I want an apple and three oranges* will always result in *an* and *apple* written as one word – *i.e.* morphosyntactic consistency) – whereas *and* and *three* may or may not be written as one word – *i.e.* morphosyntactic inconsistency).

In the case of prosodic word division, therefore, we might expect three types of unit:

- Function morphemes that are never written as separate words (the case of *an* in the above examples);

³⁰ Example (73) is infelicitous because emphasis is contrastive with *three*. This is to say that it is the quantity of apples and oranges respectively that is being compared. However, *an* only implies singularity, but does not denote it: *one* is needed to denote singularity. Accordingly, (73) would need to become *I want one apple and three oranges* to be felicitous.

- Function morphemes that may be written either as independent words or together with other morphemes (the case of *and* in the above examples);
- (Morphemes (functional or lexical) that are in principle graphematically independent, and are only written together with prosodically deficient morphemes (e.g. *before* in (64) above).

1.7.3.4. Correspondence with syntactic boundaries

At §1.5 above it was observed that there is a relationship between prosodic phrasing and morphosyntax. Specifically, the left/right edges of prosodic phrases are expected to align with the respective edges of the maximal projections of corresponding syntactic phrases. Therefore, while the univertation of smaller functional morphemes is potentially indicative of graphematic wordhood according to prosodic words, general alignment of prosodic words with the left/right edges of syntactic phrases can be taken as evidence of graphematic wordhood targeting prosodic phrases. By contrast, the lack of such alignment can be taken as evidence against such a correspondence. This line of argumentation will turn out to be important both for discounting graphematic word~prosodic phrase correspondence in Ugaritic (§8.2) and in favour of this correspondence in the case of the Phoenician inscription KAI 10 (Chapter 4).

1.7.3.5. Distinguishing word division strategies

How should we test which account is correct? If division by FUNCTION + CONTENT WORD units arises via the phonological level, so as to mark accent groups, we might expect to see at least one of the following phenomena:

- Separation of units according to phonological properties rather than their syntactic or semantic properties. In this case, we would expect to see words with the same kind of function, e.g. PREPOSITIONS, treated differently based on their phonological properties.
- Evidence of sandhi phenomena within marked word-level units, but not between them.

Conversely, if the orthography arises via the lexical semantic and morphosyntactic level of representation of the language, we would expect:

- Separation of units according to semantic and syntactic properties, independent of their phonology. In this case, we would expect to see words with the same kind of function treated in the same way, regardless of their phonological properties. For example, we would expect to see all prepositions treated in the same way, independent of their capacity to carry stress accent.
- Lack of external sandhi phenomena.

Word division could target linguistic levels other than morphosyntax or prosody. It could, for example, target the morpho-semantics, separating units on the basis of

lexical semantics. The rule could be, for example, that a word divider is inserted only after a morpheme that contributes new lexical semantic information. Example (68) with morphemes divided in this way would become:³¹

(74) *I want an apple and three oranges.*

It can be seen that such an orthography of word division yields similar results to word division based on prosody, as in (70). Note in particular the lack of separation between *an* and *apple* and *and* and *three* in both orthographies. This is because function words are often prosodically deficient (§1.4.2.1); semantic word division will therefore yield results very similar to prosodic word division insofar as function words are not separated by the orthography. The difference, however, is in the extent to which lexical and function words are rigorously distinguished in terms of word division: semantic word division would in principle not separate function words at all. By contrast, prosodic word division would be expected to separate some function words, since cross-linguistically the set of function morphemes does not map exactly on to the set of prosodically deficient morphemes (§1.4.2.1).

However, conformity to a definition of graphematic wordhood implies only compatibility with such a definition, rather than that the constraints of that strategy alone motivate word division where it occurs. For example, the proposal of a graphematic word division strategy for a particular writing system, *e.g.* that a graphematic word must comprise at least two consonants (§1.4.5.4) could be falsified by providing a counter example, namely, a single consonant morpheme demarcated by word dividers. But the lack of such counter examples would not necessarily prove that the unit that is marked out by punctuation or spaces can be accounted for at the graphematic level only. If the orthographic word were in fact a representation of the prosodic word, and the language were subject to a phonological constraint such that a prosodic word must be a full syllable CVC, a phonological explanation would also account for the mandatory graphematic weight of two consonants.

In order to establish the linguistic level of a particular word division strategy, it is important to find minimal pairs, that is pairs of morphemes that differ according to only one feature, for example in terms of their semantics, or prosodic independence, and compare their word division orthographies. For example, let us consider again the morphemes *for* and *before* in the sentence *I went out for tea before dinner* (64).

Since both *for* and *before* are function morphemes they are semantically parallel. Semantic word division would therefore not distinguish *for* from *before*:

³¹ This analysis assumes that in this semantic orthographic of word division, pronouns are regarded as function morphemes, on the grounds of referring to elements already introduced in the (implicit) discourse.

(75) *I went out for tea before dinner.*

Furthermore, *for* and *before* are morphosyntactically parallel, in that they are both prepositions. Accordingly, a morphosyntactic writing system would also not distinguish these two morphemes, as we see in the standard English representation of (64):

(76) *Bill went for tea before dinner.*

For and *before* are not, however, prosodically equivalent. In (64) *for* is not accented and forms a prosodic word with *tea*, while *before* is accented and stands as a prosodic word in its own right. Prosodic word division would, therefore, treat the two differently:

(77) *Bill went for tea before dinner.*

For and *before* are, furthermore, not graphematically equivalent, in that they consist of one and multiple³² graphematic syllables respectively. In a hypothetical writing system where a minimal graphematic word consisted of at least two graphematic syllables, ⟨*for*⟩ would be subminimal, while ⟨*before*⟩ would meet the criterion of minimality for the graphematic word. This would lead to word division identical to that seen in the case of prosodic word division at (77):

(78) *Bill went for tea before dinner.*

In this case, therefore, different minimal pairs need to be found, *e.g.*:

(79) *Bill made toys for children.*

The treatment of such minimal pairs in a given writing system is therefore in principle diagnostic of the word division strategy adopted.

The expected differences in writing system outcomes may be summarised as follows:

- **Prosodic** Word division is a function of the morphemes' prosodic properties: prosodically weak function morphemes are expected to be graphematically dependent, while prosodically strong function morphemes are expected to be graphematically independent. Lexical morphemes, insofar as these are prosodically strong, are expected to be graphematically independent.

³² Depending on one's definition of the graphematic syllable, ⟨*before*⟩ could be argued to have either two or three graphematic syllables.

- **Semantic** Word division is a function of the morphemes' semantic properties, viz. their functional/lexical status: function morphemes are expected to be graphematically dependent, while lexical morphemes are expected to be graphematically dependent.
- **Morphosyntactic** Word division is a function of the morphemes' morphosyntactic freedom: the less free, and the more morphological the morpheme, the more likely it is to be graphematically dependent.
- **Graphematic** Word division is a function of the morphemes' graphematic properties, e.g. graphematic weight: in the West Semitic context graphematically 'heavier' morphemes are expected to be graphematically independent, while graphematically 'lighter' morphemes are expected to be graphematically dependent.

1.7.4. External evidence

In addition to the internal lines of evidence adduced in §1.7.3, the study will also make use of two sources of external evidence:

- Cognate languages, especially Tiberian Hebrew;
- Cognate writing systems, especially that of Ancient Greek.

1.7.4.1. Cognate languages: Tiberian Hebrew

In the case of Northwest Semitic, an important additional line of evidence for prosody comes from the Biblical Hebrew tradition. This tradition provides a remarkably detailed representation of the phonology of the language, including prosodic divisions into prosodic words and prosodic phrases. Furthermore, this prosodic tradition has very ancient roots. This is because a number of recitation traditions of the Biblical text, with very ancient roots, were written down in the middle ages (Khan 2020).

The Tiberian Masoretic tradition is a complex of written and oral traditions, consisting of the following component parts (after Khan 2013):

- Consonantal text of the Hebrew Bible
- Layout and codicological form of the manuscripts
- Paragraph divisions (Hebrew פסקאות *pisqa'ot* or פרשיות *parašiyot*)
- Accent signs (or cantillation)
- Vocalisation
- Marginal notes on the text
- Masoretic treatises
- Orally transmitted reading tradition

For present purposes, I am concerned with the first, fourth and fifth items on this list, namely, the consonantal text, the accent signs (cantillation), and the vocalisation:

- **Consonantal text** The consonant graphemes used to represent the consonants as well as some vowels, in their capacity as *matres lectionis*.

- **Accent signs** The signs used to indicate the musical character of the Biblical text as chanted, as well as to identify the position of word stress (Khan 2001, 80). In addition to these functions, many have taken the system of Tiberian accents to indicate syntactic and/or semantic units in the text (Yeivin 1980, 158–159; Breuer 1982; Aronoff 1985; Khan 2001; cf. summary in Drescher 2013). Janis (1987, 10) and Drescher (1994; 2009; 2013), by contrast, take the primary underlying framework of the accents to be that of the uttered form of the text, that is, prosody, with syntax and semantics being interpreted from that. This view has the advantage that it is able to account for cases where the system of accents does not line up with the syntax or the semantics (Drescher 1994; 2013). As we will see, there are close parallels in this regard between the system of accents and word-level unit demarcation.
- **Vocalisation** Diacritical marks representing vowel phonemes, as well as indicating phonological features of the consonants, *e.g.* spirantisation. It should be said that the accent signs do on occasion disagree with the vocalisation, especially in the matter of pause. For disagreements of this kind see *e.g.* Revell (2015; 2016) and DeCaen & Drescher (2020). See also immediately below.

All three traditions, viz. those of the consonantal text, the vocalisation and the cantillation, may be said to have their origins in antiquity and are interrelated, although each may be said to be at least to some extent independent of the others.³³ This independence is shown by mismatches between the traditions. The independence of the consonantal and vocalisation traditions is indicated explicitly in the Masoretic Text by means of *qere/ketiv* readings throughout the text, although these cases do not exhaust the dissonance. That the vocalisation and cantillation traditions are also independent of one another is shown, *inter alia*, by the fact that pausal forms in the vocalisation tradition do not match the application of disjunctive accents in the cantillation tradition (Revell 2016; DeCaen & Drescher 2020; Khan 2020, 50; DeCaen & Drescher also citing Revell 2015). Explanations for this divergence differ. Revell holds that the two traditions have somewhat independent origins. On the other hand DeCaen & Drescher (2020) argue that the mismatches between vocalisation and cantillation can be accounted for by the fact that the Tiberian system of accents has no way to

³³ This is not, however, to say, that all phenomena associated with prosody are equally ancient. For example, it is not clear for Hebrew exactly when spirantisation of post-vocalic stops, a sandhi phenomenon which takes place within the prosodic phrase, occurred (Kantor 2017, 183). It is likely that, by the time Origen composed the *Secunda* in the 2nd/3rd century CE, at least the labial and dental stops had undergone spirantisation (Kantor 2017, 184). Whether or not the velar stops /k/ and /g/ had also done so by this point is not certain (Kantor 2017, 184). This of course means that in the earliest stages of Hebrew, spirantisation was not a feature of Hebrew phonology. Note, however, that the lack of sandhi phenomena, for example, does not negate the validity of the prosodic word, or the prosodic phrase, as linguistic entities. Rather, it is the prior existence of these prosodic units that later allowed the sandhi phenomena to be realised.

distinguish between two levels of prosodic analysis, namely, the PROSODIC PHRASE and the INTONATIONAL PHRASE. In principle a combination of these factors could be responsible.³⁴ Finally, the system of accents and the consonantal text also diverge where there is a mismatch between the accents and paragraph division (Khan 2020, 50).

That the consonantal tradition of the Masoretic Text represents a recension of the Biblical text of considerable antiquity is demonstrated by the Dead Sea Scrolls (DSS), the text of many of whose manuscripts is extremely close to that of the Masoretic Text, to the extent of being termed ‘proto-Masoretic’ (Khan 2013; Tov 2015, 21). The *terminus ante quem* of the consonantal text of the Masoretic Text can, therefore, be put at the 3rd century BCE (Khan 2013). Furthermore, the same proto-Masoretic manuscripts demonstrate that the orthography of the Masoretic Text had also been fixed by that point (Khan 2013). For our purposes, then, the orthographic word division practices that we see in the consonantal component of the Masoretic Text must have a *terminus ante quem* of the same period.

The two oral traditions also have their origins in antiquity. The antiquity of the vocalisation is indicated, *inter alia*, by the presence in the consonantal text of *qere* readings from elsewhere in the Bible, as at Jos 21 and 1Chr 6 (Khan 2020, 57), where the *qere* of the former can be no later than the composition of the *ketiv* of the latter. The cantillation tradition can be shown to be similarly ancient. In particular, a manuscript of the Greek translation of the Hebrew Bible has been found with divisions matching the distribution of pausal accents in the Tiberian tradition (Revell 1971; Khan 2020, 51; cf. also Revell 1976), showing that ‘the basis of the system of cantillation represented by the later accents was already firmly established in the second century B.C.’ (Revell 1971, 222).

The antiquity of the elements of the Tiberian tradition are important for the present study, since *mutatis mutandis* the conclusions reached on the nature of graphematic word division in Tiberian Hebrew could in principle hold for much earlier written varieties of Northwest Semitic. These possibilities are explored in later chapters.

Two word-level units are distinguished in Tiberian Hebrew (Dresher 1994, esp. 9; 2009; 2013). The first is the GRAPHEMATIC WORD, that is, the unit delimited by space in unpointed manuscripts. The second is the PROSODIC WORD, that is the unit delimited by spaces in pointed manuscripts. Prosodic words may consist of one or more graphematic words linked by *maqfef*. It is generally accepted that the prosodic word is the domain for stress assignment in Tiberian Hebrew, and that a single prosodic word will have a single primary stress (Yeivin 1980, 228; Khan 2020, 509). However, the nature of the graphematic word is less easily defined. Unlike modern European languages, division into graphematic words does not correspond to division into morphosyntactic words, running, in syntactic terms, from ‘a word to a phrase’ (Ravid 2012, 113).

³⁴ I am grateful to an anonymous reviewer for this last point.

1.7.4.2. Cognate writing system: Ancient Greek

The fact that graphematic words in alphabetic Ancient Greek – whose writing system is cognate with those of the Northwest Semitic languages under study here – correspond to prosodic words provides external support for the claim that graphematic words in those Northwest Semitic writing systems also represent prosodic words (Part IV). This evidence is not, however, as strong as the language, writing system and language family internal criteria discussed above (§1.7.3, §1.7.4.1). As will be described in greater detail, the prosody of Ancient Greek was likely somewhat different from that of contemporary Northwest Semitic languages (§13.5.1). Nevertheless, the fact that a contemporary and cognate writing system, such as alphabetic Greek, can demonstrate a prosodic word-based word division strategy, provides at least circumstantial support for the same suggestion for contemporary Northwest Semitic.

1.8. Outline

The goal of the study is to identify the ORL of word division in the Northwest Semitic writing systems of Ugaritic alphabetic cuneiform, alphabetic Phoenician and the consonantal orthography of Tiberian Hebrew. My central claim is that, with one exception, word division in these writing systems marks out a prosodic unit – rather than morphological, syntactic, graphematic or morphosyntactic – and that this unit is more often than not to be identified with the prosodic word. Morphological and syntactic units are understood secondarily via the representation of the phonology.

I do not, however, set out to establish the (prosodic) rules according to which prosodic words are separated from one another. This is, however, the logical next step beyond the present endeavour, one that would entail the in-depth study of the prosodic hierarchy as it is manifest in Northwest Semitic languages in particular contexts.

The study is divided into four parts. Part I addresses word division in a subset of Phoenician inscriptions, where I argue that word division strategies target both the prosodic word, that is, the actual prosodic word in context, and the prosodic phrase. In Part II, I investigate the ORL of word division in Ugaritic texts. There at least two orthographies may be found, targeting the ACTUAL PROSODIC WORD in context, and the morphosyntactic word. In Part III I move on to consider word division in the Tiberian consonantal text, where I propose that the ORL of word division is an abstract prosodic word unit, that I term the MINIMAL PROSODIC WORD. Furthermore I show that this orthography is extant from very early in the history of Northwest Semitic writing systems, that is, from the first half of the 1st millennium BCE.

Finally, Part IV considers the same question in Ancient Greek inscriptions between the 8th century and 5th century BCE. For Greek I propose that the two prosodic word division strategies found in Phoenician and in Hebrew/Moabite respectively may be found, that is, word division targeting actual prosodic words and minimal prosodic words.

The conclusion explores the implications of the study's findings for writing systems research and the study of the relationship between Northwest Semitic writing systems and that of Greek in particular.

1.9. Presentation of texts

1.9.1. *Scripts*

Any discussion of ancient texts requires those texts to be presented in some form. In the case of many, if not most, inscriptions discussed in the present volume, it was not possible to obtain high resolution images of the original inscriptions. Consequently, it is the modern editions of those texts that constitute the sources of those texts (other texts used are given at the appropriate location):

- Ugaritic: KTU³ (Dietrich, Loretz & Sanmartín 2013);
- Tiberian Hebrew: Westminster Leningrad Codex (<https://tanach.us/>);
- Phoenician, Moabite and epigraphic Hebrew: KAI⁵ (Donner & Röllig 2002);
- Greek: see §13.2.

Modern editors use a variety of approaches in presenting ancient texts. KTU³, for example, provides a transcription in Roman characters. By contrast, KAI⁵ provides a transcription into Hebrew square script. In modern editions of Greek inscriptions the text is usually provided in transcription into minuscules, often also including accents plus/minus breathings. Needless to say, in all of these approaches there is a degree of anachronism, which seeks to balance the expectations of readers against faithfulness to the original text.

In the present study, all texts, regardless of the source language and writing system, are presented with a transcription into Roman script, to help with accessibility. In addition, Biblical Hebrew texts are provided in square script; Ugaritic texts are provided in an Ugaritic typeface (Noto Ugaritic);³⁵ Phoenician, Moabite and pre-Masoretic Hebrew are provided in a Phoenician typeface (Noto Phoenician); finally, Greek inscriptional texts are provided in uncials. The few Proto-alphabetic and Linear B texts are not given in a typeface representing the original script. Proto-alphabetic texts are given in two Romanised transcriptions, while Linear B texts are provided with both syllabic and normalised transcriptions. The purpose of presenting the texts in this way is in no way to suggest that the representation of these texts with modern typefaces is an adequate substitute for autopsy of the original materials. Rather, the goal is to help the reader visualise, albeit in a more idealised fashion than would be the case on the original documents, the place and nature of the punctuation. Hence whitespace is only used here if it is used in this way in the original inscription.

³⁵ Noto fonts may be obtained here: <https://www.google.com/get/noto/>.

Representations of texts in non-Roman scripts, as they appear in numbered examples, are accompanied by a directional arrow, viz. → or ←. The purpose of this arrow is to indicate to the modern reader the text's direction of writing *as it is presented in the present volume*. For the most part the direction of the arrow will also correspond to the direction of writing in the original document. This is so for Ugaritic, Hebrew (epigraphic and Tiberian), Phoenician and Moabite. However, in Greek texts the directions do not necessarily correspond in so-called 'boustrophedon' texts, viz. texts where the direction of writing alternates. Such cases are indicated *ad loc*.

1.9.2. Markings in the text

Modern editions of epigraphic material generally indicate whether a text is restored or partially preserved. Once again, these facts are indicated by a variety of means in modern texts. In the present study, irrespective of the means used in the source edition, restored text is enclosed within square brackets [...] in both representations of the text; partially preserved text is indicated in the transcription into Roman characters either by underdotting, in the case of Greek, or by placing a ring above the letter concerned, in the case of Northwest Semitic. Text presumed omitted by the author is given between corner brackets <...>. Readings corrected by an editor are given in parentheses (...).

In Greek inscriptions no accents are original, and so these are not included on majuscule Greek transcriptions. Accents are placed on Roman transcriptions, following the text given, where accents are provided in the edition.

In addition to indicating restored and partially preserved text, KTU³ indicates redundant text and erasures/corrections. As it happens several instances of supposedly redundant text relate to punctuation between items that should not be punctuated. Supposedly redundant punctuation is included without special marking, since the goal of the study is to attempt to explain the text as we have it. Erasures and corrections are also not indicated; the reader is referred to the edition for this information.

1.9.3. Indication of word and line division in glosses

A range of punctuation signs are used to indicate word division in the writing systems under consideration (§1.3.2). In Roman transcription, however, all word division is marked by means of the following subscript symbol: _(ω). Line division is indicated by the following subscript symbol: _(λ).

1.9.4. Glossing

Each text is provided with annotated glosses in English. For the most part, glosses follow the Leipzig Glossing Rules (<https://www.eva.mpg.de/lingua/resources/glossing-rules.php>). Terms not found in the Leipzig Glossing Rules are indicated in the List of Abbreviations. Affixes are joined to their host by a dash, e.g. *h-byt* [the-house]; clitics are joined to their host by an equals sign, e.g. *b=byt* [in=house]; clitics joined by *maqṣef* in Hebrew are joined to their host by the sign =, e.g. *ʿl=byt* [on=house].

PART I

Phoenician inscriptions

Chapter 2

Introduction

2.1. Overview

Part I of this study concerns the use of word dividers in Phoenician inscriptions from the 1st millennium BCE. The goal is to discover what kind of linguistic unit they mark out, that is, to identify the ORL of graphematic word division. I argue that at least two principles of graphematic word division may be identified, targeting the prosodic word and the prosodic phrase respectively. A chapter is devoted to each of these word division strategies, Chapter 3 and Chapter 4 respectively.

Before each of these strategies are presented and analysed, the present chapter provides an introduction to word division in Northwest Semitic inscriptions more generally, starting with a review of the literature of the topic (§2.2), an overview of the corpus to be analysed (§2.3) as well as discussion of their linguistic and sociocultural identity (§2.4). After a description of the situation in proto-alphabetic (§2.5), there follows a discussion of the principles of word division shared between all the inscriptions analysed (§2.6), before a description of the areas in which they diverge (§2.7).

2.2. Literature review

2.2.1. Millard (1970) and Millard (2012b)

In his seminal article, Millard (1970) provides a detailed survey of the evidence for word division in Northwest Semitic inscriptions, concluding with the following summary of the ‘principles of word division’:

Words are separated from each other except for

- (1) single-letter proclitic particles (*e.g.* k, w, b)
- (2) sometimes the *nota accusativi*
- (3) sometimes bound forms (construct+regnant noun, infinitive absolute+regnant verb)
- (4) sometimes the third person plural suffix

Millard’s survey is very helpful in identifying the general tendencies. However, there is, as with the scholarship on Ugaritic, an adumbration of the fact that 2, 3 and 4 are only tendencies through their prefacing with the adverb ‘sometimes’.

Millard (2012b, 25) notes that the Mesha^c stele provides ‘the earliest lengthy example of the script, displaying clearly the practice of regular word division by a point’. There is an implicit contrast here between the ‘regular’ practice of the Mesha^c stele, and the concomitantly ‘irregular’ practice in the Phoenician inscriptions.

Millard (1970) does not directly address the question of the ORL of word division. More recently, however, in Millard (2012b, 25), he states that Hebrew scribes practiced word division ‘normally with a point after each word, except when they were bound together grammatically’. The implication of the use of the term ‘grammatical’ is, at least to the present author, that it is morphosyntactic, rather than prosodic, factors that lead to the obligatory orthographic cliticization of words like $\text{-} \text{ב} \text{-}$ ‘in’, $\text{-} \text{ה} \text{-}$ ‘the’ and $\text{-} \text{ו} \text{-}$ ‘and’, and, furthermore, that orthographic wordhood in Hebrew is a function of morphosyntax, rather than prosody.

2.2.2. *Lehmann (2005) and Lehmann (2016)*

Lehmann (2016, 38*) identifies three word division practices in Northwest Semitic inscriptions (cf. Naveh 1973b, 207):

- Scriptio continua
- Spaced writing
- Graphic separator writing

He is quick to emphasise that exactly how these systems worked and are interrelated is not fully understood (Lehmann 2016, 37–38*). He states (Lehmann 2016, 38*):

[O]ne simple but fundamental question seems to be as yet unanswered: What was the respective functional background of these three practices which, at first glance, seem to be mutually contradictory—or what, if any, was their mutual interrelationship?

The answer comes down to how graphematic wordhood was conceived of by the scribes of the time (Lehmann 2016, 38*). In grappling with this issue, the question of (in)consistency raises its head. Thus he attributes to the apparent (morphosyntactic) inconsistency of the application of word division in Northwest Semitic inscriptions the lack of interest that has tended to be shown to the subject (Lehmann 2016, 39*). Inconsistency in ‘metrics’ and ‘rhythm’ are also used to show that other parameters must be at work (Lehmann 2005, 94).

Despite the uncertainties, Lehmann offers a tentative suggestion as to the domain word division, namely that ‘the signs generally known as word dividers ... often seem to be mere delimitation marks for prosodic breath units or morpho-grammatical and other units’ (Lehmann 2016, 37*). In so doing, Lehmann leaves the door open to both prosodic and morphosyntactic explanations. He concludes, in general terms, that the word divider is a ‘low-level supra-segmental graphic delimitation mark’, proposing the term ‘low-level graphic separation mark’ to describe it (Lehmann 2016, 38*).

One further point should be made in the context of the present study, namely, that for Lehmann, the default situation is for the elements of construct chains to be

univerbated (Lehmann 2005, 86–87). This will become significant later, where we will see that, while this is certainly possible, it is not justifiable to describe it as a default situation (see esp. §3.4.2).

2.2.3. Zernecke (2013)

Zernecke (2013) addresses the question of whether or not the sequence $\text{L}^{\text{A}}\text{I}+\text{L}^{\text{O}}\text{A}$ $b^{\text{t}}\text{=gbl}$ ‘Lady of Byblos’, occurring several times in the inscriptions from Byblos, can be regarded as proper name. One of the criteria used to establish this is the separate vs. non-separate writing of the elements of a construct chain.

For Zernecke, the default situation in Northwest Semitic inscriptions is that construct chain elements are not separated (Zernecke 2013, 237; for a similar view, see Lehmann 2005, 86–87). Zernecke points out (p. 238), however, that in KAI 6 the elements of $\text{mlk} \cdot \text{gbl}$ ‘king of Byblos’ are written separately. Similarly, in KAI 7 Zernecke observes (pp. 238–241) that the elements of the construct chains $\text{mlk} \cdot \text{gbl}$ ‘king of Byblos’ and $\text{ymt} \cdot \text{šp}^{\text{t}}\text{b}^{\text{[c]}}\text{l}^{\text{l}}$ ‘days of Šipitba‘l’ are separated by word dividers.

By contrast, in KAI 6 the sequence $b^{\text{t}}\text{tgbl}$ is written without word divider (Zernecke 2013, 235–238). Similarly, in KAI 7, the elements of $b^{\text{t}}\text{tgbl}$, along with those of the filiations $\text{bn}^{\text{t}}\text{lb}^{\text{[l]}}$ ‘son of ʾEliba‘l’ and $\text{byh}^{\text{t}}\text{mlk}$ ‘son of Yehimilk’ are not separated.

Zernecke argues that the fact that $b^{\text{t}}\text{tgbl}$ is univerbated in these two inscriptions, whilst the (non-filiation) construct chains are not, is evidence in favour of the lexicalisation of $b^{\text{t}}\text{gbl}$. For Zernecke’s argument to work, there must therefore be an equation of graphematic wordhood and lexical wordhood in these inscriptions (cf. Zernecke 2013, 238).

2.2.4. Steiner (2016)

Steiner (2016) addresses the relationship between the degree of phonemic spelling in a Northwest Semitic writing system and *scriptio continua*, arguing that the latter is a consequence of the former. While Steiner contributes many useful insights, and his thesis is very attractive, his analysis is somewhat hampered by his view of word-level prosody as flat (Steiner 2016, 326; with quotation from Christophe, Gout, Peperkamp & Morgan 2003, 585):

Scriptio continua seems odd to us today, but is actually a phonetically accurate representation of the stream of speech. As every phonetician knows, when casual speech is recorded on a spectrogram, it becomes apparent that there is an ‘absence of obvious acoustic markers at word boundaries, such as silent pauses.’ ... Thus, the absence of word dividers in the orthography of our earliest alphabetic texts goes hand in hand with the frequent phonemic representation of sandhi phenomena that ignore word boundaries.

While it may be true that word boundaries do not coincide with silent pauses (although for the observation that this is not true at slow dictation speed, see Devine & Stephens

¹ Donner & Röllig (2002) do not see a word divider here.

1994, 271), it does not follow that prosody is flat. As the overview at §1.4.2 has shown, prosody even at the word level is generally assumed to be hierarchical. Furthermore, Steiner's statement does not take account of the fact that phonology at the phrasal level and higher does involve periods of silent pause (cf. e.g. Devine & Stephens 1994, 271, 432–433; DeCaen 2004; DeCaen & Drescher 2020). It would therefore not necessarily follow that *scriptio continua* is a direct representation of the stream of speech, especially in inscriptions including more than one prosodic phrase.

2.2.5. Summary

As the foregoing survey of word-level unit division in Northwest Semitic inscriptions shows, it can be said that:

- There is some agreement that univibration is expected between the elements of construct chains;
- Beyond one feature, namely, the writing of monoconsonantal prefix particles together with the following morpheme, word division is inconsistently applied;
- The target linguistic level of word division, whether phonological or morphosyntactic, is not established.

Contrary to these areas of tacit agreement, I will argue in this part of the study that word division in these inscriptions is in fact not inconsistent, provided one sees the target level of word division as prosodic rather than morphosyntactic. Seen in this way, *inter alia*, the awkward dissonance between the expectation that construct chains should be univibrated and the fact that they are often not univibrated, falls away. Furthermore, prosody will be seen to provide a framework for understanding why monoconsonantal prefixes should always be written together with the following morpheme, while longer elements show more sporadic behaviour.

2.3. Corpus

Part I of the study considers in detail the word division orthography of a set of Phoenician royal inscriptions from the 1st millennium BCE:

2.3.1. Early 1st-millennium BCE royal inscriptions from Byblos

The inscriptions in this group are ʔAḥirom (KAI 1),² Yeḥimilk (KAI 4), Šipiṭbaʿl (KAI 7), of which the oldest is KAI 1 (Friedrich, Röllig & Amadasi Guzzo 1999, 3).³ The dating of the inscriptions is contentious, with estimates ranging from the 10th to 8th century BCE.⁴ The exact dating of the early Byblos inscriptions does not matter for my

² For the circumstances of the discovery of this inscription, see Millard (1986, 390).

³ The other royal inscriptions from Byblos are also brought into consideration, but since they are either rather fragmentary (KAI 5, 6) or very short (KAI 2, 3, 8), they are less useful for a study of word division practice.

⁴ On dating see Sass (2017), esp. pp. 129ff., who dates the inscriptions to the 9th century BCE. Cf. others,

purposes. What is important is that these are some of the earliest linear inscriptions of length with word division. The language of the inscriptions is the Byblian dialect of Phoenician,⁵ and word division is denoted by means of a point.

2.3.2. A royal inscription from Sam'al (Ya'diya), northern Syria, Kilamuwa (KAI 24)

This inscription is dated to the second half of the 9th century BCE (Donner & Röllig 1968, 30; Brown 2008, 348; Noorlander 2012, 202, 204). The language of the inscription is generally understood to be Phoenician (Brown 2008, 345; Noorlander 2012, 204), despite the presence of at least one feature shared with Aramaic, namely the use of *br* for 'son', lines 1 and 9, instead of *bn*, seen elsewhere in Phoenician (Noorlander 2012, 208). The choice of Phoenician for the language of the inscription is noteworthy, in view of the use of Sam'alian and Aramaic elsewhere in Sam'al, as well as during Kilamuwa's reign (Swiggers 1982; Brown 2008, 345; Noorlander 2012, 204).⁶ Word division is denoted by means of a point.

2.3.3. A mid-1st-millennium BCE inscription from Byblos, Yehawmilk (KAI 10)

This inscription is somewhat later than the others considered here, dated to the Persian period, *i.e.* the 5th or 4th century (Lehmann 2005, 73). The inscription also differs from the other Phoenician inscriptions in that word division is denoted by means of spaces (Lehmann 2005). The inscription is included because the linguistic level targeted by word division is, as I will argue, not the prosodic word, but the prosodic phrase (see Chapter 4).

The inscriptions are chosen for the following reasons:

- They are complete or almost so;
- They are relatively lengthy;
- They all show word division.

All except KAI 24 are from Byblos. KAI 24 is included because, although it is not from Byblos, it is:

- Very lengthy and well preserved;
- From a similar period to KAI 1, 4 and 7, and therefore useful for comparison with them.

Finally, all except KAI 10 are from the early 1st millennium BCE. KAI 10 is included because in terms of word division it presents a different ORL from the other inscriptions, and therefore provides an interesting counterpoint to them.

including Millard (cited by Sass), who still place them in the 10th century. For further information on KAI 4 and on Phoenician inscriptions in general, see Richey (2019).

⁵ For details of the differences between Byblian Phoenician and other varieties of the language, see page references at Friedrich, Röllig & Amadasi Guzzo (1999, 263), as well as Krahmalkov (2001, 8–9). See also Steele (2011, 189–195) for discussion of relevance to Phoenician inscriptions on Cyprus.

⁶ On the language of the inscription, see also Collins (1971).

2.4. Linguistic and sociocultural identity of the inscriptions

As noted in the previous section, the language of the inscriptions investigated here is Phoenician. It should be noted, however, that the designation is, both from a sociocultural and linguistic perspective, problematic.

On the linguistic side, what is referred to by scholars as ‘Phoenician’ comprises in fact a dialect continuum, initially manifest in a number of city states along the Levantine coast and its immediate hinterland, and then later across a number of polities across the Mediterranean basin. Thus the variety peculiar to Byblos was somewhat set apart from those of Tyre and Sidon (cf. Krahmalkov 2001, 8–9). Furthermore, Samʿalian Phoenician has its own peculiarities, notably the use of *br* for ‘son’ instead of *bn* as elsewhere.⁷ Nevertheless, a linguistic unity, *i.e.* bundles of isoglosses, can be said to have held in a region along the Levantine coast (see for further discussion see Röllig 1983, 84–88; Xella 2017, 153–169). This linguistic unity may or may not deserve the term ‘Phoenician’.

The difficulty on the sociocultural side is that our concept of ‘Phoenicia’ is more a product of modern reception than of historical reality. As Lehmann (2020, 72) puts it:

There is not, and never has been, any ethnic, political or ‘national’ entity that understood or labelled itself as ‘Phoenicia’, nor has there ever been any Phoenician people.

Also of relevance to the (linguistic) inquiry of ‘Phoenicianness’ is the so-called ‘Phoenician script’ (on the place of the term ‘script’ within writing systems research, see §1.3.2 above). The term has persisted, and is often cited as the first or a very early instance of a standardised linear segmental script (cf. Rollston 2014; Gnanadesikan 2017, 17). The suitability of the term has, however, been questioned (Lehmann 2020). In particular, the historically central position that has been afforded to the inscriptions of Byblos is not necessarily merited (Lehmann 2020, 72; quoting Millard 2012a, 411). The reason for the prominent position of the Byblian corpus in the present study is simply that they are an early and relatively homogeneous corpus of inscriptions featuring the use of the word divider. At §2.2 we saw that the graphematic means of indicating word division varies across inscriptions. It is perfectly possible, as Lehmann (2016) implies, that particular functions are tied to specific forms of the word divider at a particular place and time. The small size of the dataset explored here does not allow for the exploration of this issue. Rather, we will be concerned simply to identify what functions can be associated with word dividers, whatever form they may take, leaving correlation between form and function to future study.

To sum up, notwithstanding the complexity of the issues associated with linguistic and sociocultural identity in handling the material culture of the late 2nd and early 1st millennia BCE, it suffices for our purposes to observe that varieties of Northwest

⁷ For other inscriptions with linguistic features sitting across the divide between Canaanite and Aramaic, cf. esp. Deir ‘Alla (KAI 312), classified as ‘Aramaic’ in Donner & Röllig (2002); cf. discussion in Beyer (2012, 123–126), Pat-El & Wilson-Wright (2015).

Semitic language are attested in lapidary inscriptions using segmental writing systems in such a way that modes of graphematic word division may be compared. It does not, of course, follow necessarily that the practices observed in this small corpus must necessarily pertain elsewhere. Nevertheless, it is hoped that the analysis of the inscriptions in this part of the study might form a helpful basis of comparison with linear alphabetic inscriptions elsewhere.

2.5. Proto-alphabetic

Although not the subject of detailed analysis in this study, it is helpful for context to glance back to the presumed forebear of the linear alphabetic script(s) with which this chapter is concerned, by addressing word division in the proto-alphabetic inscriptions. Many of these show no word division marking, notably the Serābîṭ el-khâdim corpus (Naveh 1973b, 206, with references). It is, however, attested there. Thus Cross (1984, 71) states:

The vertical stroke used as a division marker in Old Canaanite is also familiar from examples as early as ca. 1500 (the Tell Nagila sherd), the 14th–13th centuries (the St. Louis seal, the Lachish bowl; cf. the Lachish ewer with three vertical dots) and later.

Despite the often short nature of proto-alphabetic inscriptions from the 2nd millennium BCE,⁸ word division practices, where observed, are largely consistent as those observed in the 1st millennium BCE.

Word division can be seen, for example, in the following inscriptions (word dividers are transcribed with a point, regardless of the original form of the word divider):

(80) Tell Nagila sherd (text van den Branden 1966, 135; trans. after van den Branden 1966, 135). Dated to approx. 1500 BCE (Naveh 1973b, 206; Cross 1984, 74); see also Amiran & Eitan (1965, 121–123, with photo), Leibovitch (1965) and van den Branden (1966):

nḥhw · w
nḥ=hw_{<w>} w
 rest=his and
 ‘his rest and’

(81) St. Louis Seal (text after Cross 1984; trans. Hamilton 2002, 38). Dated to the 14th century BCE (Cross 1984, 74); see also Hamilton 2002, 38, with photo; Cross 1984, 71–72, 74, with drawing; Albright 1966, 11:⁹

⁸ On issues of dating these inscriptions, see Haring (2020, 56–58). For the view that proto-alphabetic was already widespread as a writing system before first attestations, see Haring (2020, 57).

⁹ Benjamin Sass regards this inscription as a forgery (Haring 2020).

lbš · | · ʿrqy
 l=bš_{<w>} <λ> ʿrqy_{<w>}
 to=PN Arkite
 '(belonging to) Bš the Arkite'

For present purposes, note especially in (81) the orthographic prefix status of ll 'to'. Note too the double word divider.

(82) Lachish Bowl 1 (text Puech 1986). Dated to the 13th century BCE (Puech 1986, 18); see also Puech (1986, 18–19) (Puech seems to print a final word divider):

bšlšt · ym · yrḥ
 b=šlšt_{<w>} ym_{<w>} yrḥ_{<w>}
 on=third day month
 'on the third of the month [X (Y?)]'

As with (81), note in (82) the orthographic prefix preposition, in this case *b* 'on'. Note also the separation of the construct chain elements *ym* 'day' and *yrḥ* 'month'.

(83) Lachish Ewer (text and trans. per Cross 1967, 16*). Dated to the 13th century BCE (Hestrin 1987, 212); see also Hestrin (1987), Cross (1967, 16*) and Cross (1954):

mtn · šy · l[rb]ty
 mtn_{<w>} šy_{<w>} l=rbt=y_{<w>} ʾlt_{<w>}
 PN offering to=lady=my DN
 'Mattan. An offering to my Lady ʾElat' [*i.e.* offering given by Mattan]

(84) Qubur al-Walaydah Bowl (text Greene 2017). 12th century BCE (Greene 2017, 46); see also Greene (2017), Berlejung (2010) and Cross (1980):

šm[p^c]l · ʾyʾl · š · [n?/š]
 šmp^cl_{<w>} ʾyʾl_{<w>} š_{<w>} n?/š_{<w>}
 PN PN
 'šmp^cl [son of] ʾyʾl ...' (For interpretations of š, see Greene 2017, 46)

As in (82), the two elements of the construct chain in (84) *šm[p^c]l* and *ʾyʾl* are separated by a word divider.

These proto-alphabetic inscriptions are clearly fragmentary and/or very short. Nevertheless, it is possible to deduce the following:

- The orthographic prefixation of monoconsonantal prepositions was already well established;
- Construct chain elements need not be written as one unit.

As we will see shortly, both of these practices are a hallmark of those seen in the material from the 1st millennium BCE.

2.6. Shared characteristics of word division

The inscriptions considered in Part I have in common that monoconsonantal prefix particles are graphematically univerbated with the following morpheme. Thus in the 'Aḥirom inscription (KAI 1), the following items are written together as one word: the relative pronoun 𐤆 <z> ; 𐤁𐤍 <bn> 'son', in synthetic genitive construction with the word that follows; the conjunction 𐤅 <k> 'when, as'; and the preposition 𐤁 'in'.

- (85) 𐤆𐤀𐤇𐤆 *z=p^l* 'which made' (1)
 (86) 𐤁𐤀𐤍𐤅 *k=št-h* 'when he placed him' (1)
 (87) 𐤍𐤆𐤀𐤁 *b=ʿlm¹⁰* 'in eternity' (1)
 (88) 𐤆𐤀𐤍 *w=ʾl* 'and if' (cf. Donner & Röllig 1968, ad loc.) (2)
 (89) 𐤍𐤇𐤆𐤍𐤅 *b=mlkm* 'among kings' (2)
 (90) 𐤍𐤇𐤍𐤍 *w=skn* 'and a commander' (2)
 (91) 𐤍𐤇𐤇𐤍𐤅 *b=sknm* 'among commanders' (2)

Furthermore, monoconsonantal suffix pronouns are univerbated with the preceding morpheme, e.g. 𐤁𐤁𐤀 *'b-h* 'his father' (KAI 1.1) and 𐤍𐤀𐤍𐤍 *šnt-w* 'his years' (KAI 4.5). Otherwise most multiconsonantal morphemes are written as separate graphematic words.

2.7. Divergence in word division practice

Where the inscriptions differ among themselves is in the treatment of multiconsonantal morphemes. These range from not being univerbated at all, to very long chains being together. Thus, for example, the following items are not written together with surrounding words:

- Demonstrative adjective 𐤆𐤆 *zn* 'this'
- Pronoun 𐤀𐤁 *h* 'he'
- Preposition 𐤇𐤆𐤆 *lpn* 'in the presence of'

To the extent that scholars have been concerned with word division in Northwest Semitic inscriptions, attention has tended to focus on the treatment of construct chains, and, to a lesser degree, on the univerbation of a verb with a following morpheme. However, these represent only a portion of sequences that may be

¹⁰ Note the ambiguity in 𐤍𐤆𐤀𐤁 <b^ʿlm> – 'lords' or 'in eternity' – which only arises because the preposition 𐤁 'in' is written together with the word that follows.

univerbated. In the inscriptions considered for this chapter, *i.e.* KAI 1, 4, 7, 10 and 24, the following sequences may be univerbated:

- Noun + Noun in construct (NPs in apposition are also univerbated in KAI 10)
- Noun + Adjective/Adjective + Noun
- Preposition + Adjective/Noun
- Particle + following morpheme (Particle + *x*)
- Verb + following morpheme (Verb + *x*)

On the basis of the extent to which these sequences are univerbated, the following word division orthographies can be identified in these texts:

- A minority (*i.e.* less than half) of potentially univerbated sequences are in fact univerbated;
- A majority (*i.e.* more than a half) of potentially univerbated sequences are in fact univerbated.

I will argue that: a) division in the Byblian inscriptions and Kilamuwa (*i.e.* KAI 1, 4, 7, 24) corresponds to the separation of ACTUAL PROSODIC WORDS (Chapter 3); b) word division in the second group corresponds to the separation of PROSODIC PHRASES (Chapter 4).

Chapter 3

Prosodic words

3.1. Introduction

The goal of this chapter is to establish the ORL of word division in a small set of early 1st millennium BCE Phoenician inscriptions: KAI 1, 4, 7 and 24. I argue that word division in these inscriptions corresponds to the separation of prosodic words, on the following grounds:

- Graphematic weight of unverbated function words (§3.3);
- Morphosyntactic profile of unverbated syntagms (§3.4);
- The domain of sandhi assimilation (§3.5);
- Comparison with the composition and distribution of prosodic words in Tiberian Hebrew (§3.6).

First, however, I provide an overview of the distribution of word division in these inscriptions, by considering the rate at which syntagms that may be unverbated are in fact so written (§3.2).

3.2. Distribution of word division

The hypothesis to be tested is that graphematic unverbation in Phoenician corresponds to prosodic units in Tiberian Hebrew, specifically, that of prosodic words. In order to provide a means of comparing patterns of unverbation both among Phoenician inscriptions, and between Phoenician inscriptions and Tiberian Hebrew, in terms of prosodic unities, it was first necessary to identify syntagms that are typically associated with prosodic wordhood in Tiberian Hebrew.

For these purposes the following syntagms were identified:

- Noun + Noun (in construct phrases)
- Prep + Noun (in prepositional phrases)

In addition, in order to facilitate comparison with prosodic phrasing (see next chapter), the following were also included:

- Noun + Adj. / demonstrative (in adjectival / demonstrative phrases, excluding monoconsonantal demonstratives)
- Adj. + Noun (in adjectival phrases)
- Verb + Noun (in the same clause)

Table 3.1 reveals that in the inscriptions under consideration, less than half of these syntagmatic sequences that are associated with prosodic words or phrases in Tiberian Hebrew are graphematically univerbated. As we will see (§3.6) this compares favourably with the distribution of *maqfef* in Tiberian Hebrew.

Table 3.1: Distribution of word division and univerbation in Phoenician inscriptions (KAI 1, 4, 7, 24)

	KAI 1		KAI 4		KAI 7		KAI 24	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Separated	12	85.71	10	71.43	5	62.5	28	82.35
Univerbated	2	14.29	4	28.57	3	37.5	6	17.65
Total	14		14		8		34	

The table also provides evidence of a considerable range in the proportion of univerbated sequences that are in fact so treated, from 14.3% (KAI 1) to 37.5% (KAI 7). The small numbers of tokens is surely a factor in the degree of variation. What emerges clearly, however, is that in each case the majority of sequences that might in principle be univerbated are in fact written as separate words. Any account of word division in these inscriptions must be able to take account of this fact.

3.3. Graphematic weight of function words

3.3.1. Introduction

At §2.6 I noted that all the inscriptions considered in this part of the study share the property that monoconsonantal prefix clitics and suffix pronouns are univerbated with their respective neighbouring morphemes. This, as we shall see, is a property shared by almost all alphabetic Northwest Semitic writing systems. At §1.7.3.1 I suggested that the treatment of prosodically or graphematically heavier function morphemes is an important diagnostic for the target of word division orthographies. In particular, where we find that light function morphemes are regularly univerbated, but heavier function morphemes experience this more sporadically, we have evidence, on a typological basis at least, for a word division orthography that targets ACTUAL PROSODIC WORDS. It is the goal of the present section to examine the treatment of these heavier function morphemes in the orthography of word division. Two sets of heavy function morphemes are considered: in §3.3.2 I consider multiconsonantal prepositional morphemes before nouns, while in §3.3.3 I consider multiconsonantal particles.

3.3.2. Multiconsonantal prepositions

Biconsonantal prepositions are attested both as independent graphematic words, as well as sequences univerted with the first element of the governed NP. With the preposition *l* ‘over’ there is one instance of univertation:

(92) 𐤋𐤍𐤋𐤐 *l=gb_{<w>}* (KAI⁵ 4.6)

By contrast there are two instances of separation:

(93) 𐤋𐤍𐤋𐤐 *l_{<w>} gb_{<w>}* (KAI⁵ 1.2)

(94) 𐤋𐤍𐤋𐤐 *l_{<w>} gb_{<w>}* (KAI⁵ 7.5)

The two instances of the triconsonantal compound preposition *lpn* ‘before’ are separated from the following morpheme:

(95) 𐤋𐤍𐤋𐤐𐤋𐤍 *lpn_{<w>} ḡb_{<w>}* ‘before Byblos’ (KAI⁵ 1.2)

(96) 𐤋𐤍𐤋𐤐𐤋𐤍𐤋𐤐 *lpn_{<w>} l=gb_{<w>} gb_{<w>}* ‘before the gods of Byblos’ (KAI⁵ 4.7)

In KAI 24 we see the vacillation within the same inscription with the preposition *km* ‘as’:

(97) KAI⁵ 24.6

𐤋𐤍𐤋𐤐𐤍𐤎𐤍𐤎 ←

km=š_{<w>} klt_{<λ>}

like=fire devouring

‘Like fire that devours ...’ (trans. Collins 1971, 185)

(98) KAI⁵ 24.13

𐤍𐤎𐤍𐤎𐤍𐤎𐤍𐤎 ←

km_{<w>} nbš=ytm_{<w>} b=ṁ_{<w>}

like attitude=orphan towards=mother

‘Like the attitude of an orphan towards their mother’ (trans. with reference to Collins 1971, 187 and Donner & Röllig 1968, 31, 34)

3.3.3. Multiconsonantal particles

Multiconsonantal particles may be graphematically univerted with the following morpheme in KAI 24, e.g.:

(99) KAI⁵ 24.14

𐤌𐤔𐤕𐤁𐤌𐤀.𐤓𐤕𐤁𐤀𐤓𐤓𐤓.𐤓𐤕𐤁𐤀𐤓𐤓𐤓 ←

*mškbm*_{<ω>} ʾl=yk**bd**_{<ω>} l=bʿr**rym**_{<ω>}

PN NEG=honour to=PN

‘Let the *Mškbm* not honour the *Bʿrrm*’ (trans. after Collins 1971, 187; see also Donner & Röllig 1968, 31)

(100) KAI⁵ 24.11

𐤒𐤕𐤅.𐤓𐤕.𐤓𐤕.𐤓𐤕.𐤓𐤕 ←

*w=my*_{<ω>} *bl=ħz*_{<ω>} *pn*_{<ω>} ʾlp_{<ω>}

and=who NEG=saw NEG OX

‘And anyone who has not seen an ox’

Again, however, this is not a rule, as the following examples with the negative particle *bl* (cf. KAI24.11Inscr) and the relative particle ʾš, respectively, show:

(101) KAI⁵ 24.12

𐤓𐤕.𐤓𐤕.𐤓𐤕.𐤓𐤕.𐤓𐤕 ←

*w=my*_{<ω>} *bl*_{<ω>} *ħz*_{<ω>} *ktn*_{<ω>}

and=who not saw tunic

‘And anyone who has not seen a tunic’

(102) KAI⁵ 24.15

𐤁𐤓𐤒𐤕.𐤓𐤕.𐤓𐤕.𐤓𐤕.𐤓𐤕 ←

*bʾl*_{<ω>} ʂ**md**_{<ω>} ʾš_{<ω>} l=g**br**_{<λ>}

PN PN REL to=PN

‘*Bʾl ʂmd* who is *Gbr*’s’ (trans. Collins 1971, 187)

Outside of KAI 24, there is only one example of a multiconsonantal particle, 𐤅𐤕 ʾl ‘if’, at KAI 1.2, which is written separately from the following morpheme:

(103) KAI⁵ 1:2

𐤓𐤕.𐤓𐤕.𐤓𐤕.𐤓𐤕.𐤓𐤕 ←

*w=ʾl*_{<ω>} *mlk*_{<ω>} *b=mlkm*_{<ω>}

and=**if** king among=kings

‘And if a king among kings ... [should rise up against Byblos]’ (trans. with ref. to Donner & Röllig 1968, 2)

3.3.4. *Implications for the target of word division in Phoenician*

Per §1.4.2.2, a distinction can be made between clitics that are never stressed and clitics that may or may not be stressed according to the context. Furthermore, at §1.4.2.4 I argued that the crosslinguistic constraint of foot binarity should lead one to expect monomoraic morphemes to belong to the never-stressed category, while morphemes satisfying foot binarity should be capable of carrying primary stress. If such a distribution can be observed in the graphematic distribution of word division, this can be taken as evidence for word division targeting prosodic words. In these terms a prosodic rationale for differential word division of light vs. heavy particles seems plausible: since a multiconsonantal particle must comprise at least two morae, consisting, minimally, of a vowel and a final consonant, it should be capable of carrying a prosodic word's primary stress. By contrast, a monoconsonantal particle is likely to be monomoraic, and so incapable of carrying the primary stress.

3.4. Morphosyntax of unverbated syntagms

3.4.1. *Introduction*

At §3.3 we saw that multiconsonantal function morphemes, viz. prepositions and particles, are only sporadically unverbated with neighbouring morphemes. This is in contrast to their monoconsonantal counterparts (§2.6), which are always unverbated with a neighbouring word. I took the contrast between the possibility of word division after such morphemes, on the one hand, and the obligatory unverbation of monoconsonantal morphemes, on the other, as evidence of a typological basis for word division targeting ACTUAL PROSODIC WORDS (§3.3).

In determining the linguistic target of word division, it is also helpful to consider these data from a morphosyntactic perspective. As I argued at §1.7.3.3, a morphosyntactic system of word division is expected to separate consistently morphemes in accordance with morphosyntactic boundaries, treating morphosyntactically like elements in the same manner, *e.g.* systematically separating all prepositions from the following morphemes.

The fact that in the Phoenician inscriptions considered here monoconsonantal prefix prepositions, along with other monoconsonantal function words, are regularly unverbated with the appropriate neighbouring morpheme, whilst their multiconsonantal counterparts are not, shows that the writing system does not treat all morphosyntactically like morphemes in the same way. In the present section we explore this phenomenon further, by considering the word division orthography of: nouns in construct (§3.4.2); verb-initial syntagms (§3.4.3); NPs in apposition (§3.4.4); and other noun + modifier phrases (§3.4.5).

Finally, I discuss the one instance in the texts under consideration where unverbation occurs at a clause boundary (§3.4.6).

3.4.2. Nouns in construct

In the Byblian inscriptions (KAI 1, 4 and 7) construct chains may be written as single graphematic words:¹

- (104) 𐤋𐤑𐤏𐤕𐤍 *mlk=gb_l*_{<w/λ>} ‘king of Byblos’ (KAI⁵ 1.1, 4.1)
 (105) 𐤁𐤏𐤍𐤅𐤓𐤍𐤏𐤕 *bn=ʾhrm*_{<w>} ‘son of ʾhrm’ (KAI⁵ 1.1)
 (106) 𐤋𐤑𐤏𐤕𐤍𐤏𐤕 *bn=ʾlbʿl*_{<w>} ‘son (of) ʾlbʿl’ (KAI⁵ 7.2)
 (107) 𐤋𐤑𐤏𐤕𐤍𐤏𐤕 *bʿlt=gb_l*_{<λ>} ‘Lady of Byblos’ (KAI⁵ 7.4)
 (108) 𐤋𐤑𐤏𐤕𐤍𐤏𐤕𐤍𐤏𐤕 *ymt=špṭbʿl*_{<w>} ‘days of špṭbʿl’ (KAI⁵ 7.5)²

However, all three inscriptions provide examples where the elements of construct phrases are written separately:

- (109) 𐤏𐤍𐤏𐤕𐤍𐤏𐤕𐤍 *tmʾ mḥnt*_{<w>} ‘camp commander’ (KAI⁵ 1.2)
 (110) 𐤁𐤏𐤍𐤏𐤕𐤍𐤏𐤕 *bʿl šmm*_{<w>} ‘Lord of the Skies’ (KAI⁵ 4.3)
 (111) 𐤏𐤍𐤏𐤕𐤍𐤏𐤕𐤍𐤏𐤕 *ymt yḥmlk*_{<w>} ‘the days of yḥmlk’ (KAI⁵ 4.5)
 (112) 𐤋𐤑𐤏𐤕𐤍𐤏𐤕𐤍 *mlk*_{<w>} *gb_l*_{<w>} (KAI⁵ 7.2)

Finally, there is one example of a construct phrase where part is univertated, but the other is not:

- (113) 𐤋𐤑𐤏𐤕𐤍𐤏𐤕𐤍𐤏𐤕𐤍 *mḥrt*_{<w>} *ʾl=gb_l*_{<λ>} ‘the council of the gods of Byblos’ (KAI⁵ 4.3)

The same vacillation between univertation and the lack thereof can be seen in the much longer inscription KAI 24. Compare the following two spellings of the same phrase *br hy* ‘son of PN’:

- (114) KAI⁵ 24.1

𐤏𐤍𐤏𐤕𐤍𐤏𐤕𐤍𐤏𐤕𐤍 ←

ʾnk_{<w>} klmw_{<w>} br_{<w>} hy<ʾ>_{<λ>}

I PN son PN

‘I am Kilamuwa son of Hy’

- (115) KAI⁵ 24.9

𐤏𐤍𐤏𐤕𐤍𐤏𐤕𐤍𐤏𐤕𐤍 ←

ʾnk_{<w>} klmw_{<w>} br=hy^ʾ_{<w>}

I PN son=PN

‘I am Kilamuwa son of Hy’

¹ Note, in addition, the line break between 𐤋𐤑𐤏//+𐤋𐤑𐤏 *bʿlt//gb_l* ‘Lady of Byblos’ (KAI 4.3–4).

² Zernecke (2013, 239) sees a line divider between *ymt* and *špṭbʿl*.

It is worth pointing out that the inconsistency exhibited here is at odds with various statements in the literature to the opposite effect, *e.g.* Lehmann (2005, 89):³

Internal spacing or dotted division of bound construct + regnant forms or attributive phrases is, as far as we know, unusual in North-Semitic texts as long as they provide a minimum of freedom to skip spaces or dividing dots.

Compare also Friedrich, Röllig & Amadasi Guzzo (1999, §219):⁴

The old inscriptions with word division usually write genitive constructions without division as a single word.

Rather, Millard (1970, 15), summarising the situation in Northwest Semitic generally, is more to the point:

Words are separated from each other except for ... sometimes bound forms (construct + regnant noun...).

3.4.3. Verb-initial syntagms

From KAI 1, 4, 7 and 24, unverbated verbal syntagms only occur in KAI 24, *e.g.*:

(116) KAI⁵ 24.3

𐤋𐤐𐤂.𐤋𐤁𐤏.𐤁𐤌𐤏𐤏𐤏 ←

$kn=bmh_{\langle\omega\rangle}$ $w=bl_{\langle\omega\rangle}$ $p^q_{\langle\omega\rangle}$

was=PN and=not did

'There was *Bmh*, and he did nothing.'

Once again, however, counterexamples may be found within the same inscription. Thus while the verb 𐤏𐤏 *kn* 'be' in (116) is written together with the following morpheme, in the same line the same verb is written separately from the following morpheme:

(117) KAI⁵ 24.3

𐤋𐤐𐤂.𐤋𐤁𐤏.𐤇𐤁.𐤁𐤌𐤏𐤏𐤏.𐤏𐤏 ←

$w=kn_{\langle\omega\rangle}$ $^yb_{\langle\omega\rangle}$ $hy_{\langle\omega\rangle}$ $w=bl_{\langle\omega\rangle}$ $p^q_{\langle\omega\rangle}$

was=PN father[my] PN and=not did

'There was my father *Hy*, and he did nothing.'

³ Cf. Lehmann (2005, 86): 'In many Northwest Semitic inscriptions bound construct+regnant noun forms have no dividing dot or space between their elements.'

⁴ Original: 'Die alten Inschriften mit Worttrennung schreiben Genetivverbindungen gewöhnlich ungetrennt wie ein Wort.'

There is one example of a longer (verb initial) univerbated phrase:

(118) KAI⁵ 24.7

𐤎𐤌𐤕[𐤕]𐤀𐤕𐤌𐤕𐤌𐤀𐤔𐤕𐤏𐤓𐤀𐤕𐤕 ←

w=³dr=^ly=mlk_{<w>} d[n]nym_{<w>}

and=had_power=over=me=king GN

'The king of the *Dnym* had power over me' (trans. after Collins 1971, 185)

It should be pointed out that none of these univerbated verbal syntagms consist of infinitive absolute + regnant verb (Millard 1970, 15, §2.2.1). Furthermore, it parallels what we find in Ugaritic, in syntagms like the following:

(119) KTU³ 1.2:I:24

→ 𐤁 𐤍 𐤕 𐤕 𐤕 𐤕 𐤕 𐤕 𐤕 𐤕 𐤕 𐤕 𐤕 𐤕 𐤕 𐤕 𐤕 𐤕 𐤕 𐤕 𐤕

b=hm_{<w>} yg^r=b^l_{<w>}

on=them reproach.PST=DN

'Ba^l reproached them' (cf. del Olmo Lete & Sanmartín 2015, 287)

The univerbation of verbal syntagms in Northwest Semitic inscriptions is, therefore, a more general phenomenon than has previously been acknowledged.

Despite the widespread morphosyntactic inconsistency evidenced here, KAI 1, 4, 7 and 24 are not wholly inconsistent in morphosyntactic terms. In particular, in two classes of syntagm, namely NPs in apposition and Noun + Modifier/Determiner phrases, the elements are always written separately.

3.4.4. NPs in apposition

In contrast to nouns in construct, NPs in apposition are not univerbated in these inscriptions. Consider the following sequences of NPs in apposition in KAI 1 and KAI 4, where each NP is written as a separate graphematic word:

(120) KAI⁵ 1:1

𐤏𐤕𐤕.𐤎𐤓𐤕𐤕𐤕.𐤕𐤕𐤕𐤕𐤕.𐤕𐤕𐤕𐤕𐤕.𐤕𐤕𐤕𐤕𐤕.𐤕𐤕𐤕𐤕𐤕.𐤕𐤕𐤕𐤕𐤕.𐤕𐤕𐤕𐤕𐤕.𐤕𐤕𐤕𐤕𐤕.𐤕𐤕𐤕𐤕𐤕 ←

[²]tb^l_{<w>} bn=³hrm_{<w>} mlk=gb^l_{<w>} l=³hrm_{<w>}

PN son=PN king=TN for=PN

³b=h_{<w>}

father=his

'²tb^l, son of ³Aḥirom, king of Byblos, for ³Aḥirom his father'

(121) KAI⁵ 4:1

𐤋𐤑𐤓𐤕𐤗𐤓𐤕𐤗𐤓𐤕𐤗𐤓𐤕𐤗𐤓𐤕𐤗𐤓𐤕 ←

bt_{<ω>} z=bny_{<ω>} yḥmlk_{<ω>} mlk=gbl_{<ω>}
 buildings which=built PN king=TN

‘Buildings that Yehimilk king of Byblos built’ (trans. with ref. to Donner & Röllig 1968, 6)

3.4.5. Noun + Modifier Phrases (incl. demonstrative determiners)

Adjective modifiers in Phoenician are placed to the right of the nouns they modify, as expected in a head-initial language, e.g.:

(122) KAI 4:7

𐤗𐤖𐤁𐤓𐤕𐤗𐤓𐤕𐤗𐤓𐤕 ←

l=pn_{<ω>} ᵛl=gbl_{<ω>} qdšm_{<ω>}
 [before [gods=TN holy_{NP-PP}]]

‘before the holy gods of Byblos’ (trans. with ref. to Donner & Röllig 1968, 6)

Adjectives generally agree in definiteness with the head noun, e.g. *h-ᵛlnm h-qdšm* ‘the holy gods’ (KAI 14.9) (Friedrich, Röllig & Amadasi Guzzo 1999, 212 §299). Thus (122) is unusual in this respect (Friedrich, Röllig & Amadasi Guzzo 1999, 212 §299).

In the texts under consideration here, Noun + Adjective modifier phrases occur in KAI 4 and KAI 24; in each case the modifier is written as a separate graphematic word. We have seen this already at (122) immediately above. Compare also:

(123) KAI⁵ 4:6

𐤓𐤕𐤗𐤓𐤕𐤗𐤓𐤕 ←

k=mlk_{<ω>} ṣdq_{<ω>}
 as=king righteous

‘as a just king’ (trans. with ref. to Donner & Röllig 1968, 6)

(124) KAI⁵ 24.5–6

𐤁𐤕𐤗𐤓𐤕𐤗𐤓𐤕𐤗𐤓𐤕 ←

𐤗𐤕

bm=tk_{<ω>} mlkm_{<ω>} ᵛd_{<λ>} rm_{<ω>}
 in=midst kings mighty

‘in the midst of mighty kings’ (trans. with ref. to Collins 1971, 184)

In Phoenician from other sites the demonstrative determiner is 𐤌 z, although other variations are also found (Friedrich, Röllig & Amadasi Guzzo 1999, 67–68 §113a). Old

Byblian, however, has a different set: a masculine form 𐤆𐤍 *zn* and a feminine 𐤆𐤌 *z'* (Friedrich, Röllig & Amadasi Guzzo 1999, 69 §113b). The plural in all cases is 𐤆𐤍𐤌𐤍 *zl* (Friedrich, Röllig & Amadasi Guzzo 1999, 68–69 §113a). In KAI 10 both sets of determiners – *zn~z'* and *z* – are used.

Demonstrative determiners are treated under modifiers in this study. This is, firstly, because demonstrative determiners in Phoenician occur to the right of the determined NP, as do modifiers. In this way they differ from other determiners, including the article 𐤀 *h*, and the quantifier 𐤆𐤍𐤌𐤍 *kl*, which occur to the left of the NP, e.g. *kl*_{<ω>} *mplt*_{<ω>} (4.2), *h-spr* (24.15).

This is not to say that demonstrative determiners behave exactly as adjective modifiers. In Phoenician it is exceedingly rare for a demonstrative determiner to carry an overt marker of definiteness in the form of the definite article: Friedrich, Röllig & Amadasi Guzzo (1999, 213 §300.2) provides only one example, at KAI 40.3 *h-smlm h-ʾl* ‘these pictures’.

In the very early text KAI 1, even the NP is not marked for definiteness:

(125) KAI⁵ 1:2

𐤆𐤍.𐤆𐤍𐤆.𐤆𐤍𐤌𐤍 ←

*w=ygl*_{<ω>} *ʾrn*_{<ω>} *zn*_{<ω>}
and=uncover **tomb** **this**

‘(And if a king among kings ... should rise up ...) and uncover this tomb’ (trans. with ref. to Donner & Röllig 1968, 2)

In texts where the definite article is used, the NP carries the definite article, but not the demonstrative, e.g.:

(126) KAI⁵ 4:2–3

𐤆𐤍𐤌𐤍.𐤆𐤍𐤌𐤍.𐤆𐤍𐤌𐤍.𐤆𐤍𐤌𐤍 ←
𐤆𐤍

*kl*_{<ω>} *mplt*_{<ω>} *h-btm*_{<λ>} *ʾl*_{<ω>}
all **ruins** **the-buildings** **these**

‘All the ruins of these buildings’

(127) KAI⁵ 24.14

𐤆𐤍𐤌𐤍𐤍.𐤆𐤍𐤌𐤍.𐤆𐤍𐤌𐤍.𐤆𐤍𐤌𐤍 ←

*mškbm*_{<ω>} *ʾl=ykbd*_{<ω>} *l=bʿrrm*_{<ω>}
PN NEG=honour to=PN

‘Let the *Mškbm* not honour the *Bʿrrm*’ (trans. after Collins 1971, 187)

(130) KAI⁵ 24.14–15

𐤇𐤃𐤆𐤄. 𐤇𐤃𐤆𐤄𐤆𐤇. 𐤇𐤃𐤆𐤄𐤆𐤇. 𐤇𐤃𐤆𐤄𐤆𐤇. 𐤇𐤃𐤆𐤄𐤆𐤇. 𐤇𐤃𐤆𐤄𐤆𐤇. 𐤇𐤃𐤆𐤄𐤆𐤇 ←
 𐤇𐤃𐤆𐤄𐤆𐤇. 𐤇𐤃𐤆𐤄𐤆𐤇. 𐤇𐤃𐤆𐤄𐤆𐤇. 𐤇𐤃𐤆𐤄𐤆𐤇. 𐤇𐤃𐤆𐤄𐤆𐤇. 𐤇𐤃𐤆𐤄𐤆𐤇. 𐤇𐤃𐤆𐤄𐤆𐤇. 𐤇𐤃𐤆𐤄𐤆𐤇. 𐤇𐤃𐤆𐤄𐤆𐤇. 𐤇𐤃𐤆𐤄𐤆𐤇. 𐤇𐤃𐤆𐤄𐤆𐤇. 𐤇𐤃𐤆𐤄𐤆𐤇. 𐤇𐤃𐤆𐤄𐤆𐤇. 𐤇𐤃𐤆𐤄𐤆𐤇. 𐤇𐤃𐤆𐤄𐤆𐤇. 𐤇𐤃𐤆𐤄𐤆𐤇.

*mškbm*_{<w>} 𐤇𐤃𐤆𐤄𐤆𐤇_{<w>} *l=bʿrrym*_{<w>} *wbʿrr*_{<λ>} *m*_{<w>}
 PN not=honour to=PN and=PN

𐤇𐤃𐤆𐤄𐤆𐤇_{<w>} *l=mškbmwmy*_{<w>} *yšht*_{<w>} *h-spr=z*_{<w>}
 not=honour to=PN=and=who destroy the-inscription=this

*yšht*_{<w>} *rʿš*_{<w>} *bʿl*_{<w>} *šmd*_{<w>} ʿš_{<w>}
 destroy head DN DN who

*l=gbr*_{<λ>}
 to=PN

'Let the *Mškbm* not honour the *Bʿrrm*, and the *Bʿrrm* not honour the *Mškbm*. And whoever destroys this inscription, let *Bʿl Šmd* who is *Gbr*'s destroy his head.' (trans. after Collins 1971, 187)

Collins (1971, 186) verifies this as follows:

mškbm · 𐤇𐤃𐤆𐤄𐤆𐤇 · *l bʿrrm*
w bʿrrm · 𐤇𐤃𐤆𐤄𐤆𐤇 · *l mškbm*
w my · *yšht* · *h spr z* ·
yšht · *rʿš* · *bʿl* · *šmd* · ʿš · *l gbr*

However, the lack of word divider after *l=mškbm* perhaps points to the following versification:

mškbm · 𐤇𐤃𐤆𐤄𐤆𐤇 · *l bʿrrm*
w bʿrrm · 𐤇𐤃𐤆𐤄𐤆𐤇 · *l mškbm w my* ·
yšht · *h spr z* ·
yšht · *rʿš* · *bʿl* · *šmd* · ʿš · *l gbr*

The poetic effect would be clear: by univerbating *l=mškbm* with *w=my*, and thereby keeping *w=my* in the previous colon with *l=mškbm*, the writer is able to begin both elements of the next bicolon with *yšht*.

3.4.7. Implications for the target of word division in Phoenician

Although there appears to be some consistency in the treatment of NPs in apposition and Noun + Modifier/Determiner phrases, the fact that there is inconsistency in a number of syntagm types points to a lack of isomorphy between graphematic word division and morphosyntactic structure. This is consistent with what we saw in the case of prefix particles (§3.3), where multiconsonantal morphemes are optionally univerbated with the following morpheme.

Morphosyntactic inconsistency is of course negative evidence. This is to say that it constitutes evidence as to what word division does not do, viz. identify morphosyntactic units. It does tell us what word division actually does do. Although we have provided some positive evidence on typological grounds for word division corresponding to ACTUAL PROSODIC WORDS (§3.3.4), it is also desirable to have direct evidence for this. For this we look to two further pieces of evidence: sandhi assimilation (§3.5) and morphosyntactic comparison with *maqṣef* phrases in Tiberian Hebrew (§3.6).

3.5. Sandhi assimilation

Sandhi assimilation is associated with construct phrases headed by *bn* ‘son (of)’ (§3.4.2) in Phoenician. When *bn* occurs immediately before ⟨ʔ⟩, it is written with final ⟨n⟩:

(131) 𐤁𐤍𐤅𐤆𐤇𐤈 *bn=ʔhrm* ‘son of ʔhrm’ (KAI⁵ 1.1)

(132) 𐤁𐤍𐤅𐤆𐤇𐤈 *bn=ʔbl* ‘son (of)’ (KAI⁵ 7.2)

By contrast, when written before ⟨y⟩ or ⟨k⟩, the ⟨-n⟩ is not written:

(133) 𐤁𐤍𐤅𐤆𐤇𐤈 *b=yḥ[mlk* ‘son of Yehawmilk’ (KAI⁵ 6.1)

(134) 𐤁𐤍𐤅𐤆𐤇𐤈 *b=klby* ‘son of Klby’ (KAI⁵ 8)

I take these latter examples to be representations of sandhi assimilation of /n/. This is part of a more general phenomenon of the representation of sandhi assimilation in early Northwest Semitic orthography (Steiner 2016, 321–326).⁵ Outside of Old Byblian, Phoenician examples are restricted to inscriptions from Cyprus, where sandhi assimilation occurs in Preposition + Noun sequences, as well as in construct phrases.

Of sandhi assimilation in Preposition + Noun sequences the following is an example (cf. Friedrich, Röllig & Amadasi Guzzo 1999, §251):

(135) KAI⁵ 33.2

𐤁𐤍𐤅𐤆𐤇𐤈𐤉𐤊𐤋𐤌𐤍𐤎𐤏𐤐𐤑𐤒𐤓𐤔𐤕𐤖𐤗𐤘𐤙𐤚𐤛𐤜𐤝𐤞𐤟𐤠𐤡𐤢𐤣𐤤𐤥𐤦𐤧𐤨𐤩𐤪𐤫𐤬𐤭𐤮𐤯𐤰𐤱𐤲𐤳𐤴𐤵𐤶𐤷𐤸𐤹𐤺𐤻𐤼𐤽𐤾𐤿 ←

$[s]mlt_{\langle\omega\rangle}$	$ʔz_{\langle\omega\rangle}$	$ʔš_{\langle\omega\rangle}$	$ytn_{\langle\omega\rangle}$	$w=ytn'_{\langle\omega\rangle}$
image	this	which	gave	and=set_up

$m=nḥšt_{\langle\omega\rangle}$	$yʔš_{\langle\omega\rangle}$
from=bronze	PN

‘This image which the Yʔš gave and set up out of bronze’ (trans. with ref. to Donner & Röllig 1968, 51)

⁵ Cf. the related phenomenon of consonant coalescence (Steiner 2016, 313–321), e.g. 𐤁𐤍𐤅𐤆𐤇𐤈 ⟨mlkty⟩ is equivalent to *mlk=ky* ‘king of Kition’ (KAI 33.2) (cited in Friedrich, Röllig & Amadasi Guzzo 1999, 56, §99a). This, however, is not attested in the Old Byblian corpus or in KAI 24, and is therefore not treated here.

In the next example, sandhi assimilation of /-m/ occurs in construct chain:

- (136) $\omega\tau\tau\text{L}+\tau\tau\tau\text{L}\tau\text{A}\text{K}\text{L}\tau$ $w=l=^{\text{d}}mlkm=ptlmys$ ‘and to the Lord of Kings Ptolemy’
(KAI⁵ 42.2)

In this example $^{\text{d}}mlkm$ is for $^{\text{d}}dn=mlkm$ [lord=kings] (Steiner 2016, 317, citing Harris 1936, 30).

In Tiberian Hebrew sandhi assimilation is much more restricted than sandhi spirantisation, which operates at the level of the prosodic phrase (§1.4.2.2). Here sandhi assimilation is limited to the boundary between the preposition מִן *min* ‘from’ and the following morpheme (Steiner 2016, 323) (see also §11.1.4); it does not occur at the boundary of בן *bn* ‘son (of)’ and the following morpheme. Nevertheless, insofar as מִן *min* when assimilated forms a single prosodic word with the following morpheme, sandhi assimilation there too is restricted to the level of the prosodic word.

In both Phoenician and Tiberian Hebrew, therefore, sandhi assimilation appears to operate at the level of the prosodic word.⁶ Accordingly, the fact that in KAI 6 and 7 we see sandhi assimilation within a univertated construct phrase is consistent with word division targeting prosodic words.

Corroborating evidence for this claim would come from $\langle n \rangle$ occurring at a boundary where we would not expect to find internal sandhi phenomena. Ideally this would be in the same inscription where we also find assimilation, *i.e.* in KAI 6 or 7, but such a sequence is not attested there. We do, however, have such a sequence in KAI 1:

- (137) KAI⁵ 1:1

$\text{L}\text{O}\text{S}+[\text{K}]\cdot\text{L}\text{O}\text{G}\text{L}\cdot\text{Y}\text{A}\text{K} \leftarrow$

$^{\text{r}}n_{\langle\omega\rangle}$ $z=p^{\text{q}}_{\langle\omega\rangle}$ $[\text{?}]tb^{\text{q}}_{\langle\omega\rangle}$
sarcophagus which=made PN

‘Sarcophagus that $^{\text{r}}tb^{\text{q}}$ made’ (trans. with ref. to Donner & Röllig 1968, 2)

It remains to account for the alternation between the assimilated and non-assimilated variants given above. As Friedrich, Röllig & Amadasi Guzzo (1999, 56) note, the assimilation occurs ‘außer wenn der folgende Name mit einem Laryngal beginnt’. This is the same distribution that we saw with מִן *min* in Tiberian Hebrew (§11.1.4). The distribution of the morpheme בן *bn* ~ ב *b* therefore parallels that of Tiberian

⁶ Steiner (2016, 317–318) classes these cases as instances of ‘external sandhi’, reserving the term ‘internal sandhi’ for assimilation/coalescence occurring at the ‘morpheme boundary between stem and affix’ (Steiner 2016, 321). For Steiner, then, internal sandhi is sandhi at a morpheme boundary within morphosyntactic words, including between stem and affix, while external sandhi is sandhi at a morpheme boundary between morphosyntactic words. Under the definitions of Zwicky (1985), therefore, it can be argued that sandhi assimilation is a case of ‘internal sandhi’, while spirantisation is both an internal and an external sandhi process.

Hebrew מִן *min* ~ מִי *mi-* in Tiberian Hebrew (§11.1.4). The lack of assimilation in these contexts could be explained by the weakening of /h/ and /ʔ/ to zero, for which there is evidence even in the inscriptions from Serābît el-khâdim (Steiner 2016, 326–328).

3.6. Comparison of composition and distribution with prosodic words in Tiberian Hebrew

3.6.1. Distribution

External evidence for the semantics of word division in Phoenician comes from distributional comparison with Tiberian Hebrew. In order to assess the suggestion that graphematic words target prosodic words, the distribution of the morphemes making up graphematic words in the inscriptions are directly compared with the distribution of *maqṣef* phrases in Tiberian Hebrew (cf. §1.4.2.6 above). The results of this comparison are given in Table 3.2.⁷

Table 3.2: Comparison of Tiberian Hebrew morphemes joined by *maqṣef* vs. separated by spaces (Gen 14:1–3; 2Kgs 1:3, 8:16–18) with Phoenician morphemes either unverbated vs. separated by word dividers (KAI 1, 4, 7, 24)

	Phoenician		Tiberian Hebrew	
	Freq.	%	Freq.	%
Separated	55	78.57	27	72.97
Unverbated	15	21.43	10	27.03
Total	70		37	

The table indicates that the degree to which possibly unverbated morpheme sequences in the Phoenician inscriptions are in fact unverbated parallels the degree to which morpheme sequences that have the potential to be joined by *maqṣef* in Tiberian Hebrew are in fact joined by *maqṣef*. This parallel distribution speaks in favour of the two units corresponding. This is to say that the distributional evidence supports the view that graphematic words in Phoenician correspond to prosodic words in Tiberian Hebrew.

3.6.2. Syntagms capable of belonging to the same prosodic word

Closer inspection of individual texts supports the observations made on a macro-level. In Tiberian Hebrew it is possible to find, within a similarly small section of text, minimal pairs of each kind that we find in the Phoenician royal inscriptions. These comparisons provide further evidence of the isomorphy between prosodic words in Tiberian Hebrew and graphematic words in Phoenician.

⁷ The source texts for the Hebrew Bible for the quantitative parts of the investigation in Chapter 3 and Chapter 4 were TanakhML (<https://www.tanakhml.org/>) and BHS (<https://www.academic-bible.com/>).

3.6.2.1. Nouns in construct

At §3.4.2 I adduced the following minimal pairs of construct chains both with and without word division intervening between the elements. These are repeated here for convenience:

(138) 𐤋𐤓𐤁𐤀𐤌𐤍 *mlk*_{<ω>} *gbl*_{<ω>} ‘king of Byblos’ (KAI⁵ 7.2)

(139) 𐤋𐤓𐤁𐤀𐤌𐤍𐤁𐤁𐤋𐤍 *mlk=gbl*_{<ω/λ>} ‘king of Byblos’ (KAI⁵ 1.1, 4.1)

Now compare these graphematic word minimal pairs with the following minimal pair of prosodic words in Tiberian Hebrew:

(140) 2Kgs 18:16

וַיִּתְּנֵם לְמֶלֶךְ אַשּׁוּרִי ←

(*w=ytn=m*_ω) (*l=mlk*_ω) (*ʔšwr*_ω)

and=gave=them to=king TN

‘And he gave them to the king of Assyria.’

(141) 2Kgs 18:17

וַיִּשְׁלַח מֶלֶךְ-אַשּׁוּרִי אֶת-תַּרְתָּן ←

(*w=yšlh*_ω) (*mlk=ʔšwr*_ω) (*ʔt=trtn*_ω)

and=sent king=TN OBJ=Tartan

‘And the king of Assyria sent Tartan’ (KJV)

In both Phoenician and Tiberian Hebrew we see the same syntagm, namely, *mlk TN*, with graphematic word and prosodic word divisions intervening, respectively.

3.6.2.2. Verb-initial syntagms

At §3.4.3 we provided the following examples of unverbated and non-unverbated verb-initial syntagms in Phoenician, repeated again for convenience:

(142) KAI⁵ 24.3

𐤋𐤓𐤁𐤀𐤌𐤍 𐤁𐤁𐤋𐤍 𐤁𐤁𐤋𐤍 ←

*kn=bmh*_{<ω>} *w=bī*_{<ω>} *pī*_{<ω>}

was=PN and=not did

‘There was *Bmh*, and he did nothing.’

(143) KAI⁵ 24.3

לֹא-וַיַּעַשׂ-בְּאִתִּי-וְאֵל-אֲבִי ←

w=kn_{<ω>} b_{<ω>} hy_{<ω>} w=bl_{<ω>} p^l_{<ω>}
 was=PN father[my] PN and=not did

‘There was my father Hy, and he did nothing.’

This contrast is also attested at the level of the prosodic word in Tiberian Hebrew, again in relatively close proximity:

(144) 2Kgs 9:18

בָּא-הַמַּלְאָךְ עַד-הֵם ←

(b^ʾ=h-ml^k_ω) (d=hm_ω)

came=the-messenger up_to=them

‘The messenger came to them’ (KJV)

(145) 2Kgs 15:19

בָּא פּוּל מֶלֶךְ-אַשּׁוּר ←

(b^ʾ) (pwl_ω) (mlk=ʾšwr_ω)

came PN king=TN

‘Pul the king of Assyria came’ (KJV)

3.6.2.3. Multiconsonantal prepositions and particles

In Phoenician we saw that it is perfectly possible for multiconsonantal prepositions and particles either to be univervbated with the following morpheme, or to be written independently (§3.3). In Tiberian Hebrew, we find exactly the same phenomenon at the level of the prosodic word.

PREPOSITION + NOUN

Compare (93) and (94) above with:

(146) 2Kgs 18:14

וַיִּשֶׂם מֶלֶךְ-אַשּׁוּר עַל-חֶזְקִיָּה מֶלֶךְ-יְהוּדָה ←

(w=yśm_ω) (mlk=ʾšwr_ω) (ʾl=hzyh_ω) (mlk=yhw_ω)

and=placed king=TN over=PN king=Judah

‘And the king of Assyria appointed unto [*lit.* placed on] Hezekiah, the king of Judah’ (KJV)

(147) 2Kgs 18:13

עָלָה סְנַחֲרִיב מֶלֶךְ-אַשּׁוּר עַל כָּל-עָרֵי יְהוּדָה ←

(ʿlh_ω) (snhryb_ω) (mlk=šwr_ω) (ʿ_ω) (kl=ʿry_ω)
 went_up PN king=TN **against** **all=cities**

(yhwdh_ω)

Judah

‘(Now did) Sennacherib king of Assyria come up **against all the cities** of Judah’
 (after KJV)

PARTICLE + X

Compare (100) and (101) with:

(148) 2Kgs 10:5

לֹא-נַמְלִיךְ אִישׁ ←

(lʾ=nmlyk_ω) (ʿyš_ω)
not=make_king.1PL man

‘We will not make any man king’ (ERV)

(149) 2Kgs 10:4

הִנֵּה נִשְׁנִי הַמְּלָכִים לֹא עָמְדוּ לְפָנָיו ←

(hnh_ω) (šny_ω) (h-mlkym_ω) (lʾ_ω) (ʿmdw_ω)
 behold two the-kings **not** **stand.3PL**

(lpn=yw_ω)

before=him

‘Behold, the two kings could not stand before him’ (RSV)

3.6.2.4. Longer prosodic word/graphematic word sequences

It is even possible to find partial parallels of (118), that is, of correspondence between Tiberian Hebrew prosodic words and Phoenician graphematic words not only in terms of composition and distribution, but also in terms of length. Example (118) is repeated here for convenience:

(150) KAI⁵ 24.7

wzγ[γ]A·ʿLʿzLōʿAʿʿ ←

w=ʿdr=ʿl=y=mlk_(ω) d[n]nym_(ω)
 and=had_power=over=me=king GN

‘The king of the Dnny_m had power over me’ (trans. after Collins 1971, 185)

As such it is not only the case that the elements which are unverbated have a similar distribution, but the number of items that can co-exist in a unverbated syntagm is also parallel, *e.g.*:

(151) 2Sam 15:2

כָּל־הָאִישׁ אֲשֶׁר־יָהָה לְדָוִד ←

(*kl=h-ʿyš_ω*) (*ʿšr=yhyh=lw=ryb_ω*)

every=the-man who=be=to=him=dispute
'any man that had a controversy' (KJV)

Both (151) and (118) feature the sequence:

CONJ + V + PP + NP

The difference between the Phoenician and Hebrew examples is that in (118) the NP consists of a construct phrase of which the second element is not unverbated with the foregoing graphematic word, whereas in (151) the NP element consists of only a single element.

3.6.3. Syntagms split across more than one prosodic word

3.6.3.1. Noun-Adjective syntagms

Thus far only positive correspondences in the distribution and composition of prosodic words in Tiberian Hebrew and graphematic words in Phoenician have been considered. However, it is also possible to find correspondences between word division in Phoenician and Tiberian Hebrew prosodic wordhood in negative terms, that is, in terms of the syntagms that are not unverbated.

At §3.4.5 it was found that modifiers in Phoenician are not unverbated with the NPs they modify, albeit with the caveat that the dataset is small. However, in Tiberian Hebrew, adjective modifiers only rarely belong to the same prosodic word as the NP they modify. A search of Noun-Adjective sequences joined by *maqef* in the Hebrew Bible yielded 286 instances. By contrast, such sequences joined by a conjunctive accent number 1280.⁸

On closer inspection, a significant proportion of the instances of conjunction by *maqef* are in fact construct phrases, that is, where the adjective is substantivised, *e.g.* כָּל־חַי *kl=h_y* 'every living (thing)' (Gen 3:20): syntagms with כָּל *kl* account for 98 instances. By contrast, these account for only two instances of Noun-Adjective syntagms joined by a conjunctive accent. An adjective modifier in Tiberian Hebrew is therefore much more likely to belong to the same prosodic phrase as its head NP than to belong to the same prosodic word. The distribution of word division in

⁸ Search conducted using software written by the author on the basis of morphological analysis in MorphHb (<https://github.com/openscriptures/morphhb/tree/master/wlc>), using the corpus of prose books listed at §7.2.

adjective modifier phrases therefore closely parallels what we see in the Phoenician royal inscriptions.

3.6.3.2. Demonstrative-NP syntagms

Word division of demonstrative determiners in Phoenician also parallels that which we see in Tiberian Hebrew.

Definiteness in Phoenician is not grammaticalised to the extent that it is in Biblical Hebrew, where the demonstrative determiner is marked by the definite article (van der Merwe, Naudé & Kroeze 2017, 301), implying a syntactic status for the definite article as in Modern Hebrew (Danon 1996, 4–5). Thus, while in Phoenician the demonstrative determiner generally carries no overt marker of definiteness (§3.4.5), in Tiberian Hebrew the article appears on both the determined NP and the determiner pronoun, *e.g.*:

(152) Gen 19:14

מִן־הַמִּקְוֹם הַזֶּה ←

mn=h-mqwm h-zh

from=the-place the-this

‘from this place’

Thus demonstrative determiners in and of themselves, in both Phoenician and Hebrew, are syntactically as opposed to, semantically speaking, modifiers.

It is therefore significant that the sequence Noun + *h-zh* joined by *maqef* is found only 12 times in the Hebrew Bible:⁹

(153) Exod 3:21

וְנָתַתִּי אֶת־הַנֶּחֱמָה לְעַם־הַזֶּה ←

w=ntty_ω ’t=hn_ω h-‘m=h-zh_ω

and=give.1SG OBJ=favour the-people=the-this

‘And I will give this people favour’ (KJV)

By contrast, the same sequence, Noun + *h-zh*, joined by a conjunctive accent, is much more common, occurring 436 times, *e.g.* הַיּוֹם הַזֶּה בְּעֵצָה (b=‘sm_ω h-ywm_ω h-zh_{ωφ}) ‘In the selfsame day’ (Gen 7:13, KJV).

Both adjectives and demonstrative determiners show a very low degree of incorporation into the prosodic word of their syntactic heads. This behaviour

⁹ Exod 3:21; Isa 8:11, 9:15, 26:1, 29:14; Jer 16:5; 1Sam 30:8; 1Kgs 12:6; 2Kgs 6:18; 2Chr 1:10; Obad 1:20; Hag 2:14. All but one example involve biconsonantal Nouns. The exception is 1Sam 30:8: הַגִּדְוֹד־הַזֶּה *h-gdwd=h-zh_ω* ‘this band’.

contrasts with the pattern that we see with construct phrases, where univibration is more frequently found. This relative distribution is again consistent with the morphosyntax of word division in KAI 1, 4, 7 and 24.

3.7. Conclusion

This chapter has brought to bear the following lines of evidence concerning the nature of prosodic word division in five early 1st-millennium BCE inscriptions written in Phoenician:

- Graphematic weight of function words participating in univibration;
- Morphosyntax of univibration;
- Coincidence of sandhi assimilation with graphematic words in Phoenician;
- Comparison of the composition and distribution of graphematic words in Phoenician with prosodic words in Tiberian Hebrew.

All four lines of evidence I have argued are either consistent with, or indicate, that graphematic words in these inscriptions correspond to prosodic words in Phoenician.

I noted at §3.2 that any account of word division in the Phoenician inscriptions must take account of the fact that under half of those sequences that are capable of univibration are in fact so written. The fact that in Tiberian Hebrew prosodic words show a similar rate of incorporation of univibratable items is evidence of the reasonability of this proposal.

In one instance univibration was found to lie across the boundary between clauses (§3.4.6), a fact that is not compatible with prosodic wordhood. I accounted for this by observing that the effect of univibrating across the clausal boundary was to facilitate a poetic effect, allowing two consecutive cola to begin with the sequence *yšht*. It seems likely, then, that graphematic words in many cases represent poetic units on the level of prosodic words, but which, in the verse structure, may not be identical to the prosodic words of ‘everyday’ speech.

Chapter 4

Prosodic phrases

4.1. Introduction

4.1.1. Overview

Chapter 3 has argued that the ORL of word division in KAI 1, 4, 7 and 24 is the prosodic word. As a counterpoint to this, the present chapter considers the ORL of word division in the Yehawmilk inscription (KAI 10). I argue that, in contrast to the inscriptions analysed in the previous chapter, the ORL of word division in KAI 10 is the prosodic phrase. I argue this on three grounds:

- The rate of univibration of elements that are univibrated in other Phoenician inscriptions;
- The syntactic status of univibrated syntagms;
- Comparison of the composition and distribution of graphematic words with the composition and distribution of prosodic phrases in Tiberian Hebrew.

In what follows, I first provide an overview of the distribution of word division in KAI 10 at the macro-level (§4.1.2). From this it is apparent that the distribution is markedly different from that seen in the case of prosodic word level word division. The detail of this is demonstrated in syntactic terms (§4.2): while in inscriptions with prosodic word level word division, the boundaries of graphematic words rarely coincide with syntactic boundaries, this is often the case in KAI 10. This is consistent with prosodic phrase level word division, since, from a cross-linguistic perspective, word division based on prosodic phrases should lead to an alignment between prosody and syntax (§1.5.1).

Table 4.1: Distribution of word division and univibration in Phoenician (KAI 10)

	<i>Freq.</i>	<i>%</i>
Separated	8	17.02
Univibrated	39	82.98
Total	47	

4.1.2. Distribution of word division

Of the set of POS sequences in the Phoenician royal inscriptions considered in Chapter 3, more than half were not in fact univibrated. In the Yehawmilk inscription (KAI 10), however, this situation is reversed. See Table 4.1, where it is shown that a full 82.98% of the POS sequences considered are in fact written together (figures based on

transcription in Lehmann 2005, 84). Concomitant with this is the fact that graphematic words are generally much longer in KAI 10 than in the other Phoenician royal inscriptions considered above, *e.g.* (all examples from KAI 10 follow the text of Lehmann 2005):

(154) KAI 10.8

ʔLʔʔḡḡz+zḡLḡḡ+Lḡḡ ʔḡḡ+ ←
*tbrk*_(ω) *bʔlt=gbl=ʔyt=yḡwmlk*_(ω)
 may_she_bless **Lady=Byblos=OBJ=PN**
 ‘May the Lady of Byblos bless Yehawmilk’

The inscription also presents a number of other features of word division which distinguish it from its Phoenician royal colleagues. These are now discussed in turn, by morphosyntactic collocation type.

4.2. Syntax of unverbated syntagms

In contrast to what we saw in the case of prosodic word level word division (Chapter 3), in KAI 10 there is a high degree of isomorphy between graphematic structure and syntax.

4.2.1. Construct phrases

All construct phrases are unverbated in their entirety in KAI 10 (compare Lehmann 2005, 87–90), if line division is not taken as an indication of word division:¹

(155) KAI 10.1

Lḡḡḡḡzḡḡ LḡḡʔLʔ ʔLʔʔḡḡ ʔḡḡ ←
*ʔnk*_(ω) *yḡwmlk*_(ω) *mlk=gbl*_(ω) *bn=yḡrbʔl*_(ω)
 I PN **king=TN** **son=PN**
 ‘I am Yehawmilk, king of Byblos, son of Yḡrbʔl’

(156) KAI 10.1–2

ʔLʔʔʔLʔḡḡḡḡḡḡ ←
 Lḡḡḡ
*bn=bn=ʔʔmlk=mlk*_(λ) *gbl*
 [son=son=PN_{NP}]=[king TN_{NP}]
 ‘grandson of ʔʔmlk, king of Byblos’

¹ That line division is not indicative of graphematic wordhood in this inscription is suggested by three instances in the inscription where a line break intervenes between morphemes: ʔ_(λ) ḡ ‘land’ (10–11), *mz*_(λ) *bḡ* ‘altar’ (11–12) and *ts*_(λ) *r* (13–14).

The universal univerbation of construct phrases in KAI 10, and consequent alignment between graphematic word division and syntax, contrasts strongly with their treatment in the other Phoenician inscriptions, where we saw that construct chains are in many cases not written as single graphematic words (§3.4.2). (For the word division in *pth hṛṣ* (10.4, 5, 12), identified by Lehmann (2005, 87–89) as a construct chain, see further §4.2.3 below.)

4.2.2. Noun + Modifier Phrases (incl. Demonstrative Determiner Phrases)

There are no adjectival modifiers in KAI 10. There are, however, demonstrative determiners aplenty. In Chapter 3 we saw that these are not graphematically univerbated with their phrases unless determined by the monoconsonantal *z*. In KAI 10, by contrast, the picture is more mixed. The monoconsonantal demonstrative *z* is always univerbated with its phrase, just as it is in the other inscriptions, e.g.:²

(157) KAI 10.5

𐤋𐤆𐤏𐤕𐤔𐤕𐤏𐤕 ←

$l=pn=pth-y=z_{(w)}$

[over=face=[opening/inscription-my_{NP}]=this_{DP-PP}]

‘opposite this opening/inscription of mine’ (trans. with ref. to Donner & Röllig 1968, 12–14)

(158) KAI 10.2

𐤋𐤏𐤕𐤔𐤕𐤏𐤕𐤔𐤕 ←

$h-rbt=b^l t=gbl_{(w)}$

the-Great.F.SG=Mistress=TN

‘the Great Lady, Mistress of Byblos’

Of the biconsonantal determiners, *zn* and *z'*, however, there are examples both of univerbation and separation. Thus for *zn* we have the following minimal pair:³

(159) KAI 10.4

𐤕𐤏𐤕𐤔𐤕𐤏𐤕𐤔𐤕 ←

$h-mzbh=nhšt_{(w)}$

[the-[altar_{NP}]=[bronze_{NP}]

‘this bronze altar’

$zn_{(w)}$

this_{DP}]

² Parallel at line 10: $w=l^c n=c^c m=r^c s=z$.

³ Similarly for *z'* compare $w=h^c r^c p^c t_{(w)}$ $z'_{(w)}$ (line 6) with $w=l^c t=c^c r^c p^c t=z'_{(w)}$ (line 12).

(160) KAI 10.12

𐤆𐤋 𐤏𐤓𐤁𐤄 𐤄[+𐤓+𐤋𐤐𐤑] ←

$[w=ʿlt=pt]h_{\langle\omega\rangle}$ $hrš_{\langle\omega\rangle}$ $zn_{\langle\omega\rangle}$
 and=[over=[opening/inscription gold this_{DP-PP}]]
 ‘and over this golden opening/inscription’

Insofar as multiconsonantal determiners may be univerted with their phrases, the orthography of KAI 10 differs in an important way from that of the other Phoenician royal inscriptions.

4.2.3. NPs in apposition

NPs consisting of more than one element in apposition may be written separately. Accordingly, in the first line of the inscription, the construct phrases in apposition are separated by spaces:

(161) KAI 10.1

𐤋𐤐𐤑𐤓𐤄𐤅𐤆𐤇𐤈 𐤉𐤐𐤑𐤒𐤓𐤔𐤕 𐤖𐤗𐤘𐤙𐤚𐤛 ←

$ʾnk_{\langle\omega\rangle}$ $yḥwmlk_{\langle\omega\rangle}$ $mlk=gb_{\langle\omega\rangle}$ $bn=yḥrbʿl_{\langle\omega\rangle}$
 $[I_{NP}]$ $[PN_{NP}]$ $[king=TN_{NP}]$ $[son=PN_{NP}]$
 ‘I am *Yḥwmlk*, king of Byblos, son of *Yḥrbʿl*’

However, it is possible for NPs in apposition to be written without an intervening space. This is most often the case with the sequence $(h-)rbt(-y)=bʿlt=gb_{\langle\omega\rangle}$ ‘(the/my) Lady, Mistress of Byblos’, e.g.:⁴

(162) KAI 10.2

𐤉𐤐𐤑+𐤋𐤐𐤑+𐤓𐤔𐤕 ←

$h-rbt=bʿlt=gb_{\langle\omega\rangle}$
 the-Great.F.SG=Mistress=TN
 ‘the Great Lady, Mistress of Byblos’

A special case of apposition is that where the second element gives the material of which the first element is made (cf. Friedrich, Röllig & Amadasi Guzzo 1999, 309a; Kautsch 1910, §131d). In KAI 10 these may or may not be univerted. Compare:

⁴ Parallels at 10.2, 3, 3–4. An intervening space is also found at 10.7. The analysis follows Lehmann (2005, 96). Compare also the Ugaritic syntagm $rbt_{\langle\omega\rangle} ʾātrt=y_{\langle\omega\rangle}$ ‘Great Lady, *ʾAtrt* of the Sea’ (KTU 1:4:I.21, III:27, IV:40, V:2; 1.6:I:39, 47, 53).

(163) KAI 10.4

𐤆𐤋 + 𐤍𐤁𐤅𐤇𐤁𐤆𐤋𐤏𐤕𐤏𐤕 ←

h-mzbh=nhšt _{<w>} *zn* _{<w>}
 [the-[**altar**_{NP}]=**[bronze**_{NP}] this_{DP}]
 ‘and this bronze altar’

(164) KAI 10.4

𐤁+𐤍𐤏𐤕𐤏𐤕 𐤆𐤋𐤏𐤕𐤏𐤕 ←

w=h-ptḥ _{<w>} *ḥrṣ=zn* _{<w>}
 and=[the-[**opening**_{NP}] **[gold**_{NP}]=this_{DP}]
 ‘and this golden opening’

Compare also 𐤏𐤕𐤏𐤕=*pth* _{<w>}, *ḥrṣ=zn* _{<w>} (10.5) and [*w*𐤏𐤕=*pt*]*ḥ* _{<w>}, *ḥrṣ* _{<w>}, *zn* _{<w>} (10.12) with *w=h*𐤏𐤕=*ḥrṣ* _{<w>} (10.5). Lehmann (2005, 87–89), for reasons that are not fully explained, takes these as nouns in construct, and therefore needs to account for the fact that the noun sequence at (164) is not unverbated. See also §4.3.2.1 below.

4.2.4. Particle-initial sequences

The relative particle 𐤏𐤕 *ṣ* is unverbated with the following morpheme in six out of the seven occurrences (Lehmann 2005, 87), e.g.:⁵

(165) KAI 10.5

𐤆𐤋𐤏𐤕𐤏𐤕 𐤁+𐤍𐤏𐤕𐤏𐤕 𐤏𐤕𐤏𐤕 ←

𐤏𐤕=*l=pth* _{<w>} *ḥrṣ=zn* _{<w>}
 which=[over=[[opening/inscription_{NP}] [gold_{NP}]=this_{DP-PP}]
 ‘which is over this golden opening/inscription’

The one instance where 𐤏𐤕 *ṣ* is not unverbated with the following morpheme is the following:⁶

⁵ *ṣ* also occurs at the boundary between lines 4 and 5.

⁶ Here the text of KAI is given, with the spacing from Lehmann (2005). Lehmann (2005) gives the text as 𐤏𐤕 *bḥ*[·] *hz*, i.e. 𐤏𐤕 *bḥ*[·] *h-z*, glossing *which is-in-this-court*. Lehmann’s reading is of interest, since it would provide an example of determination on the demonstrative pronoun, which is otherwise extremely rare (§3.4.5).

(166) KAI⁵ 10.4

𐤊𐤎 [] 𐤍𐤎 ←

𐤍_{<ω>} b[.]𐤎= z_{<ω>}

REL this

‘which ... this’

4.2.5. Prepositional phrases

In the Phoenician inscriptions considered in the previous chapter, prepositions are occasionally found to be univerbated with the following morpheme, as at (92) and (97). By contrast, prepositions in KAI 10 are in all but one instance univerbated with the following morpheme. In most cases the resulting graphematic words align with the right edges of their respective PPs (text according to Lehmann 2005), e.g.:

(167) 𐤋𐤍𐤋𐤏+𐤍𐤎𐤍𐤎 mmlkt=ʕ=gbl_{<ω>} [majesty=over=Byblos] (KAI 10.2)(168) 𐤋𐤍𐤎_{<λ>} +𐤋𐤏𐤍𐤎+𐤍𐤎𐤎𐤎 l=rbt-y=bʕt_{<λ>} gbl_{<ω>} [to=Great.F-my=mistress_{<λ>} Byblos] (KAI 10.3–4)(169) 𐤍𐤎𐤎𐤎𐤎𐤎 𐤍𐤎=ʕ=hm_{<ω>} [which=over=them] (KAI 10.6)

This extends to PPs headed by compound prepositions, e.g. ʕlpn ‘over the face of’ and btkt ‘in the middle of’:

(170) 𐤋𐤎𐤎+𐤎𐤎𐤎𐤎 ʕ=pn=pth-y=z_{<ω>} [over=face=inscription-my=this] (KAI 10.5)(171) 𐤍𐤎𐤎+𐤍𐤎𐤎𐤎 𐤍𐤎=b=tkt=ʕbn_{<ω>} [which=in=middle=stone] (10.5)

By contrast, where such prepositions occur in KAI 1 and 7, viz. (95) and (96) they are written as separate graphematic words.

There are, however, examples where a space intervenes before the right edge of the PP, e.g.:

(172) 𐤎𐤎𐤎𐤎 𐤎+𐤎𐤎𐤎𐤎 𐤍𐤎=ʕ=pth_{<ω>} hrs=zn_{<ω>} [which=over=inscription/opening_{<ω>} gold=this_{<ω>}] (KAI 10.4) Finally, in one instance, the preposition is followed by a space:(173) [𐤋𐤍]𐤎𐤎𐤎𐤎𐤎 𐤎𐤎+𐤎 ʕpn_{<ω>} kl=ʕln=g[bl]_{<ω>} [before=all=gods=Byblos] (KAI 10.16)

The general picture, therefore, is that phrases headed by prepositions are written as single graphematic words, albeit with important exceptions where either a space is written immediately after the preposition (173) or within the PP (172).

4.2.6. Verb-initial sequences

With only one exception (see further below), verbs are written together with at least one of their core arguments. In some cases the argument(s) in question are

pronominal, and might be expected to be written as a single graphematic word even in the consonantal text of the Hebrew Bible:

(174) KAI 10.2

לגל+לוג+גגג ג+לוגמא ←

š=p^ltn_(ω) h-rbt=b^ll=gbl_(ω)

REL=made.F.SG=me the-Great=Lady=TN

‘I whom the Great Lady of Byblos made’

In other cases, however, this would not be expected, as in the next example where the independent pronoun גגא nk is unverbated with the foregoing verb:

(175) KAI 10.2

גגאאא ←

ז+גא+א

w=qr²=nk_(λ) t=rbty=b^ll=gbl

and=call=I OBJ=great.F=Lady=Byblos

‘And I called on the Great Lady of Byblos’

Once again it is worth highlighting that the unverbated verbal syntagms cited here do not involve infinitive absolute + regnant verb syntagms (cf. §2.2.1, §3.4.3). Univerbation on this scale is parallel to what we have seen elsewhere in Ugaritic and Phoenician (§3.4.3).

What sets KAI 10 apart, however, is that certain unverbated verbal syntagms are very long indeed, and may comprise an entire phrase/clause, e.g.:

(176) KAI 10.7–8

ז+גא+א+אאאאאאא ←

לגל+לוג

km_(ω) š=qr²t=t=rbty_(λ) b^ll=gbl_(ω)

as when=called.1SG=OBJ=great.F lady=Byblos

‘When I called on the Great Lady of Byblos’

(177) KAI 10.8

אאאאאא ←

w=p^ll=y=n^lm_(ω)

and=did=to-me=good

‘and she did good to me’

Note that, as with construct phrases, there is a close alignment between the (right) edge of the verbal phrase, and the right boundary of the graphematic word. In one instance, however, a verbal form is written on its own as an independent graphematic word:

(178) KAI 10.8

תברק לטגבל ית יהו מלך ←

*tbrk*_{<ω>} *b'lt=gbl=ʔyt=yhwmk*_{<ω>}

[*may_she_bless*]_v [*Lady=Byblos*]_{NP}]=[*OBJ=PN* *OBJP=VP*]

'May the Lady of Byblos bless Yehawmilk'

This example is discussed further at §4.3.2.6 below.

4.2.7. Implications for the ORL of word division

The distribution of graphematic word division in KAI 10 has a close mapping to syntax, much closer than that seen in the inscriptions considered in the last chapter. Such a close mapping between graphematic wordhood is what one would expect were graphematic wordhood to target prosodic phrases (§1.5.1), but is far removed from the kind of distribution we obtain where word division corresponds to prosodic words (cf. §3.4). To demonstrate this connection further, in the following section I directly compare the distribution of prosodic phrases in Tiberian Hebrew with that of graphematic words in KAI 10.

4.3. Comparison with prosodic phrases in Tiberian Hebrew

4.3.1. Distribution of graphematic word division compared with prosodic phrases in Tiberian Hebrew

As the foregoing discussion has shown, the distribution of word division is very different from what we see elsewhere in the Phoenician royal inscriptions. It is reasonable to infer, therefore, that word division in KAI 10 targets a different linguistic level from the prosodic words that we saw in Chapter 3. It emerges clearly that word division in KAI 10 has a much greater degree of correlation with syntax than was found to be the case in the other Phoenician royal inscriptions. However, the agreement with syntax is not total.

One possibility is again that word division targets a level of the prosodic hierarchy. Since the level above the prosodic word is the prosodic phrase, it is reasonable to consider whether this is in fact the unit that is targeted by word division in this inscription. In favour of such an alignment is comparison with the distribution of conjunctive phrases, which correspond to prosodic phrases, in Tiberian Hebrew (Table 4.2).

Table 4.2: Comparison of Tiberian Hebrew morphemes joined by conjunctive accents or maqqef vs. disjunctive accents (Gen 14:1-3; 2Kgs 1:3, 8:16-18) with Phoenician morphemes either univerbated or separated by word dividers (KAI 10)

	Tiberian Hebrew		Phoenician	
	Freq.	%	Freq.	%
Separated	7	17.95	8	17.02
Univerbated	32	82.05	39	82.98
Total	39		47	

The results in the table are promising, since the percentage of univerbation of POS sequences considered in Tiberian Hebrew, 82.05%, is comparable to that seen in KAI 10, viz. 82.98%. The results in the table confirm that, at least in terms of scale, an explanation on the level of the prosodic phrase is much more likely than one on the level of the prosodic word.

4.3.2. Graphematic words with close parallels to prosodic phrases in Tiberian Hebrew

4.3.2.1. Construct phrases

When words are separated according to prosodic words, we have seen that there is no particular expectation that construct phrases should be represented as single graphematic words (§1.5.2, §1.5.4). This corresponds with what we saw in the Phoenician inscriptions discussed in the previous chapter (§3.4.2, §3.6.2.1). However, in KAI 10, construct phrases are always written as a single graphematic word (§4.2.1). An advantage of a prosodic phrase level account of word division in KAI 10, therefore, is that it explains why construct phrases would be univerbated: as we saw in Tiberian Hebrew (§1.5.3), construct chains are usually part of the same prosodic phrase.

Construct chains comprising more than two elements, however, may be broken up into one or more prosodic phrases (cf. Yeivin 1980, 174), e.g.:

(179) Num 11:11

הַיָּהוָה אֵלֵינוּ כָּל־הָעָם אֶת־מִשְׁנֵהְנוּ ←
 (ʔt≅msʔ_φ) (kl≅hʿm h-zh_φ)
 OBJ≅[[burden all≅the-people_{NP}] this_{DP}]
 ‘the burden of all this people’

If ʔš=ʔl=pth_{(w)} hrš=zn_{(w)} (10.5) is indeed an example of nouns in construct (cf. §4.2.3) then (179) constitutes a close parallel to KAI 10.5: both examples take the following form (where the bracketed divisions correspond to prosodic phrases in Tiberian Hebrew and graphematic words in KAI 10):

(PTCL + NP) + (NP + ADJ.DEM)

4.3.2.2. Noun + Modifier Phrases (incl. Demonstrative Determiner Phrases)

We saw previously (§3.6.3) that demonstrative determiners are much more likely to form part of the same prosodic phrase as their syntactic heads than to be incorporated into the same prosodic word. This distribution aligned with the distribution of graphematic words in KAI 1, 4, 7 and 24, where modifiers are never incorporated into the same graphematic word as their syntactic heads. By contrast, in KAI 10 modifiers may or may not be incorporated into the same graphematic word as their syntactic heads.

Since by far the most frequent situation is for a modifying demonstrative to occur in the same prosodic phrase as the syntactic head, finding parallels for (159) is not difficult, e.g.:

(180) Deut 32:49

← עֲלֵה אֶל-הַר הָעֵבְרִים הַזֶּה
 (ʔlh_{ωφ}) (ʔl=hr_ω h-ʔbrym_{ωφ} h-zh_{ωφ})
 go.IMP to=mountain the-TN **the-this**
 ‘Go up to this Mount Abarim’ (trans. after KJV)

The Tiberian Hebrew sequence Noun + *h-zh* where each element belongs to a different prosodic phrase is even rarer than the same sequence joined by *maqef*, with a mere 60 instances. Furthermore, in a number of these, the demonstrative does not modify the noun. Nevertheless, it is possible to find parallels to (160), e.g.:

(181) Num 21:25

← וַיִּקַּח יִשְׂרָאֵל אֶת כָּל-הָעָרִים הָאֵלֶּה
 (w=yqh_{ωφ}) (yśrʔl_{ωφ}) (ʔt_ω kl=h-ʔrym_{ωφ}) (h-ʔlh_{ωφ})
 and=took GN OBJ all=the-cities **the-these**
 ‘And Israel took all these cities’ (KJV)

The following example provides a minimal counterpart to (180):

(182) Num 27:12

← עֲלֵה אֶל-הַר הָעֵבְרִים הַזֶּה
 (ʔlh_{ωφ}) (ʔl=hr_ω h-ʔbrym_{ωφ}) (h-zh_{ωφ})
 go.IMP to=mountain the-TN **the-this**
 ‘Go up to this Mount Abarim’ (after KJV)

While rare, therefore, the prosodic phrase level separation of demonstrative determiner from its syntactic head is paralleled in Tiberian Hebrew.

4.3.2.3. NPs in apposition

At (161) we saw that the set of appositional NPs in KAI 10.1 are written separately from one another (§4.2.3). We see the same phenomenon in Tiberian Hebrew prosodic phrases involving lists of names and their titles, *e.g.*:

(183) Gen 14:1

← וַיְהִי בַיּוֹם אֲמֶרְפֶּל מֶלֶךְ-שִׁנְאָר אַרְיוֹךְ מֶלֶךְ אֵלְסָר כְּדָרְלָעוֹמֶר מֶלֶךְ עֵילָם
וַתִּדְעַל מֶלֶךְ גּוֹיִם:

(w=yhy_φ) (b=ymy_φ) (ʔmrpl mlk=šnʿr_φ) (rywk_φ)
and=happened in=days PN king=TN PN

(mlk ʔlsr_φ) (kdrʔmr_φ) (mlk ʔylm_φ)
king TN PN king TN

(w=tdʔ_φ) (mlk gwym_φ)
and=PN king TN

‘And it happened in the days of Amraphel, king of Shinar, Arioch, king of Ellasar, Chedorlaomer, king of Elam, and Tidal, king of nations’ (trans. after KJV)

In (183) all but one of the appositive $mlk=TN_{\omega}$ phrases comprise their own prosodic phrases. The exception is the first of these, namely, $ʔmrpl_{\omega} mlk=šnʿr_{\omega}$, where both the initial personal name and the appositive NP are contained in the same prosodic phrase. This shows that, while there is a tendency for NPs in apposition to belong to separate prosodic phrases, this principle is not always adhered to.

In KAI 10, too, we saw that NPs in apposition are not *always* written separately, as with the sequence $mzbh=nhšt$ (163). With NPs in apposition of this kind – that is, where the second element gives the material out of which the first is made – the elements may either belong to the same prosodic phrase, or to different ones. Compare the phrases ($mnrwt_{\omega} h-zhb_{\omega\phi}$) and ($w=nrt_{\omega} =hm_{\omega\phi}$) ($zhb_{\omega\phi}$) in the following example:

(184) 1Chr 28:15

← וַיִּשָׁקֵל לְמִנְרֹת הַזָּהָב

(w=mšql_{ωφ}) (l=mnrwt_ω h-zhb_{ωφ}) (w=nrt_ω =hm_{ωφ}) (zhb_{ωφ})
and=weight for=lamps the-gold and=light=their gold

‘[he gave] the weight for the candlesticks of gold, and for their lamps of gold’ (after KJV)

4.3.2.4. Particle-initial phrases

In KAI 10 relative pronouns are always written together with the following morpheme. Table 4.3 gives the frequencies for the accentuation of the equivalent Hebrew relative

particle, אֲשֶׁר ^{ʔsr}.⁷ From the table it can be seen that in the vast majority of instances (63% + 27.9% = 90.9%), אֲשֶׁר ^{ʔsr} has either a conjunctive accent or *maqfef*, meaning that it belongs to the same prosodic phrase or prosodic word as the following morpheme. A prosodic phrase level account of word division in KAI 10 would therefore be consistent with the Hebrew data.

Table 4.3: Accentuation of relative particle אֲשֶׁר ^{ʔsr}

Accent	Freq.	%
Conjunctive	2937	63.0
<i>Maqfef</i>	1303	27.9
Disjunctive	424	9.1
Total	4664	

4.3.2.5. Verb-initial phrases

In the prosodic word-level word division we saw in Chapter 3, unverbated verb-initial syntagms are attested, but rarely. In KAI 10, by contrast, we find multi-multiconsonantal-morpheme units containing verbal forms written together as a single graphematic word, as at (177). In many cases these graphematic words correspond to whole VPs.

Univerbation of entire VPs may be found in Tiberian Hebrew prosodic phrases. (177) above can be said to be of the following form:

(185) CONJ=V + PREP + PRON.SUFF + N

The following is a parallel from Tiberian Hebrew:

(186) Exod 25:25

← וְעָשִׂיתָ לּוֹ מִסְגֶּרֶת

(w=‘śyt l-w msgrt_φ)

and=make.2SG for-it rim

‘And you shall make a rim for it’

It should be said that examples of this scale of unit exist at the prosodic word level in Tiberian Hebrew, although much more rarely. I could find no direct parallels of (177), that is, of prosodic word syntagms of the form at (185). However, one example of the following similar sequence was found:

(187) PTCL + V + PREP + PRON.SUFF + N

namely:⁸

⁷ Search conducted using software written by the author on the basis of morphological analysis in MorphHb (<https://github.com/openscriptures/morphhb/tree/master/wlc>), using the corpus of prose books listed at §7.2.

⁸ With the same phrase occurring again two verses later at 2Sam 15:4.

(188) 2Sam 15:2

כָּל־הָאִישׁ אֲשֶׁר־יְהִי־לִוְיָב ←

kl=h-γyš_ω ṣr=yhyh=l-w=ryb_ω

every=the-man who=be=to-him=dispute

'any man that had a controversy' (KJV)

However, the fact that this length of sequence is more commonly attested at the prosodic phrase rather than the prosodic word level in Tiberian Hebrew is suggestive of the former being the more appropriate level of comparison.

Finally, in one instance, *tbrk* at (178), the verb is written separately from the surrounding morphemes. If graphematic word division in KAI 10 corresponds to prosodic phrase division in Tiberian Hebrew, the separate writing of *tbrk* corresponds in Tiberian Hebrew to an initial verb carrying a disjunctive accent. Such instances do occur in Tiberian Hebrew, although they are rare, *e.g.*:⁹

(189) Num 21:25

וַיִּקַּח יִשְׂרָאֵל אֶת כָּל־הָעָרִים הָאֵלֶּה ←

(w=yqh_{ωφ}) (yśr'l_{ωφ}) (?t_ω kl=h-rym_{ωφ}) (h-'lh_{ωφ})

and=took GN OBJ all=the-cities the-these

'And Israel took all these cities' (KJV)

4.3.3. Graphematic words without close parallels to prosodic phrases in Tiberian Hebrew

4.3.3.1. Suffix + Independent pronoun

Although the parallels between prosodic phrases in Tiberian Hebrew and graphematic words in KAI 10 are striking, it is important to note some areas of disagreement. One unverbated sequence that has no parallel in Tiberian Hebrew prosodic phrases is that of a Suffix pronoun + Independent pronoun, where both pronouns agree in person and number:

(190) KAI 10.12

שְׁמִי אֲנִי יְהוֹאָב ←

šm=ank=yħwmlk_(λ)

name[my]=I=PN

'My own name is Yehawmilk' (trans. with ref. to Donner & Röllig 1968, 12, 15)

⁹ Parallels at Num 1:53; Deut 1:41; Judg 11:21; Jer 35:17, 41:10; 1Sam 2:17, 20:32; 2Sam 15:6; 1Kgs 2:23; 2Kgs 18:15; Zech 11:9.

In Tiberian Hebrew sequences of this kind, the independent pronoun constitutes its own prosodic phrase (cited Donner & Röllig 1968, 15):

(191) Num 14:32

וְפַנְיֹתֵיכֶם אֶתְּם יִפְּלוּ ←
 (w=pgry-km_φ) (ʔtm_φ) (yplw_φ)
 and=bodies-your.PL you.PL fall.3PL
 ‘But as for *your* bodies – they will fall’¹⁰

Nor is it possible to find *maqfef* sequences of this kind in Tiberian Hebrew.

A univerbated sequence Suffix pronoun + Independent pronoun is, however, attested in Ugaritic:

(192) KTU³ 1.2:IV:11–12

→ < ʔ ʔ ʔ ʔ → → →
 ʔ ʔ ʔ ʔ < ʔ ʔ
 šm-k=āt_{<λ>} ygrš_{<ω>}
 name-your.SG=you.SG PN
 ‘Your name is Ygrš’

Since the sequence is univerbated in Ugaritic, and if it is right that graphematic word division in Ugaritic corresponds to prosodic words (see Part II), a univerbated sequence of this kind can be argued to be compatible with prosodic phrasehood, since each prosodic phrase must consist of one or more prosodic words. If this is correct, it suggests that prosodic phrasehood in Phoenician, and Ugaritic, has slightly different properties from those seen in Tiberian Hebrew.

4.3.3.2. Bisection of VPs

At (178) we noted one instance where we find a verb written separately from an ensuing univerbated subject-object sequence. This is hard to parallel in Tiberian Hebrew at the prosodic phrase level. Furthermore, it is unexpected on theoretical grounds, since it entails a prosodic separation of the verb from both core arguments, which should not happen if prosodic phrases are aligned with XP^{MAX}.

The following Ugaritic parallel, where a subject-object sequence are written together, but separately from the verb, may be relevant:

¹⁰ I am grateful to an anonymous reviewer for offering this translation.

(193) KTU³ 1.2:IV:11

→ 𐤀𐤌𐤓𐤌𐤓𐤌𐤓 𐤁𐤏𐤌𐤓𐤏𐤌𐤓𐤏𐤌𐤓

ktr=šmdm_{<ω>}ynht_{<ω>}[[DN_{NP}]=[double_mace_{NP}] brought_{VP}]

'DN brought a double mace' (trans. per del Olmo Lete & Sanmartín 2015, 620)

However, if graphematic words in Ugaritic do indeed represent prosodic words, rather than prosodic phrases, the Ugaritic example does not provide much help.

4.3.4. Implications for the ORL of word division

At §4.2 we found that there is a high degree of isomorphy of graphematic and syntactic structure in KAI 10, in contrast to what we have seen elsewhere. In this section I have sought to demonstrate that the particular relationship between syntax and graphematic word division seen in this inscription finds a close parallel in the relationship between syntax and prosodic phrases in Tiberian Hebrew, despite certain areas of disagreement. However, it is worth asking whether word division in this inscription does not in fact target syntax directly, without reference to prosody.

4.4. Syntactic vs. prosodic phrase level analysis

The strongest reason for believing that graphematic word division targets a layer distinct from syntax is the fact that, while there is a strong correlation between syntactic phrasing and graphematic word division, the two are not completely isomorphic. For instance, while most PPs are unverbated in their entirety (§4.2.5), one instance is not, (173), repeated here for convenience:

(194) [𐤀𐤏]𐤏𐤏𐤏𐤏𐤏𐤏 𐤏𐤏𐤏𐤏𐤏𐤏 𐤏𐤏𐤏𐤏𐤏𐤏 𐤏𐤏𐤏𐤏𐤏𐤏 [before=all=gods=Byblos] (10.16)

Similarly, while most instances of the relative particle are unverbated with the following morpheme (§4.3.2.4), one is not:

(195) KAI⁵ 10.4

𐤀𐤏 [𐤏] 𐤏𐤏𐤏 ←

𐤏_{<ω>} b[.]𐤏= z_{<ω>}

REL this

'which ... this'

The fact that these syntactic inconsistencies are paralleled in the relationship between prosodic phrases and syntax in Tiberian Hebrew suggests that the representation of syntax is filtered through the prosodic layer of representation.

4.5. Verse form

Another reason for believing that factors beyond syntax are at play is the fact that univertation occurs across a clause boundary:¹¹

(196) KAI 10.13–15

w + ʿk + ʿ ←

lʾ + ʿlʾ ʿpʿ + lʾ ʾk + ʿkʾlʿ ʿ
ʿk + wʿ

w = ʾm = ts_{<λ>} r_{<ω>} mlʾkt_{<ω>} zʾ_{<ω>}
and = if = you_remove workmanship this

ʿlt = mqm = z = wtgl_{<λ>} mstr-w_{<ω>}
from = place = this = and = uncover cover = its

‘and if you remove this workmanship from this place or open this cover’

Here, however, we have the added difficulty that it is hard to reconcile univertation at clause boundaries with prosodic phrasehood: as we have seen (§1.5.3) clause boundaries in Tiberian Hebrew coincide with disjunctive accentuation, that is, with prosodic phrase boundaries.

We cannot, of course, exclude the possibility that the writer of the inscription intended a space where one is no longer discernible. It is in line 9 where the longest graphematic word occurs. As Lehmann (2005, 94) points out, the letters are written most closely together:

(197) KAI 10.9

ʿ + ʿk + ʿkʾpʿ + ʿlʾ ʿkʾlʾ ʾkʾlʾ ʾkʾlʾ ʾkʾlʾ ʾkʾlʾ ʾkʾlʾ ʾkʾlʾ ʾkʾlʾ ʾkʾlʾ ←

w = tʾrk = ym - w = w = šnt - w = ʿl = gbl = k = mlk = šdq = hʾ = w = ttn

and = lengthen = days - his = and = years - his = over = TN = as = king = righteous = he = and = give

‘and may (the Great Lady of Byblos) lengthen his days and his years over Byblos since he is a righteous king, and may [the Great Lady of Byblos] give ...’

However, univertation at clause boundaries is something we have seen before in Phoenician (§3.4.6), where I argued that the writer sought to generate a poetic effect by univertating across a clause boundary.

Such an explanation is conceivable in KAI 10, although a deeper understanding of Phoenician (and Ugaritic) verse form is needed before this can be substantiated.

¹¹ Parallels may be found at lines 3 and 9.

It may turn out to be the case that other discrepancies from what would be expected from a prosodic phrase level analysis (§4.3.3) can be accounted for in this way.

4.6. Conclusion

In this chapter the principles of word division in KAI 10 have been analysed. The inscription provides a counterpoint to those considered in Chapter 3, in that graphematic word division shows a much closer relationship to syntactic structure than that seen previously. I sought to account for this closer relationship to syntax by arguing that graphematic words are separated on the basis of prosodic phrases, rather than prosodic words. This was for the following reasons:

- Cross-linguistically prosodic phrases show a much greater degree of alignment with syntax than prosodic words (cf. §1.5.1);
- Most of the word division phenomena can be paralleled in prosodic phrases in Tiberian Hebrew; and
- The occasional areas of disagreement with syntax may in many cases also be accounted for with reference to prosodic phrasehood.

This is not to say that there are not still some areas of uncertainty. In particular, it appears that, as with the inscriptions considered in the last chapter, verse form has a role to play over and above that of prosodic phrasehood, a subject that deserves future investigation.

PART II

Ugaritic alphabetic cuneiform

Chapter 5

Introduction

5.1. Overview

Part II of the study concerns the use of the SMALL VERTICAL WEDGE ⟨∪⟩ as a word divider in the orthographies of Ugaritic alphabetic cuneiform. The goal is to establish ORL of the vertical wedge as a word divider¹ in mythological texts, that is, the linguistic level of word-level unit is delimited by the small vertical wedge.² The problem arises because in a minority of cases, the units so delimited can be quite long, up to four morphosyntactic words, *e.g.*:

(198) KTU³ 1.2:IV:12

→ 𐎮𐎵 𐎶𐎵𐎰𐎠𐎮𐎥𐎢𐎵 𐎶𐎵𐎶𐎵𐎠𐎮𐎥𐎢𐎵 𐎶𐎵𐎶𐎵𐎠𐎮𐎥𐎢𐎵 𐎶𐎵𐎶𐎵𐎠𐎮𐎥𐎢𐎵 𐎶𐎵𐎶𐎵𐎠𐎮𐎥𐎢𐎵

*ygrš*_{<ω>} *grš=ym=grš=ym*_{<ω>} *l=ksī-h*_{<λ>}
DN₁.VOC drive_away.PTPL=DN₂=drive_away.IMP=DN₂ from=throne-his
'*Ygrš*, who drives away Yam, drive away Yam from his throne'

This contrasts with the majority of instances, where word division corresponds, *a priori* at least, to what one might expect in Northwest Semitic lapidary inscriptions or the consonantal text of Tiberian Hebrew:

¹ It should be noted that empty space also appears between words (Horwitz 1971, 130), but the function of spacing for word division is not within the present chapter's scope. Also out of scope is the material form of the small vertical wedge, for which the reader is directed to Ellison (2002).

² Although word division constitutes the most frequent use of the small vertical wedge (Mabie 2004, 203), I use the term SMALL VERTICAL WEDGE rather than WORD DIVIDER because of its potential use for a variety of functions (cf. Mabie 2004, 204 n. 4) including both as a column marker and as a clause divider, the latter especially in the text *Yariḥ and Nikkal* (KTU 1.24) (Robertson 1999; Mabie 2004, 203, 205–211).

(199) KTU³ 1.4:VI:49

→ <P> = <C> , 𐤀 𐤃 𐤄 𐤅 𐤆 𐤇 𐤈 𐤉 𐤊 𐤋 𐤌 𐤍 𐤎 𐤏 𐤐 𐤑 𐤒 𐤓 𐤔 𐤕 𐤖 𐤗 𐤘 𐤙 𐤚 𐤛 𐤜 𐤝 𐤞 𐤟 𐤠 𐤡 𐤢 𐤣 𐤤 𐤥 𐤦 𐤧 𐤨 𐤩 𐤪 𐤫 𐤬 𐤭 𐤮 𐤯 𐤰 𐤱 𐤲 𐤳 𐤴 𐤵 𐤶 𐤷 𐤸 𐤹 𐤺 𐤻 𐤼 𐤽 𐤾 𐤿

špq_{<ω>} ilm_{<ω>} ʾlpm_{<ω>} y[n_{<ω>}

he_supplied gods Calves wine

'he supplied the calf-gods with wine' (trans. after del Olmo Lete & Sanmartín 2015, 58)

(200) KTU³ 1.114:11

→ 𐤁 𐤂 𐤃 𐤄 𐤅 𐤆 𐤇 𐤈 𐤉 𐤊 𐤋 𐤌 𐤍 𐤎 𐤏 𐤐 𐤑 𐤒 𐤓 𐤔 𐤕 𐤖 𐤗 𐤘 𐤙 𐤚 𐤛 𐤜 𐤝 𐤞 𐤟 𐤠 𐤡 𐤢 𐤣 𐤤 𐤥 𐤦 𐤧 𐤨 𐤩 𐤪 𐤫 𐤬 𐤭 𐤮 𐤯 𐤰 𐤱 𐤲 𐤳 𐤴 𐤵 𐤶 𐤷 𐤸 𐤹 𐤺 𐤻 𐤼 𐤽 𐤾 𐤿

b=hm_{<ω>} ygʾr_{<ω>} tgr_{<λ>} bt_{<ω>} il_{<ω>}

on=them reproached guardian house DN

'The guardian of the house of ʾEl reproached them.' (for trans. cf. del Olmo Lete & Sanmartín 2015, 335, 889)

The primary goal of this part of the monograph is to account for the writing of these longer units as single graphematic words, and occupies Chapter 6 through Chapter 8. I argue that the use of the word divider is compatible with the demarcation of ACTUAL PROSODIC WORDS, that is, prosodic words as they exist in a particular context.

The topic is treated as follows. Two word division orthographies are identified, which I term the 'Majority' and 'Minority' orthographies respectively. These terms, self-evidently, refer to the degree to which these orthographies are found. Although the 'Majority' orthography is found in literary texts, it is not exclusively found there, and so I avoid the term 'Literary' orthography, or similar. By the same token, the 'Minority' orthography is found in administrative texts and letters. However, it is not the only word division orthography found, there, and so I avoid the term 'Administrative' orthography, or similar.

In what follows, the phenomenon of Ugaritic univerbation is first described and exemplified in Chapter 6, with particular attention paid to the parts of speech (henceforth POS) of the morphemes joined in this way. Ugaritic univerbation and Tiberian Hebrew accentuation are then compared in quantitative terms in Chapter 7, from which analysis it emerges that univerbation has parallels with both prosodic words and prosodic phrases in Tiberian Hebrew, albeit with a closer relationship to the former than to the latter. In Chapter 8 I argue on syntactic grounds that Ugaritic univerbation corresponds to prosodic words rather than prosodic phrases. In Chapter 9 I address the *a priori* anomalous separation of monoconsonantal prefix particles.

Although the non-separation of monoconsonantal prefix particles is one of the hallmarks of Northwest Semitic writing systems, in Chapter 9 I address three contexts where these morphemes are graphematically separated from their neighbours: literary

texts adopting the ‘Majority’ orthography, non-literary texts adopting the ‘Majority’ orthography, and texts adopting the ‘Minority’ orthography. Since this is a systematic feature of the ‘Minority’ orthography, it is to this that greatest attention is paid (§9.4).

Before embarking, however, I devote the rest of the present chapter to covering preliminary issues. At §5.2 I survey the literature on word division in Ugaritic, outlining the specific issues pertaining to the semantics of word division that remain obscure, before summarising the basic patterns of word division that account for the majority of instances of its use (§5.3). The relationship to line division is then addressed briefly (§5.5), as well as the variety of contexts in which the ‘Majority’ orthography is found (§5.6), textual issues (§5.7) and the optional nature of word division (§5.8). Finally, I frame the task of the chapter in terms of the hypothesis for which I set out to provide evidence, namely that graphematic words in this Ugaritic orthography correspond to ACTUAL PROSODIC WORDS.

5.2. Literature review

Word division practices in Ugaritic are seldom addressed directly.³ To my knowledge only two works, both PhD dissertations, have been dedicated to shedding light on the matter: Horwitz (1971) and Robertson (1994). The matter is also given some air time in Tropper (2000; 2012). The following survey focuses on these contributions, with a final section surveying views sporadically expressed elsewhere in the literature.

5.2.1. Horwitz (1971)

In the context of a scholarly consensus that word division in Ugaritic is often haphazard and unpredictable, Horwitz (1971) provides a detailed analysis of the question of the regularity of word division in Ugaritic texts, showing that it is much more predictable than previously assumed. In particular, he demonstrates that the distribution of the word divider is not random, and that it regularly separates word-level units (Horwitz 1971, 69–72). Horwitz shows that irregular cases of word division – that is, in our terms, those not adhering to the basic patterns of word division outlined above – can be grouped into one of four categories (Horwitz 1971, 130):

- The word divider is written within a word;
- The word divider is written at the end of a line;
- A single word spans two lines;
- An expected division of words is not marked at all.

The fourth category comprises the great majority of instances of irregularity (Horwitz 1971, 83, 130).

³ For this sentiment see also Robertson (1999, 92) and Ellison (2002, 398f.). To my knowledge the situation has changed little since Robertson’s and Ellison’s work.

Horwitz (1971, 130–131) sees his greatest contribution as showing that the distribution of word division varies between texts. This is the case both between genres and within genres. Thus, among mythological texts irregularities occur disproportionately in KTU (=CTA) 13 and 24 (Horwitz 1971, 84). However, the texts are irregular in different ways. Thus in KTU 24 line division does not correspond to word division at all, although there are no instances of a word divider occurring at the end of the line, while in KTU 13 there are no examples of multiline words, but several instances of the word divider occurring at the end of the line (Horwitz 1971, 84).⁴ Among non-mythological texts different scribes have different practices, with some much more likely to treat w- as a separate word than mythological ones (Horwitz 1971, 107). This latter phenomenon I address in the next chapter.

In poetic texts, the focus of the present study on Ugaritic, Horwitz (1971, 101) points out that the perception of inconsistency in the use of the word divider is based on the erroneous assumption that ‘the distribution of the small vertical wedge is conditioned solely by its function as a word divider’ and that, therefore, ‘the probability that markers actually divide two words remains constant from one text to another’. However, if one instead entertains the possibility that there might be a correlation between the poetic structure and the use of the word divider, then apparent inconsistencies can instead be revealed as features of that structure (Horwitz 1971, 101–104). Horwitz (1971, 92) gives the following couplet as an example:

(201) KTU 1.17:VI:25 (text following Horwitz 1971, 92)

→ $\text{w-} \text{btlt} \text{nt}$
 $\text{w-} \text{btlt} \text{nt}$

$w=t_{(w)}$ $btlt_{(\lambda)}$ $nt_{(w)}$
 and=answered maiden DN
 ‘and the Maiden ‘Anat answered’

(202) KTU 1.17:VI:26 (text following Horwitz 1971, 92)⁵

→ $\text{irš} \text{l} \text{g} \text{zr}$

$irš=hym_{(w)}$ $l=âqht_{(w)}$ $g\acute{z}r_{(\lambda)}$
 ask.IMP=life PN youth

‘Ask for life, O ‘Aqhat the youth’ (trans. Horwitz 1971, 92)

Horwitz’s claim is that the number of word dividers per *stichos* of a verse is constant (Horwitz 1971, 92; cf. also Mabie 2004, 204–205). Thus the *stichoi* in examples (218) and (219) have four word dividers each, demarcating three graphematic words.

⁴ KTU 19 also has a larger instance of irregularity (Horwitz 1971, 84–85).

⁵ KTU³ places a word divider after *irš*.

If true, this means that the function of the small vertical wedge as a word divider is modulated by its function as a metrical device. We will discuss some issues attendant to this claim at §5.7 and §5.8, as well as provide some additional evidence for the word divider functioning in this way at §8.4.

For the purposes of the present chapter we make two final points of significance for this study. First, Horwitz makes the intriguing observation that the ‘inconsistency’ in the use of the word divider has parallels with *maqṣef* in Tiberian Hebrew, namely, that in both cases the writing of construct pairs may, but need not, include the word divider/*maqṣef* (Horwitz 1971, 66). Regarding other POS sequences, such as Noun + Verb, or Verb + Noun, however, he states that, ‘The reason why these combinations occasionally lack the small vertical wedge eludes us at present.’ Although Horwitz does not develop the point, the similarity with Tiberian Hebrew *maqṣef* is important for the present analysis, and we will return to it at various points as the study progresses.

Horwitz hints at another important relationship too, namely, that between the word divider and word stress (Horwitz 1971, 90) stating:

[T]he role of the small vertical wedge in the metric pattern at Ugarit remains unknown. Perhaps it is a stress marker.

The suggestion of a relationship between the word divider and stress has occasioned perhaps too emphatic a response, given the cautious manner of its presentation (see §5.2.4). Nevertheless, given the centrality of word stress to prosodic wordhood cross-linguistically (§1.4.2.2; cf. Horwitz 1971, 10), the observation is nonetheless crucial for the present study, where I argue that graphematic words in the Ugaritic ‘Majority’ orthography correspond to prosodic words (§5.9).

5.2.2. Robertson (1994)

Robertson (1994) comprises a study of the ‘secondary system’ of signs in three ancient Near Eastern languages: Old Assyrian, Ugaritic and Ancient Egyptian (Robertson 1994, 2). The ‘secondary system’ of signs is the set of signs used to ‘organize’ the ‘primary system’ of signs – that is, those used to convey the ‘words of the language’ – into ‘easily readable sentences’ (Robertson 1994, 1). In terms of Ugaritic, the sign of primary interest to Robertson is the ‘word divider’ (=small vertical wedge) (Robertson 1994, 8–9).

Robertson (1994) makes little use of Horwitz (1971), since she rejects what she understands to be Horwitz’s assumed identification of a minimal marked unit in Ugaritic as a ‘word’ (Robertson 1994, 10–11).

Robertson’s description focuses on a variety of instances where univerbation occurs, including:

- Monoconsonantal preposition/conjunction/relative **𐎠** d- + Noun (pp. 226–229, 231–232)

- Nouns in a construct genitive construction (pp. 229–231)
- Conjunctions (pp. 232–234)
- Prepositions (pp. 235–238)
- Verbs (pp. 238–240)

While Robertson's work is helpful for helping to identify the basic patterns of word division (these are summarised at §5.3 below) she ignores the irregular cases that constitute the prime focus of Horwitz's work. In particular, she does not address the issue of univertation of verbs with surrounding morphemes, other than to acknowledge that the phenomenon occurs (pp. 259–260). The grounds for ignoring 'unusual cases' for the purpose of formulating rules is that they occur in less than 5% of instances (Robertson 1994, 241–243). However, as Robertson herself acknowledges (p. 241), a rule established on the basis of 95% of instances is not necessarily 'complete'. Robertson and Horwitz therefore talk past one another on this question.

Robertson does, however, attempt to address the ORL of word dividers in the languages she considers (pp. 360–365). Although Robertson does not appear to define the term 'word', she implicitly adopts a morphosyntactic definition (Robertson 1994, 360), stating '[Word dividers] do not delineate the boundaries of words, unless one proposes to redefine the term "word" so that what they do mark does fit the definition of "word"'. For Robertson, the purpose of word division in all the orthographies she considers, bar one, is to mark 'a word or the grammatical relationships between words, and the purpose of the system is to differentiate the close grammatical relationship from the more distant one' (Robertson 1994, 361). (The one exception to this is the Ugaritic non-literary orthography, to which I devote a separate chapter, and defer discussion of this claim to there.)

Concluding that the 'word' division that she considers is not in fact, for her, word division at all, it is necessary to establish what kind of unit is in fact demarcated (p. 361). Robertson suggests that this unit is related to verse structure, on the grounds that another system of word division also appears to have been in use, constituting for Robertson a 'true word divider system'. More specifically, Robertson speculates that the unit marked out in the mythological texts 'may well have had a unit based at least in part on a sound length value which may have been related to some aspect of the verse structure'.

5.2.3. Tropper (2000) and Tropper (2012)

Tropper does not to my knowledge devote a specific study to word division, but the matter is addressed in his monumental grammars (Tropper 2000, 68–70; 2012, 68–70).⁶ Intriguingly he does not cite either of the other two studies discussed here, and so does not engage directly the issues arising in them, nor does he tackle directly the

⁶ It is perhaps indicative of the low level of attention that word division has received in Ugaritic scholarship as a whole that, in grammars numbering 1056 and 1068 pages respectively, the matter is only accorded two and a half pages.

question of the ORL of word division. However, he does observe (p. 68:21.411e) that a graphematic word in Ugaritic may consist of ‘two word forms that in terms of content belong closely together and/or form a single accentual unit’,⁷ thus invoking prosody as a relevant governing factor for the operation of word division in Ugaritic texts. Tropper observes, furthermore, that the first word in units written together in this way have a tendency to consist of monosyllabic, *i.e.* biconsonantal, stems. Note that the statement is also to some extent equivocal, in allowing for the possibility that (presumably semantic) ‘Inhalt’ (‘content’) is also relevant. Furthermore, a syntactic/semantic explanation potentially stands behind the statement (Tropper 2012, 69:21.412i) that ‘Word dividers can be lacking in certain lines or word combinations in texts that otherwise have regular dividers’.⁸ However, Tropper does not make clear what exactly constitutes a ‘Wortverbindung’ in this context.

5.2.4. Word division and metre

Two notable studies in the second half of the twentieth century considered the nature of metre in Ugaritic poetry (Margalit 1975; Stuart 1976), both proposing a metrical basis for Ugaritic and/or Hebrew poetry. These stand against other studies arguing that neither Ugaritic nor Hebrew poetry are metrical, at least in the sense in which the term is used for Greek and Latin poetry (Young 1950; Pardee 1981; Wansbrough 1983).

The possibility of a relationship between word division and metre is also suggested, as we have seen, by Horwitz (§5.2.1). Horwitz’s suggestion that the use of the word divider as a metrical device has, however, received short shrift in the literature (Pardee 1981, 123 n. 36; Wansbrough 1983, 222 n. 6). I am not here primarily concerned with the nature of metre in Ugaritic poetry, although this study’s findings should have implications for it. Rather, for present purposes I should note that Horwitz’s suggestion of a relationship between graphematic word separation and word stress has not been accepted.

Pardee (1981, 123 n. 36) objects to this idea ‘because of the number of particles and pronominal suffixes which are marked off by it, and because of the number of independent words which are not marked off by it’. Of the first Pardee gives the example of *w l. ʿpr* ‘and to the dust’ (KTU 1.2:IV:5). Of the second he gives the example of *ktr šmdm* (KTU 1.2:IV:11). I will have occasion to discuss both instances further below (§9.2.3, §6.6). For now, however, it suffices to say that I consider neither a fundamental obstacle to seeing a relationship between the distribution of the small vertical wedge and that of (prosodic) word stress.⁹

⁷ Original: ‘zwei Wortformen, die inhaltlich eng zusammengehören und/oder eine Akzentheit bilden’.

⁸ Original: ‘Worttrenner können in Texten, die sonst regelmäßige Trenner aufweisen, in bestimmten Zeilen bzw. Wortverbindungen fehlen’.

⁹ The univerbation of ‘independent words’ is not in itself problematic, as will become clear. The particular instance cited by Pardee is, however, exceptional in that it involves the bisection of the VP.

5.2.5. *Other views of the use of the word divider*

More recent work on Ugaritic, with the exception of Tropper (2012), has tended to eschew detailed discussion of the word divider, and its use is generally presented as inconsistent. Thus, for example, Sivan (2001, 11) states ‘The Ugaritian scribes were not consistent in dividing words’ (cf. similar remarks in Wansbrough 1983, 222; Huehnergard 2012, 22).¹⁰ Alternatively, word division is treated in passing, e.g. Pardee (2003–2004b, 25):

Historically, there would most often been a vowel at a lexical boundary (i.e., the first word would have ended with a case or mood vowel) and such boundaries are usually indicated graphically by the word-divider.

5.2.6. *Summary*

To summarise, Horwitz (1971) has demonstrated in both general and specific terms that word division is not random, and suggests a relationship between the use of word division and verse structure based on the irregular cases that he discusses. On the other hand Robertson (1994) focuses on the regular cases in the poetic/mythological texts, and suggests that the unit so demarcated is in part related to ‘sound length value’ and/or ‘verse structure’. Tropper (2000) and (2012) frames the matter in similar terms, although leaves the door open to both prosodic and syntactic/semantic explanations. Among the few scholars who have addressed in detail the question of word division in Ugaritic, the following can therefore be said to be a summary of the consensus:

- The use of the small vertical wedge is not random, and in most cases corresponds to the demarcation of word-level units, although other uses are attested;
- It is possible to formulate statements describing an important majority of instances of word division in Ugaritic alphabetic cuneiform;
- There is a relationship between word division and accent/prosody;
- There is a relationship between the lack of word division and the construct chain;
- More generally, there appears to be a relationship between unverbated sequences a syntactic-semantic units.

However, the precise semantics of the small vertical wedge as a word divider, that is, its ORL (Sproat 2000), remain obscure, since none of the three scholars are specific about what kind of word-level unit is demarcated by the small vertical wedge as a word divider, whether a prosodic, morphosyntactic or semantic, or a combination of these. This is despite, in the case of Horwitz, examining the possible answers to this question at a theoretical level in some depth. It is the goal of the present chapter to

¹⁰ Wansbrough (1983, 222): ‘The problem there is the random and hence indeterminate functional load of that device [i.e. the word divider]’; Segert (1984, 78): ‘The one-consonant prepositions b- “in,” l- “to,” and k- “as” are usually written together with the following noun.’ There is no discussion of the distribution of the variants written with or without the word divider.

provide greater clarity on this question, by identifying the linguistic level where such sequences exist as a unit.

5.3. Basic patterns of word division and univerbation

The use of the small vertical wedge as a word divider in the ‘Majority’ orthography does not follow many strict rules. In fact the only hard-and-fast ‘rule’ appears to be that monoconsonantal suffix pronouns and suffix discourse clitics are not separated from the foregoing morpheme. Both rules are exemplified in the following:

(203) KTU³ 1.6:VI:10–11

→ 𐤀𐤃 𐤀𐤃 𐤅𐤃 𐤅𐤃 𐤅𐤃 𐤅𐤃 𐤅𐤃 𐤅𐤃 𐤅𐤃 𐤅𐤃 𐤅𐤃 𐤅𐤃 𐤅𐤃
 𐤁 𐤃 𐤄 𐤅 𐤆 𐤇 𐤈 𐤉 𐤊 𐤋 𐤌 𐤍

p=hn_{<w>} ʾāḥ-y=m_{<w>} ytn_{<w>} bʿl_{<λ>} spu-y_{<w>}
 and=behold brothers-my=PTCL gave DN food-my
 ‘And behold Baʿl gave my brothers as my food’

There are, however, strong tendencies (for exceptions, see below, §5.4).

First, monoconsonantal prefix particles are regularly written together with the following morpheme:

(204) KTU³ 1.2:IV:5

→ 𐤀𐤁 𐤁 𐤃 𐤄 𐤅 𐤆 𐤇 𐤈 𐤉 𐤊 𐤋 𐤌 𐤍 𐤎 𐤏 𐤐 𐤑 𐤒

l=ʾārṣ_{<w>} ypl_{<w>} ʾl-n(-)y_{<w>}
 to=ground fell military_forces-our/my
 ‘our / my forces fell to the ground’ (trans. del Olmo Lete & Sanmartín 2015, 50)

(205) KTU³ 1.5:II:11

→ 𐤁𐤃 𐤁 𐤃 𐤄 𐤅 𐤆 𐤇 𐤈 𐤉 𐤊 𐤋 𐤌 𐤍 𐤎 𐤏 𐤐 𐤑 𐤒

bht_{<w>} l=bn_{<w>} ilm=mt_{<λ>}
 hail INTERJ=son DN=DN

‘Hail, O son of ʾEl, Môt’ (for interpretation cf. Pardee 2003, 266; del Olmo Lete & Sanmartín 2015, 482)

Second, combinations of two monoconsonantal prefixes (usually clausal + prepositional) are written together. In many cases, this combination is itself univerbated with the following morpheme, as with 𐤅 *w-* and 𐤁 *b-* in the next example:

(206) KTU³ 1.2:IV:3

→ 𐤨𐤨 𐤨 𐤨𐤨 𐤨 𐤨𐤨 𐤨 𐤨𐤨 𐤨 𐤨𐤨

*w=b=ym*_{<w>} *mnḥ=l=ʾabd*_{<w>}

and=in=DN calm=not=lack

‘And in Yam calm was not lacking’ (trans. del Olmo Lete & Sanmartín 2015, 7)

Third, morphemes consisting of two or more consonants are usually separated from the surrounding morphemes. This tendency is not in principle affected by whether or not the morpheme in question is in a dependent or appositive relationship with another morpheme in the context. Thus in the following examples we see nouns in apposition (*ilm*_{<w>} *ʾlpm*, cf. Tropper 2012, 828), nouns in construct (*bt*_{<w>} *il*) and nouns dependent on a biconsonantal preposition (*ʿm*_{<w>} *ʾahy*) all written as separate words:

(207) KTU³ 1.4:VI:49

→ 𐤧𐤶𐤺 𐤧𐤶𐤺 𐤧𐤶𐤺 𐤧𐤶𐤺 𐤧𐤶𐤺 𐤧𐤶𐤺 𐤧𐤶𐤺 𐤧𐤶𐤺

*špq*_{<w>} *ilm*_{<w>} *ʾlpm*_{<w>} *y[n]*

he_supplied gods calves wine

‘he supplied the calf-gods with wine’ (trans. after del Olmo Lete & Sanmartín 2015, 58)

(208) KTU³ 1.114:11

→ 𐤨𐤨 𐤨𐤨 𐤨𐤨 𐤨𐤨 𐤨𐤨 𐤨𐤨 𐤨𐤨 𐤨𐤨

𐤨𐤨 𐤨𐤨 𐤨𐤨 𐤨𐤨

*b=hm*_{<w>} *ygʿr*_{<w>} *tgr*_{<w>} *bt*_{<w>} *il*_{<w>}

on=them reproached guardian house DN

‘The guardian of the house of ʾEl reproached them.’ (for trans. cf. del Olmo Lete & Sanmartín 2015, 335, 889)

(209) KTU³ 1.5:I:25

→ 𐤨𐤨 𐤨𐤨 𐤨𐤨 𐤨𐤨 𐤨𐤨 𐤨𐤨 𐤨𐤨 𐤨𐤨

*w=štṭ*_{<w>} (*w=štm*_{<w>}) *ʿm*_{<w>} *ʾah[ly]*_{<w>} *yn*_{<w>}and=[drink [with [brothers-my_{NP}]_{PP}]

‘(invite me both to eat meat with my brothers) and to drink wine with my brothers’ (trans. per del Olmo Lete & Sanmartín 2015, 620)

That morpheme length is a critical factor is shown by the fact that while monoconsonantal prepositions are generally written together with the following morpheme (see (203) above), monoconsonantal prepositions that have been extended

by a suffix particle, such as 𐤓 -m, are generally written as separate words (cf. Horwitz 1971, 4; quoting Gordon’s summary in Gordon 1965). The following example gives the two cases in a minimal pair:

(210) KTU³ 1.14:I:31-32 (Example given at Huehnergard 2012, 87)

→ 𐤁𐤍𐤓 𐤓𐤁𐤕𐤕𐤗 𐤓𐤁𐤕𐤕𐤗 𐤓𐤁𐤕𐤕𐤗 𐤓𐤁𐤕𐤕𐤗 𐤓𐤁𐤕𐤕𐤗 𐤓𐤁𐤕𐤕𐤗
 𐤁𐤍𐤓 𐤓𐤁𐤕𐤕𐤗 𐤓𐤁𐤕𐤕𐤗 𐤓𐤁𐤕𐤕𐤗 𐤓𐤁𐤕𐤕𐤗 𐤓𐤁𐤕𐤕𐤗 𐤓𐤁𐤕𐤕𐤗

b^om_{<ω>} bky-h_{<ω>} w=yšn_{<λ>} b^odm^c-h_{<ω>}
 in weeping-his and=he_slept in=shedding-his

nhmmt_{<ω>}
 deep_sleep

‘in his weeping he fell asleep, in his tear-shedding deep sleep’ (trans. after del Olmo Lete & Sanmartín 2015, 973)

5.4. Exceptions to the basic patterns of word division

As already noted, the basic patterns of word division identified in the previous section are strong tendencies, rather than hard-and-fast rules. Thus, for each tendency, exceptions can be found.

First, although as remarked monoconsonantal prefix particles are regularly univerted with a following morpheme, they may on occasion be separated:

(211) KTU³ 1.1:III:4

→ 𐤓𐤁𐤕𐤕𐤗 𐤓𐤁𐤕𐤕𐤗 𐤓𐤁𐤕𐤕𐤗 𐤓𐤁𐤕𐤕𐤗 𐤓𐤁𐤕𐤕𐤗 𐤓𐤁𐤕𐤕𐤗 𐤓𐤁𐤕𐤕𐤗

w_{<ω>} rgm_{<ω>} l=kt[r_{<ω>}
 and say.IMP to=DN
 ‘And say to Ktr ...’

Furthermore, while it is usually the case that a clitic chain is written together with the following morpheme, this need not be so, and the prefix combination is often treated as a graphematic word in its own right. The following example again gives a minimal pair:

(212) KTU³ 1.14:I:24-25

→ 𐤓𐤁𐤕𐤕𐤗 𐤓𐤁𐤕𐤕𐤗 𐤓𐤁𐤕𐤕𐤗 𐤓𐤁𐤕𐤕𐤗 𐤓𐤁𐤕𐤕𐤗 𐤓𐤁𐤕𐤕𐤗 𐤓𐤁𐤕𐤕𐤗
 𐤓𐤁𐤕𐤕𐤗 𐤓𐤁𐤕𐤕𐤗 𐤓𐤁𐤕𐤕𐤗 𐤓𐤁𐤕𐤕𐤗 𐤓𐤁𐤕𐤕𐤗 𐤓𐤁𐤕𐤕𐤗 𐤓𐤁𐤕𐤕𐤗

w=b=kī-ḥn_{<ω>} šph_{<ω>} yitbd_{<λ>} w=b_{<ω>}
 and=in=entirety-their family perished and=in

*phyr-h*_{<ω>} *yrt*_{<λ>}

totality-its succession

‘in their entirety the family perished, and in its totality the succession’ (trans. del Olmo Lete & Sanmartín 2015, 659)

Second, although morphemes consisting of two or more consonants are usually graphematically separated from following morphemes, they may on occasion be written together with them. We have seen examples of this at *ilm=mt* at (205) and *mnḥ=l=ābd* at (206).

Since the separate writing of monoconsonantal prefixes is a feature of the ‘Minority’ orthography, discussion of this is deferred until Chapter 9. The present chapter is therefore concerned with providing a framework for understanding the third phenomenon, viz. graphematic words spanning, in Tiberian Hebrew terms, multiple minimal prosodic words.

5.5. Line division

The small vertical wedge is generally not found at line ends, although there are exceptional instances of this, especially in KTU 1.13 and 1.19 (Horwitz 1971; Tropper 2012, 69). Since these texts are not considered in the present study, for our purposes line division is taken to entail graphematic word division. This is in fact not always the case, since occasionally words are spread across lines, e.g.:

(213) KTU³ 1.1:II:19–20

→ <P – M < E
[E – T , M M M M T ,

*št=b=ʿp*_{<λ>} [*rm*_{<ω>} *ddym*_{<ω>}

put=in=steppe harmony

‘Put harmony in the steppe’ (trans. del Olmo Lete & Sanmartín 2015, 171)

However, since in the vast majority of instances line division corresponds to the use of the small vertical wedge as a word divider elsewhere, the writers did not feel it necessary to specify the matter by explicit use of the small vertical wedge at line ends (for the adoption of this approach see also Horwitz 1971, 30; Robertson 1999, 93–94).

5.6. Contexts of use

What I term here the ‘Majority’ orthography is attested in a wide variety of contexts including on lapidary inscriptions, literary works, esp. epic poetry (see ex. (198)–(213) above), as well as non-literary documents including correspondence and administration, as the following examples show:

(214) **Inscribed stela** RS 6.028:1–2 (text Bordreuil & Pardee 2009, 218)

→ ⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡
⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡

*pgr*_{<ω>} *d=š^cly*_{<λ>} *zn*_{<ω>} *l=dgn*_{<ω>}
mortuary_sacrifice that=offered PN to=DN

*b^cl-h*_{<ω>}
lord-his

‘Mortuary sacrifice that ‘Uzzinu offered to Dagan his lord’ (trans. Bordreuil & Pardee 2009, 218)

(215) **Legal/Administration** KTU³ 3.12:6–7

→ ⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡
⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡

*mišmn*_{<ω>} *nqmd*_{<λ>} *mlk=úgrt*_{<ω>}
seal PN king=TN

‘seal of Niqmadu, king of Ugarit’

(216) **Correspondence** KTU³ 2.14:10–14

→ ¹⁰ ⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡
¹¹ ⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡
¹² ⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡
¹³ ⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡
¹⁴ ⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡

*w=ht*_{<ω>} *āh=y*_{<λ>} *bn-y*_{<ω>} *ysāl*_{<λ>} *tryl*_{<ω>}
and=now brother-my son-my ask PN

*p=rgm*_{<λ>} *l=mlk*_{<ω>} *šm-y*_{<λ>} *w=l=iytlm*_{<λ>}
CONJ=speak to=king name-my and=to=PN

‘Now, may my brother, my son, ask Ṭarriyilli to speak my name to the king, and to Ṿiyya-talmi.’ (trans. Huehnergard 2012, 193)

(217) **Ritual** RS 1.001:5 (text Bordreuil & Pardee 2009, 198)

→ ⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡⚡

*ālp=w=š=ilhm*_{<ω>} *gdl*_{<ω>} *ilhm*_{<ω>}
bull=and=ram=DN cow DN

‘a bull and a ram for the Ṿilāhūma; a cow for the Ṿilāhūma’ (trans. Bordreuil & Pardee 2009, 198)

For the purposes of the present analysis, we will focus on its manifestation in literary (epic) works. The reason for this is that the literary compositions provide a relatively large (for Ugaritic) corpus of homogeneous texts from which general patterns can be observed. I will, however, return to the orthography of non-literary text types at §8.5 below.

5.7. Textual issues

In our survey of Horwitz (1971) at §5.2.1 above, I highlighted Horwitz's suggestion of the small vertical wedge as a marker of verse structure. However, Horwitz's proposal is, as with any claim relating to ancient texts, reliant on readings of those texts. In this regard, it should be noted that the latest edition of these texts, Dietrich, Loretz & Sanmartín (2013), prints a text that is not compatible with Horwitz's claim, since the first *stichos* comprises three graphematic words, while the second *stichos* comprises four:

(218) KTU 1.17:VI:25

→ →

$w=t'n_{\langle\omega\rangle}$ $btlt_{\langle\lambda\rangle}$ $'nt_{\langle\omega\rangle}$
 and=answered maiden DN
 'and the Maiden 'Anat answered'

(219) KTU 1.17:VI:26

→ →

$irš_{\langle\omega\rangle}$ $hym_{\langle\omega\rangle}$ $l=aqht_{\langle\omega\rangle}$ $g'zr_{\langle\lambda\rangle}$
 ask.IMP life PTCL=PN youth

'Ask for life, O 'Aqhat the youth' (trans. after Horwitz 1971, 92)

This is not to say that Horwitz's readings should be rejected out of hand: he clearly autopsied a number of tablets himself (Horwitz 1971, iii). Furthermore, in collecting his examples of 'irregular' cases (Horwitz 1971, 31–65), he collated Herdner (1963) and the hand-copy of Virolleaud (Horwitz 1971, 30).

Owing to the difficulty of accessing the tablets themselves in the context of the coronavirus pandemic, it has not been possible for me to verify the readings of one or another scholar. The results of the study presented in this chapter were initially found using the collation in Cunchillos, Vita & Zamora (2003), which in most cases represents Dietrich, Loretz & Sanmartín (1976), and then checked against Dietrich, Loretz & Sanmartín (2013). It turned out that Dietrich, Loretz & Sanmartín (2013) is much more wont to read small vertical wedges than Cunchillos, Vita & Zamora (2003)/

Dietrich, Loretz & Sanmartín (1976). Only where small vertical wedges were absent in both Cunchillos, Vita & Zamora (2003) and Dietrich, Loretz & Sanmartín (2013) was a small vertical wedge read as absent.

5.8. Inconsistent nature of univerbation

As previously observed (§5.2), the aspect of word division in Ugaritic that has most beguiled scholars, and which perhaps accounts for the lack of attention to the matter in the literature, is its apparent inconsistency. Thus, even in the case of syntagms where univerbation is more common, it is still by no means obligatory. Consider the following two minimal pairs:

(220) KTU³ 1.2:I:24

→ $\text{𐎧𐎥𐎠𐎵} \text{ } \text{𐎶} \text{ } \text{𐎧} \text{ } \text{𐎶} \text{ } \text{𐎧} \text{ } \text{𐎶} \text{ } \text{𐎧} \text{ } \text{𐎶} \text{ } \text{𐎧} \text{ } \text{𐎶}$

*b=hm*_{<ω>} *ygʿr=bʿl*_{<ω>}

on=them reproach.PREF=DN

‘**Baʿl reproached** them’ (cf. del Olmo Lete & Sanmartín 2015, 287)

(221) KTU³ 1.114:11

→ $\text{𐎧𐎥𐎠𐎵} \text{ } \text{𐎶} \text{ } \text{𐎧} \text{ } \text{𐎶} \text{ } \text{𐎧} \text{ } \text{𐎶} \text{ } \text{𐎧} \text{ } \text{𐎶} \text{ } \text{𐎧} \text{ } \text{𐎶} \text{ } \text{𐎧} \text{ } \text{𐎶}$
 $\text{𐎧} \text{ } \text{𐎶} \text{ } \text{𐎧} \text{ } \text{𐎶} \text{ } \text{𐎧} \text{ } \text{𐎶}$

*b=hm*_{<ω>} *ygʿr*_{<ω>} *tgr*_{<ω>} *bt*_{<ω>} *il*_{<ω>}

on=them reproached guardian house DN

‘The guardian of the house of ʿEl reproached them.’ (for trans. cf. del Olmo Lete & Sanmartín 2015, 335, 889)

(222) KTU³ 1.3:IV:37

→ $\text{𐎧} \text{ } \text{𐎶} \text{ } \text{𐎧} \text{ } \text{𐎶} \text{ } \text{𐎧} \text{ } \text{𐎶} \text{ } \text{𐎧} \text{ } \text{𐎶} \text{ } \text{𐎧} \text{ } \text{𐎶} \text{ } \text{𐎧} \text{ } \text{𐎶}$

*l*_{<ω>} *ttn=pnm*_{<ω>} *m*_{<ω>} *bʿl*

now gave face towards

‘assuredly she set her face towards Baʿl’

(223) KTU³ 1.3:II:8

→ $\text{𐎧} \text{ } \text{𐎶} \text{ } \text{𐎧} \text{ } \text{𐎶} \text{ } \text{𐎧} \text{ } \text{𐎶} \text{ } \text{𐎧} \text{ } \text{𐎶} \text{ } \text{𐎧} \text{ } \text{𐎶} \text{ } \text{𐎧} \text{ } \text{𐎶}$

*tšmt*_{<ω>} *adm*_{<ω>} *šāt*_{<ω>} *špš*_{<λ>}

she_destroyed people coming_out sun

‘She destroyed the people of the rising sun’ (trans. del Olmo Lete & Sanmartín 2015, 775)

In the light of these considerations, it might be supposed that univerbation beyond the basic patterns of word division is a matter of scribal or textual error, or indeed the whim of the writer. Given the propensity for Ugaritic writers to make mistakes (Richardson 1973; Pitard 2012), this is *a priori* reasonable. As we have seen, however, (§5.2), previous work has shown that error alone cannot account for the phenomena observed.

5.9. Hypothesis: Graphematic words represent actual prosodic words

It is conspicuous that the basic patterns of word division in the ‘Majority’ orthography (§5.3) is strongly reminiscent of word division elsewhere, in particular in respect of:

- The univerbation of monoconsonantal prefix clitics with the succeeding morpheme;
- The univerbation of monoconsonantal suffix pronouns with the foregoing morpheme;
- The general distinction between monoconsonantal and multi-consonantal morphemes, whereby the former are generally univerbated with neighbouring morphemes, while the latter are generally written as independent words.

Among the basic patterns of word division identified above, the main difference with respect to word division in Tiberian Hebrew and Northwest Semitic inscriptions is the fact that prefix clitic chains can stand as graphematic words in their own right, as we saw at (212) above.¹¹

According to the general principles laid out above (§1.7.3.1), the distribution of word division is characteristic of separation by prosodic words. With reference specifically to Northwest Semitic, a prosodic basis for word division in Ugaritic would account for a number of the regularly observed phenomena, including:

- **Univerbation of monoconsonantal prefixes** since monoconsonantal prefixes are generally proclitic in Semitic languages;
- **Differential treatment of simplex and extended monoconsonantal prefixes:** *e.g.* (210). Compare Tiberian Hebrew, where extended prepositions are, unlike their simplex counterparts, capable of carrying their own accent.¹²

The most significant difference between word division in the Ugaritic ‘Majority’ orthography, and that seen in Tiberian Hebrew and inscriptions is that an important minority of graphematic words consist of considerably longer units than would be

¹¹ This too is paralleled, however, if the net is thrown wider than Northwest Semitic, since it is found in Old South Arabian (Beeston 1984: 6). (My thanks to Aaron Koller for this reference.)

¹² For differential prosodic statuses of simplex and extended prepositions in Ugaritic, see Gzella (2007b, 546), where under $\text{P} \text{—} k$ he notes: ‘*k* appears in its long form *km*, counting as a prosodic unit on its own, and could thus in theory precede a noun prefixed by another proclitic preposition.’ The comment presupposes a situation, inherited from Proto-Semitic, where monoconsonantal prepositions do not carry their own accent.

expected on the basis of the basic patterns of word division outlined so far, resulting in graphematic words spanning multiple multiconsonantal morphosyntactic words. This means that these graphematic units are also longer than those that we see in general in Tiberian Hebrew and in Northwest Semitic inscriptions.

If error or whim are not responsible (see §5.8 above), we would expect to see a linguistic motivation for what we observe. Therefore, to go beyond existing work, any underlying framework for word division in Ugaritic proposed should be able not only to tolerate and provide a context for such variation. It is the contention of the present chapter that there is indeed a linguistic motivation for the presence of longer graphematic words in the Ugaritic ‘Majority’ orthography. Specifically, I argue that the distribution of longer graphematic units in Ugaritic matches what one would expect of ACTUAL PROSODIC WORDS. I base this argument on comparison both with Tiberian Hebrew and cross-linguistic evidence more generally. The analysis is based on an analysis of the following subcorpora of the Ugaritic mythological material: the Baʿl cycle (KTU 1.1–6), the Keret epic (KTU 1.14–16) and the first tablet of Aqhat (KTU 1.17).

Note, however, I will not seek to offer a comprehensive solution able to predict the use or non-use of the small vertical wedge as a word divider in a particular case. This is to say that I will not seek to prove or disprove Horwitz’s proposal regarding the use of the small vertical wedge as a metrical device (see §5.2.1 and §5.7 above). Instead, as will be seen, I seek merely to show that the use, and especially the non-use, of the small vertical wedge is compatible with demarcating ACTUAL PROSODIC WORDS.

Chapter 6

The Ugaritic ‘Majority’ orthography

6.1. Introduction

The present chapter provides an overview of phenomena associated with word division, and the lack of it, in the Ugaritic ‘Majority’ orthography. The data underlying the generalisations are presented in Chapter 7 below.

6.2. Syntagms particularly associated with univertation

Univertation in Ugaritic tends to be found in particular syntagms, notably the following:

- Preposition + Noun
- Noun + Noun
- Verb + Preposition, Noun or Particle (*i.e.* verb first)

By contrast, univertation tends not to be found in the following syntagms:

- X + Verb (*i.e.* verb second)
- Noun + Preposition
- Pronominal Suffix + Noun

In terms of tying down the ORL of word division and univertation in Ugaritic, it is essential to account for this distribution. For now, however, we simply describe and exemplify the observed behaviour in each case.

6.3. Univertation with nouns

Univertated Noun + Noun sequences may either be nouns in apposition, *e.g.* *zbl ym* ‘prince DN’ (KTU 1.2:IV:16), or nouns in construct *gr bt il* ‘guest of the sanctuary’ (KTU 1.19:III:47).¹ The latter may in turn involve either proper or common nouns:

¹ Robertson (1994) finds no examples of univertation of nouns in apposition in her corpus, and so does not discuss the rules that determine its distribution (Robertson 1994, 231).

Univerbated construct chains may comprise more than two elements², e.g.:³

(224) KTU³ 1.2:III:8

→ 𐎧𐎥𐎵 – 𐎧𐎥𐎵 𐎠𐎢𐎣𐎧

*bḥt=zbḥ=ym*_{<λ>}

palace=prince=DN

‘the palace of prince Yam’ (trans. del Olmo Lete & Sanmartín 2015, 246)

Where a noun in apposition and a construct chain appear in sequence relating to the same referent, either the noun in apposition may be univerbated with part of the chain, or the elements in construct may be univerbated with one another. However, it is not normally the case that the whole chain is written together, e.g.:

(225) KTU³ 1.3:IV:7

→ 𐎠𐎢𐎣𐎧 – 𐎠𐎢𐎣𐎧 𐎠𐎢𐎣𐎧 𐎠𐎢𐎣𐎧 𐎠𐎢𐎣𐎧

*thm=āliyn*_{<ω>} *bḥl*_{<ω>}

message=Almighty DN

‘message of the Almighty, Baʿl’

(226) KTU³ 1.5:II:11

→ 𐎧𐎥𐎵 𐎠𐎢𐎣𐎧 𐎠𐎢𐎣𐎧 𐎠𐎢𐎣𐎧 𐎠𐎢𐎣𐎧

*bḥt*_{<ω>} *l=bn*_{<ω>} *ilm=mt*_{<λ>}

hail INTERJ=son DN=DN

‘Hail, O son of El, Môt’ (for interpretation cf. Pardee 2003, 266; del Olmo Lete & Sanmartín 2015, 482)

Other univerbated combinations involving nouns are also possible, including:

- **N + d(t) + N** e.g. *gpny dt ksp* ‘harnesses of silver’;
- **N + Adj.** e.g. *bk rb* ‘a great beaker’;
- **Num + N** e.g. *ālp ḥzr* ‘a thousand palaces’;
- **Elliptical phrases** e.g. *w ym ymm* ‘one day and two (passed)’

It may be seen from the foregoing analysis that univerbation of NPs in Ugaritic is by no means restricted to potentially fossilised expressions, such as *zbl ym* ‘prince DN’. In particular, the univerbation of presumably productive N + N construct chains such as *l bmt ḥr* ‘on the back of an ass’ points to a productive basis for the univerbation.

² For these purposes, any unit that would normally be written together, such as the combination of a monoconsonantal preposition with a noun, or a monoconsonantal preposition and any other word, is treated as a single unit.

³ Cf. also KTU 1.6:VI:30: *bn ilm t (mt)* ‘son of the gods DN’.

6.4. Univerbation with verbs

Verbs may be univerbated with nouns, prepositions or particles. What univerbated verbal syntagms share is the fact that the verbal element tends to come first:

(227) KTU³ 1.4:V:42

→ 

w=tb=l=mspr_{<w>}

and=return.IMPV.SG=to=narrative

‘And return to the narrative!’ (cf. del Olmo Lete & Sanmartín 2015, 483)

(228) KTU³ 1.2:I:24

→ 

b=hm_{<w>} yg‘r=b‘l_{<w>}

on=them reproach.PREF=DN

‘Ba‘l reproached them’ (cf. del Olmo Lete & Sanmartín 2015, 287)

In the case of nouns, this tendency applies regardless of the particular syntactic role that a given noun will play, whether subject or object. In (228) it is the subject that is written together with the verb. In the following examples we have a direct object:

(229) KTU³ 1.3:IV:37

→ 

l_{<w>} ttn=pmn_{<w>} ‘m_{<w>} b‘l

PTCL gave face towards

‘assuredly she set her face towards Ba‘l’

In contrast to univerbated syntagms where the verb comes first, those where the verb comes second are rare. The following is in fact the only case I could find in the Ba‘l epic where univerbation does not lie across a clause/colon boundary (cf. §8.4 below):

(230) KTU³ 1.1:III:11

→ 

‘m=y=twth_{<w>} iš‘d-k_{<w>}

towards=me=let_hasten steps-your

‘Towards me let your steps hasten’ (trans. del Olmo Lete & Sanmartín 2015, 929)

6.5. Univerbation with suffix pronouns

As previously mentioned and exemplified above (§5.3, (203)) monoconsonantal suffix pronouns are always written together with the preceding morpheme. This is also usually the case for biconsonantal – so-called ‘heavy’ – suffix pronouns, e.g.:

(231) KTU³ 1.2:I:29

→ 

tšū=ilm_{<ω>} rāšt=hm_{<ω>}

raised=gods heads=their

‘the gods raised their heads’ (trans. del Olmo Lete & Sanmartín 2015, 713)

Occasionally, however, a heavy suffix pronoun is written as a separate word:


(232) KTU³ 1.15:VI:6–7 (example cited Tropper 2012, 68)

→ 

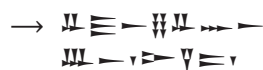
k̄m_{<λ>} rgm_{<ω>} tr̄m̄_{<ω>} rgm_{<ω>} hm_{<λ>}

like voice bull voice their

‘like the voice of a bull was their voice’

It also happens that a suffix pronoun may be univerbated with a following morpheme, although this is not frequent. We see this at (230) above and in the following example in the case of -y:⁴

(233) KTU³ 1.4:VI:36

→ 

ht-y=bnt_{<λ>} dt_{<ω>} ksp_{<ω>}

house-my=construction of silver

‘my house is a construction of silver’

6.6. Univerbation at clause and phrase boundaries

A most striking feature of univerbation in Ugaritic – and to my knowledge not previously discussed – is that it may, in a small minority of instances, lie across a syntactic phrase or clause boundary. The following examples illustrate univerbation across phrase and clause boundaries respectively:

⁴ Cf. also KTU 1.2:IV:11, 1.6:II:22.

(234) KTU³ 1.2:IV:11

→ 𐎧𐎠𐎫𐎠𐎫𐎠𐎫 𐎠𐎧 𐎠𐎧𐎠𐎧 𐎠𐎧𐎠𐎧𐎠𐎧

*ktr=šmdm*_{<ω>}*ynht*_{<ω>}[DN_{SUBJP}]=[[double_mace_{NP}] brought_{VP}]

'DN brought a double mace' (trans. per del Olmo Lete & Sanmartín 2015, 620)

(235) KTU³ 1.3:III:14–15

→ 𐎧𐎠𐎫𐎠𐎫𐎠𐎫 𐎠𐎧𐎠𐎧𐎠𐎧𐎠𐎧

𐎠𐎧𐎠𐎧𐎠𐎧𐎠𐎧𐎠𐎧 𐎠𐎧𐎠𐎧𐎠𐎧𐎠𐎧𐎠𐎧

*qryy*_{<ω>}*b=ârš*_{<λ>}*nīlḥmt=št*_{<ω>}*b=ʿprm*_{<ω>}*ddym*_{<λ>}[meet.IMP in-land war_s]=[put in-steppe harmony_s]

'Meet war in the land, put harmony in the steppe' (trans. del Olmo Lete & Sanmartín 2015, 264, 704)

To the extent that there is any relationship between syntax and word division, such a distribution of univerbation is, of course, unexpected. This is especially so in the case of inter-clausal univerbation. I reserve dedicated discussion of these instances until §8.4, that is, until after the syntax of word division has been addressed.

6.7. Summary

To summarise, in the above survey of univerbation in Ugaritic alphabetic cuneiform we have observed that graphematic univerbation is particularly associated with the following POS sequences:

- Prep + Noun
- Noun + Noun
- Verb + X (*i.e.* verb first)

By contrast, univerbation is negatively associated with the combinations:

- X + Verb (*i.e.* verb second)
- Noun + Prep
- SuffPron + Noun

We have, however, so far refrained from offering an explanation of the phenomena. What might be said to account for this distribution? Per the earlier discussion, it is plausible that the productive basis on which these NPs are univerbated is prosodic, and that the NPs so univerbated represented prosodic words. If, for the time being, we assume this to be the case, that is, that word division in Ugaritic is a representation of prosody, graphematic words can be considered to represent prosodic words or prosodic phrases. This state of affairs would have the following implications for the significance of the distribution of univerbation observed immediately above:

- Prepositions have a tendency to be prosodically associated with the nouns they govern;
- Verbs in Ugaritic have a tendency to be proclitic;
- Noun phrases have a tendency to function as a single prosodic unit.

It could, of course, be proposed that univerbation targets a syntactic relationship, rather than a prosodic one. However, against this is particularly the fact that graphematic word boundaries do not as a rule coincide with syntactic phrase boundaries.

Circumstantially, then, the distribution of graphematic univerbation in Ugaritic alphabetic cuneiform coincides with what might be expected of a prosodic phenomenon. In order to further demonstrate this, however, it is helpful to compare graphematic univerbation in Ugaritic with graphematic representations of prosodic phrasing in a closely related language. Tiberian Hebrew provides a good comparandum for this purpose, given its elaborate system of prosodic word and prosodic phrase representation through its cantillation tradition.

The one piece of evidence that might be said to speak against a prosodic interpretation is that graphematic univerbation does, on occasion, occur at phrase and clause boundaries, especially in bi- and tricola.

Chapter 7

Quantitative comparison of Ugaritic and Tiberian Hebrew

7.1. Introduction

The present chapter presents quantitative data on the distribution of graphematic univertation in the Ugaritic ‘Majority’ orthography, compared with the distribution of both *maqef* and conjunctive accents in Tiberian Hebrew. As noted previously, irregularities of word division involving the placing of the small vertical wedge where it would not be expected according to the basic patterns of word division (§5.4), including between a monoconsonantal particle and a following morpheme, were not considered in this analysis. These phenomena are addressed in Chapter 9.

7.2. Corpus

For the quantitative investigation described in the present chapter, the corpora analysed were as follows:

- **Ugaritic:** KTU 1–6, 14–17;
- **Tiberian Hebrew:** Genesis, Exodus, Leviticus, Numbers, Deuteronomy, Joshua, Judges, Ruth, 1 Samuel, 2 Samuel, 1 Kings, 2 Kings, 1 Chronicles, 2 Chronicles, Ezra, Nehemiah, Esther, Qohelet (Ecclesiastes), Isaiah, Jeremiah, Lamentations, Ezekiel, Daniel, Hosea, Joel, Amos, Obadiah, Jonah, Micah, Nahum, Habakkuk, Zephaniah, Haggai, Zechariah, Malachi.

Original analysis of the Ugaritic data was conducted on the basis of UDB (Cunchillos, Vita & Zamora 2003), and subsequently cross-checked against KTU³ (Dietrich, Loretz & Sanmartín 2013). Examples were excluded where UDB showed a text with univertation but KTU³ did not.

In the case of Ugaritic, the need to take account of syntagms in the corpus as a whole meant that it was only possible to include KTU 1–6 in the morphosyntactic collocational part of the investigation, §7.5 to §7.7 (for more information see §7.5.6).

The Hebrew text used for this part of the study was the Open Scriptures Hebrew Bible (<https://github.com/openscriptures/morphhb>) parsed using Python scripts written by the author. In the Hebrew corpus Job, Psalms and Proverbs were not

included because of the different system of accents used there.¹ In addition Song of Songs was also excluded.

7.3. Frequency of occurrence

Graphematic univerbation in Ugaritic is considerably less frequent than are either *maqfef* or the conjunctive accent in Tiberian Hebrew. This is clearly shown in Table 7.1 and Table 7.2: while conjunctive accents and *maqfef* are used in 30.87% and 14.06% of sequences respectively, Ugaritic univerbation (KTU 1–6 and KTU 14–17) is found in only 2.57% of sequences.²

Table 7.1: Frequencies of joining features (*maqfef*, conjunctive accent, disjunctive accent) in Tiberian Hebrew

Accent type	Freq.	%
Disjunctive	147315	54.61
Conjunctive	83274	30.87
<i>Maqqef</i>	37918	14.06
All sequences	269758	

Table 7.2: Frequency of univerbation (pairs of words joined) in Ugaritic (KTU 1–6 and 14–17)

	Freq.	%
Univerbation	159	2.57
All sequences	6184	

7.4. Length of phrase

Despite its lower incidence overall, the distribution of univerbation where it is found in Ugaritic parallels the distributions of both *maqfef* and conjunctive accents. In this section we consider the length of phrases joined by univerbation, *maqfef* and conjunctive accents, in Ugaritic and Tiberian Hebrew respectively, before considering their morphosyntactic context in §7.5.

Conjunctive phrases in Tiberian Hebrew range in length from two to six prosodic words. In order to facilitate comparison with Ugaritic, since the purpose of the exercise is to establish the status of graphematic words there, conjunctive phrases consisting of one or more *maqfef* phrases were not counted. Accordingly, for these purposes prosodic words are equivalent to graphematic words. The following verse gives a case of a prosodic phrase of six graphematic words in length:

¹ There are two types of accent systems in use in the Hebrew Bible, one for the so-called ‘Twenty-one books’ and the other used for the ‘Three books’. While the symbols used to demarcate phrases are distinct in each, the systems work according to the same principles. For more information, see Yeivin (1980) and Park (2020).

² For both Ugaritic and Hebrew this figure was calculated by counting the number of POS sequences linked by a particular linking feature (e.g. *maqfef*, univerbation etc.), and dividing this by the number of POS sequences irrespective of univerbation/accenuation.

(236) 1Kgs 6:1

← וַיְהִי בַשְּׁמוֹנִים עָשָׂר וָאַרְבַּע מֵאוֹת שָׁנָה

(w=yhy_ω b=šmw_ωny_ωm_ω šnh_ω w=ʾrb_ω^c mʾwt_ω)
 and=it_happened in=eighty year and=four hundred

šnh_{ω φ})

year

'And it happened in the four hundred and eightieth year'

By contrast, *maqfef* phrases range in length from two to four graphematic words, with the following exemplifying a phrase of four graphematic words:

(237) Gen 2:6

← וַהֲשִׁקָה אֶת-כָּל-פְּנֵי-הָאָרֶץ:

(w=hšqh_{<ω> φ}) (ʾt_{<ω>} =kl_{<ω>} =pny_{<ω>} =h=ʾdmh_{<ω> φ})
 and=watered OBJ=whole=face=the=ground

'(a mist) watered the whole face of the ground' (after KJV)

The relative distributions of the various phrase lengths attested for conjunctive accents and *maqfef* are given in Table 7.3 and Table 7.4 respectively.³ There it may be seen that the vast majority of phrases in both cases consist of two graphematic words, with phrases of greater sizes diminishing in number very sharply.

Table 7.3: Frequencies of conjunctive phrases of different lengths in Tiberian Hebrew

Variable	Freq.	%
2	66266	93.279
3	4265	6.004
4	466	0.656
5	43	0.061
6	1	0.001
Total	71041	

Table 7.4: Frequencies of *maqfef* phrases of different lengths in Tiberian Hebrew

Variable	Freq.	%
2	36593	96.625
3	1253	3.309
4	25	0.066
Total	37871	

In KTU 1–6 and 14–17, univertated graphematic words range from two to four units long, where a unit is a graphematic word according to the basic patterns of

³ For the Hebrew corpus, see §7.2.

word division laid out at §5.3. The longest example in the corpus studied was four morphosyntactic words in length:

(238) = (198) KTU 1.2:IV:12

→ →

ygrš_(ω) grš=ym=grš=ym_(ω)

l=ksi[?]=h_(λ)

DN₁.VOC drive_away.PTPL=DN₂=drive_away.IMP=DN₂

from=throne-his

‘Ygrš, who drives away Yam, drive away Yam from his throne’

Ugaritic univertation as a phenomenon is therefore comparable in scale to conjunctive and *maqgef* phrases in Tiberian Hebrew, and particularly so to *maqgef*. Furthermore, Table 7.5 shows that the relative proportions of graphematic word phrases of differing lengths parallels those of both conjunctive accents and *maqgef* in Tiberian Hebrew. Once again, however, the distribution is closer to that of *maqgef*.⁴

The data presented here should, however, be treated with some caution, in two respects. First, the Ugaritic dataset considered is, necessarily, owing to constraints of time and morphologically parsed data, considerably smaller than that of Tiberian Hebrew.

Table 7.5: Frequencies of univertated phrases of different lengths in Ugaritic (KTU 1-6, 14-17)

Variable	Freq.	%
2	152	98.065
3	2	1.29
4	1	0.645
Total	155	

7.4.1. Taking account of line division

A second important difference between the Hebrew and Ugaritic datasets is that the Ugaritic data is split into lines of varying lengths, while that in Hebrew is not. In Ugaritic the small vertical wedge is generally not used at line division, although there are isolated exceptions. Accordingly, there are likely to be graphematic word sequences that would be univertated were they to have been written on a longer line, or at a different point in the line. I leave it to future research to fully account of this issue. However, it is possible to go some way towards mitigating this difficulty by creating line divisions artificially in the Tiberian Hebrew data.

This was achieved by taking the following steps:

- Graphematic words in a range from two to seven units apart were randomly selected, using Python’s random number generator, to have their accents removed. This simulates line division in Ugaritic because, in the same way that in Ugaritic line division removes information about univertation, so in this modified dataset, information about the accentuation of a Hebrew word is removed.
- The modified dataset was processed in the same way as before.

⁴ Unfortunately, owing to some expected frequencies being less than five, it was not possible to perform chi-square tests on the distributions.

The effect of artificially modifying the Hebrew dataset in this way can be seen in Table 7.6 and Table 7.7.

Table 7.6: Frequencies of conjunctive phrases of different lengths in Tiberian Hebrew, with account taken of line divisions

Variable	Freq.	%
2	36908	95.513
3	1624	4.203
4	107	0.277
5	3	0.008
Total	38642	

Table 7.7: Frequencies of maqqef phrases of different lengths in Tiberian Hebrew, with account taken of line divisions

Variable	Freq.	%
2	28800	97.531
3	716	2.425
4	13	0.044
Total	29529	

It can be seen from the tables that the effect of adjusting for line division is to move the distributions of both conjunctive and *maqqef* phrases closer to that of Ugaritic univerbation. The net effect is that the distribution of Ugaritic univerbated phrase lengths becomes yet closer to that of *maqqef*.

In terms of relative phrase length, therefore, it can be concluded that the unit demarcated by word division in Ugaritic has affinities both to *maqqef* phrases and to conjunctive phrases in Tiberian Hebrew, but that it is closer to that of *maqqef* than to conjunctive phrases. In the next sections, I consider the morphosyntactic context of these phrases, comparing Ugaritic univerbation with Tiberian Hebrew prosodic word- and prosodic phrase-hood. In order to do this, however, it is first necessary to establish a quantitative method for comparing the two.

7.5. Quantifying the morphosyntactic collocation of linking features

7.5.1. Introduction

We have seen that while syntax and prosody are not isomorphic, they have an important structural relationship (§1.5). Therefore, insofar as word division is a function of prosody, it is expected that word division and the joining behaviour (univerbation, separation) of particular parts of speech (POS, *i.e.* noun, verb, preposition etc.) that are joined/separated should have a non-random collocational ‘signature’. By measuring the collocational relationship between POS and word division in Ugaritic, and comparing it with the collocational relationship between the same POS and accents in Tiberian Hebrew (*maqqef*, conjunctive, disjunctive), it is possible to locate Ugaritic univerbation with respect to Hebrew accents (*maqqef*, conjunctive, disjunctive).

7.5.2. Counting POS sequences

The simplest measure of the distribution of graphematic univerbation is to count the number of instances of particular POS sequences that are written as a single graphematic unit. For a given corpus, this value is described by the following expression:

$$\left| I_{w_a:w_b} \right| \quad \text{Eq. 1}$$

where W is the set of all POS, and $|X|$ is the size of the set X , I_w is the set of instances of a given POS w , $|I_w|$ is the size of that set, $I_{w_a:w_b}$ is the set of instances of POS a occurring to the left of POS b under a particular linking feature (univerbation, *maqfef*, conjunctive accent), and $|I_{w_a:w_b}|$ is the size of that set.

7.5.3. Proportion of Occurrence

Of course, the raw frequency information is only useful for comparison between two subcorpora if the total number of instances of the linking feature are the same in the two subcorpora. Since this is unlikely to be the case, the proportion of sequences joined by the linking feature needs to be obtained.

A Proportion of Occurrence of a POS sequence (*e.g.* Noun–Noun) for a linking feature (*e.g.* *maqfef*) is obtained by dividing the frequency of the POS sequence joined by the linking feature by the total of all POS sequences joined by the linking feature, *i.e.*:

$$\frac{\left| I_{w_a:w_b} \right|}{\sum_{w \in W} \left| I_{w_a:w_b} \right|} \quad \text{Eq. 2}$$

Table 7.8 provides an overview of the various syntagms that are joined in this way. The top four combinations, accounting for nearly three quarters (79.83%) of the total, are the following:

- Noun + Noun
- Verb + Noun
- Prep + Noun
- Noun + Verb

7.5.4. Association score A

The drawback of using POS sequence frequency as a collocational measure of comparison in isolation is that the frequency of a particular POS sequence joined by a particular linking feature, *e.g.* *maqfef* (in Hebrew) or univerbation (in Ugaritic), will be affected by the frequency of that POS combination in the corpus

Table 7.8: Frequencies and Proportions of Occurrences of unverbated syntagms in KTU1-6

POS sequence	Freq.	%
Noun + Noun	57	47.9
Verb + Noun	24	20.17
Prep + Noun	9	7.56
Noun + Verb	5	4.2
Verb + Prep	4	3.36
Noun + Ptcl	4	3.36
Noun + Adj	3	2.52
SuffPron + Pron	2	1.68
Verb + Pron	2	1.68
Verb + Ptcl	1	0.84
Noun + Prep	1	0.84
Adv + Verb	1	0.84
Ptcl + Noun	1	0.84
Verb + Adv	1	0.84
Verb + Adj	1	0.84
SuffPron + Verb	1	0.84
SuffPron + Noun	1	0.84
Ptcl + Ptcl	1	0.84
Total	119	

as a whole, regardless of whether or not the combination is unverbated. This measure can still be helpful in contexts where the distribution in the corpus as a whole is also represented, but otherwise, it is more representative to try to take account of this factor.

Suppose, for illustrative purposes, that we are interested in comparing two subcorpora, Genesis and Exodus, in terms of the distribution of *maqfef* according to the parts of speech that it joins. Suppose then that *maqfef* is particularly associated with Noun–Noun sequences, and also that Noun–Noun sequences occur more frequently in Genesis than in Exodus. There will therefore be a comparatively greater incidence of Noun–Noun sequences joined by *maqfef* in Genesis than in Exodus simply because Noun–Noun sequences occur more frequently there, and not (necessarily at least) because there is a stronger degree of association between *maqfef* and Noun–Noun sequences in Genesis.

The incidence of POS combinations in the corpus as a whole can be accounted for by calculating the ratio of the frequency of a given linking feature, *e.g.* *maqfef*, linking two POS, to the frequency of that combination occurring irrespective of the presence of the joining feature, *i.e.*:

$$\frac{|I_{w_a, w_b}|}{|I_{w_a, w_b}|} \quad \text{Eq. 3}$$

where I_{w_a, w_b} is the set of instances of POS a occurring to the left of POS b irrespective of the presence of the presence of a linking feature. The ratio can in turn be expressed as a percentage.

The ratio obtained by Eq. 3 gives a measure of association between a given linking feature, such as *maqfef*, and a particular POS sequences, *e.g.* Noun + Noun, so that a value of 1 (or 100%) would mean that all Noun + Noun sequences in the corpus are joined by *maqfef*. As an example, Table 7.9 gives these ratios for the book of Genesis in the Hebrew Bible, down to a ratio of 1%.

In what follows, the Association Score obtained by Eq. 3 will be termed Association Score A.

7.5.5. Low frequencies

From Table 7.9 it can be seen that a number of POS sequences have very high ratios, including Prep + Adv, Prep + Noun, Ptcl + Noun, Prep + Ptcl and Conj + Pron. On the face of it, this suggests a high degree of association between *maqef* and these particular syntagms.

However, the figures for these syntagms may be unreliable if the total number of instances of these particular POS sequence is very low. A small number of datapoints, in the context of a relatively large corpus, such as the book of Genesis, might have a disproportionate effect on the Association Scores.

For the syntagms with the top five Association Scores there is a wide range of prevalence in Genesis as a whole:

- Prep + Adv: 14 instances (0.07%)
- Prep + Noun: 540 instances (2.62%)
- Ptcl + Noun: 757 instances (3.68%)
- Prep + Ptcl: 256 instances (1.24%)
- Conj + Pron: 8 instances (0.04%)

The difficulty can be mitigated by setting a threshold to exclude low frequency items. For example, excluding POS combinations that account for less than 0.5% of combinations in Genesis yields the results in Table 7.10.

7.5.6. Integrating the Ugaritic data

For Biblical Hebrew, it is relatively easy to obtain an Association Score, since open morphologically parsed texts are readily available. The results in Table 7.9 and Table 7.10 were obtained using the texts of the Open Scriptures Hebrew Bible (<https://github.com/openscriptures/morphhb>) parsed using Python scripts written by the author. By contrast, such texts are not available for Ugaritic. While an electronic text does exist in the form of the Ugaritic Databank (UDB) (Cunchillos, Vita & Zamora 2003), this is not linguistically analysed. Accordingly, linguistic analysis must be done by hand. The quantity of data that can be analysed is therefore small compared to Hebrew.

Table 7.9: Association Score A for *maqef* in Genesis (Assoc. Score A > 20%)

POS Sequence	Assoc. Score A
Prep + Adv	92.86
Prep + Noun	82.59
Ptcl + Noun	79.66
Prep + Ptcl	79.3
Conj + Pron	75.0
Ptcl + Ptcl	68.71
Adv + Pron	68.42
Conj + Ptcl	62.32
Conj + Verb	58.6
Ptcl + Adv	58.33
Ptcl + Pron	57.32
Conj + Adv	57.14
Conj + Adj	53.33
Ptcl + Adj	51.22
Conj + Noun	50.0
Prep + Adj	50.0
Ptcl + Verb	32.38
Ptcl + Prep	30.43
Conj + Prep	28.12
Verb + Adv	26.17
Prep + Verb	21.15

Table 7.10: Association Score A for maqqef in Genesis (Assoc. Score A > 20%; POS prevalence in Genesis > 0.5%)

Sequence	Assoc. Score A	Pop. freq.	Pop. %
Prep + Noun	82.59	540	2.62
Ptcl + Noun	79.66	757	3.68
Prep + Ptcl	79.3	256	1.24
Ptcl + Ptcl	68.71	278	1.35
Conj + Verb	58.6	186	0.9
Ptcl + Verb	32.38	593	2.88
Ptcl + Prep	30.43	184	0.89
Verb + Adv	26.17	107	0.52

In particular, it was not practical to find all the POS sequences in the Ugaritic corpora under analysis in this chapter, namely KTU 1–6 and 14–17. Instead, smaller samples were taken, from KTU 1–6 only, and the POS sequences irrespective of univerbation were counted in these samples. The samples chosen were:

- KTU 1.2:I:20–29
- KTU 1.2:IV:5–12
- KTU 1.4:IV:8–44
- KTU 1.5:I:1–9

The sample contained 281 tokens. The small size of the sample means that the Association Scores obtained for Ugaritic are necessarily contingent on a more complete analysis being conducted. However, despite the small size of the dataset, it will be seen that the quantitative data do complement the results obtained on linguistic grounds.

7.5.7. Association Score B

The method of obtaining Association Score A for Ugaritic described at §7.5.6 carries the unfortunate entailment that the scores for Ugaritic cannot be compared directly with those obtained for Tiberian Hebrew. The problem is that Ugaritic Association Scores obtained in this way may be greater than 1. This is because only a subset of instances of w_a, w_b are included, *i.e.*:

$$\frac{|I_{w_a;w_b}|}{|i_{w_a,w_b}|} \quad \text{Eq. 4}$$

where:

$$i_{w_a,w_b} \subset I_{w_a,w_b} \quad \text{Eq. 5}$$

If $|I_{w_a;w_b}|$ is greater than $|i_{w_a,w_b}|$, the result will be greater than 1. This could clearly never be the case in the corpus as a whole, since:

$$I_{w_a;w_b} \subseteq I_{w_a,w_b} \quad \text{Eq. 6}$$

and therefore:

$$\left| I_{w_a:w_b} \right| \leq \left| I_{w_a, w_b} \right| \quad \text{Eq. 7}$$

To obtain a measure that can be used to compare directly the Hebrew Bible books with the Ugaritic corpus, we calculate Association Score B. This is the ratio of the Proportions of Occurrences (per Eq. 2) of a) POS sequence $w_a.w_b$ and b) the same sequence collocating irrespective of the linking feature, *i.e.* $w_a.w_b$, *i.e.*:

$$\frac{\left| I_{w_a:w_b} \right|}{\sum_{w \in W} \left| I_{w_a:w_b} \right|} \cdot \frac{\left| I_{w_a, w_b} \right|}{\sum_{w \in W} \left| I_{w_a, w_b} \right|} \quad \text{Eq. 8}$$

This is equivalent to:

$$\frac{\left| I_{w_a:w_b} \right|}{\sum_{w \in W} \left| I_{w_a:w_b} \right|} \cdot \frac{\sum_{w \in W} \left| I_{w_a, w_b} \right|}{\left| I_{w_a, w_b} \right|} \quad \text{Eq. 9}$$

Since by definition:

$$\left| I_{w_a:w_b} \right| \leq \sum_{w \in W} \left| I_{w_a, w_b} \right| \quad \text{Eq. 10}$$

and:

$$\left| I_{w_a, w_b} \right| \leq \sum_{w \in W} \left| I_{w_a, w_b} \right| \quad \text{Eq. 11}$$

each of these figures will always be between 0 and 1. By this measure, the Ugaritic results, and those from books of the Hebrew Bible may be compared directly.

Of course, the result of Association Score B itself can be any number greater than 0. An Association Score B of greater than 1 for a particular POS sequence says that that sequence is a greater proportion of sequences collocating with the linking feature

Table 7.11: Association Score B for POS sequences in KTU 1–6 (Assoc. Score B > 0; POS prevalence in Genesis > 0.5%)

POS sequence	Assoc. Score B
Prep + Noun	2.35
Noun + Noun	2.03
Verb + Noun	1.41
Verb + Prep	1.34
Noun + Adj.	1.18
Verb + Adj.	0.78
Noun + Ptcl	0.72
Verb + Ptcl	0.59
Ptcl + Noun	0.39
SuffPron + Verb	0.39
Noun + Verb	0.38
SuffPron + Noun	0.29
Noun + Prep	0.18

(e.g. univerbation, *maqef* etc.) than in the population as a whole, and is therefore positively associated with the linking feature. By contrast, a figure below 1 says that the POS sequence is negatively associated with the linking feature.

Association Score B has the problem encountered before (§7.5.5) that low frequency items can have a disproportionate effect on the rankings of POS sequences under the linking feature. Accordingly, it is still helpful to set a minimum threshold for instances of particular POS sequences in order to mitigate the disproportionate effect of sequences attested only a small number of times. The tables for KTU 1–6 and Genesis are given at Table 7.11 and Table 7.12 respectively.

Table 7.12: Association Score B for POS sequences in Genesis (Assoc. Score B > 0; POS prevalence in Genesis > 0.5%)

Sequence	Assoc. Score B
Prep + Noun	5.76
Ptcl + Noun	5.55
Prep + Ptcl	5.53
Ptcl + Ptcl	4.79
Conj + Verb	4.08
Ptcl + Verb	2.26
Ptcl + Prep	2.12
Verb + Adv	1.82
Noun + Noun	1.35
Noun + Adj.	0.89
Noun + Ptcl	0.83
Verb + Prep	0.64
Noun + Pron	0.39
Verb + Noun	0.36
Adj. + Noun	0.31
Verb + Ptcl	0.27
Pron + Noun	0.21
Noun + Verb	0.17
SuffPron + Ptcl	0.16
Pron + Verb	0.13
SuffPron + Prep	0.05
Verb + Verb	0.04
Noun + Conj	0.01
SuffPron + Noun	0.01
Noun + Prep	0.01

7.6. Measuring Association Score B for Ugaritic and Tiberian Hebrew

7.6.1. Ugaritic

Table 7.13 shows that graphematic univerbation is positively associated with the following POS combinations. Recall again that any number above 1 indicates a positive association between the particular combination of words and univerbation:

- Prep + Noun
- Noun + Noun
- Verb + Noun
- Verb + Prep
- Noun + Adj.

This is to say that univerbation is positively associated with verb-initial sequences, prepositional phrases and nominal combinations including both Noun + Noun and Noun + Adj. To see the significance of this distribution for the semantics of word division/univerbation in Ugaritic, it is helpful to compare the distribution with Tiberian Hebrew accentuation involving *maqfef* and conjunctive accents.

7.6.2. Tiberian Hebrew

Let us consider those POS combinations that are positively associated with *maqfef* and conjunctive accents respectively. The most important of these are given in Table 7.14. The sequences given fall into one of the following categories:

- Prep + X
- Ptcl + X
- Noun + Noun

By contrast, the equivalent POS sequence joined by conjunctive accents are given in Table 7.15. Here, the POS sequences fall into the following predominant categories:

- Verb + X
- Ptcl + X
- Noun + X (where Noun is any nominal, *i.e.* Adj. or Noun)

The main points of contact are in the fact that the sequences Ptcl + X and Noun + Noun is found joined both by *maqfef* and by conjunctive accents. A major difference between them is that Verb + X syntagms are positively associated with conjunctive accents, but not with *maqfef*.

Table 7.13: Univerbated POS sequences in KTU 1–6 (Assoc. Score B, pop. % > 0.5%)

POS sequence	Assoc. Score B
Prep + Noun	2.35
Noun + Noun	2.03
Verb + Noun	1.41
Verb + Prep	1.34
Noun + Adj.	1.18

Table 7.14: POS sequences joined by *maqfef* in Tiberian Hebrew (Assoc. Score B, POS prevalence in BH corpus > 0.5%)

POS sequence	Assoc. Score B
Prep + Noun	5.75
Prep + Ptcl	5.68
Ptcl + Noun	5.3
Ptcl + Ptcl	4.81
Conj + Verb	4.13
Ptcl + Verb	2.52
Ptcl + Prep	1.63
Noun + Noun	1.58
Adv + Verb	1.03

7.6.3. Comparing Tiberian Hebrew and Ugaritic

The Tiberian Hebrew and Ugaritic data can be compared directly according to Association Score B for those POS sequences found in both sets of data representing the distributions in the corpus as a whole (note again the considerations in §7.5.6 in respect of the partial nature of the Ugaritic data). The sequences compared were:

- Noun + Noun
- Noun + Verb
- Noun + Prep
- Noun + Ptcl
- Noun + Adj.
- Prep + Noun
- Verb + Prep
- Verb + Noun
- Ptcl + Noun

7.6.3.1. Quantitative comparison

The values for Association Score B were plotted on two bar charts. Figure 7.1 gives the POS sequences with a *positive* association with Ugaritic univertation, along with the Association Scores of their Tiberian Hebrew counterparts. In each case univertation in Ugaritic is positively associated with a syntagm that has a positive association in Tiberian Hebrew either with conjunctive accentuation (Verb + Noun, Verb + Prep, Noun + Adj.), or with *maqfef* (Prep + Noun), or with both (Noun + Noun).

Figure 7.2 gives the POS sequences with a negative association with Ugaritic univertation, along with the Association Scores of their Tiberian Hebrew counterparts. Every syntagm negatively associated with univertation in Ugaritic is also negatively associated in Tiberian Hebrew with *maqfef* Noun + Ptcl, with conjunctive accentuation (Ptcl + Noun) or with both *maqfef* and conjunctive accentuation (Noun + Verb, Noun + Prep). In three of these cases (Noun + Ptcl, Noun + Prep, Noun + Verb), the syntagm is also positively associated with disjunctive accentuation.

7.6.3.2. Verb + X syntagms

One area where Ugaritic univertation patterns with conjunctive accentuation over against *maqfef* is in Verb + X (Verb + Noun, Verb + Prep) syntagms: these are positively associated with univertation in Ugaritic, and with conjunctive accentuation in Tiberian Hebrew. By contrast, these syntagms are negatively associated with *maqfef*.

Table 7.15: POS sequences joined by conjunctive accent in Tiberian Hebrew (Assoc. Score B, POS prevalence in BH corpus > 0.5%)

POS sequence	Assoc. Score B
Verb + Noun	2.11
Adj. + Adj.	1.9
Ptcl + Verb	1.8
Adj + Noun	1.71
Verb + Ptcl	1.65
Ptcl + Prep	1.6
Verb + Prep	1.6
Noun + Adj.	1.54
Noun + Noun	1.5
Noun + Ptcl	1.36
Pron + Verb	1.25
Adv + Verb	1.13
Conj + Verb	1.11
Verb + Verb	1.06

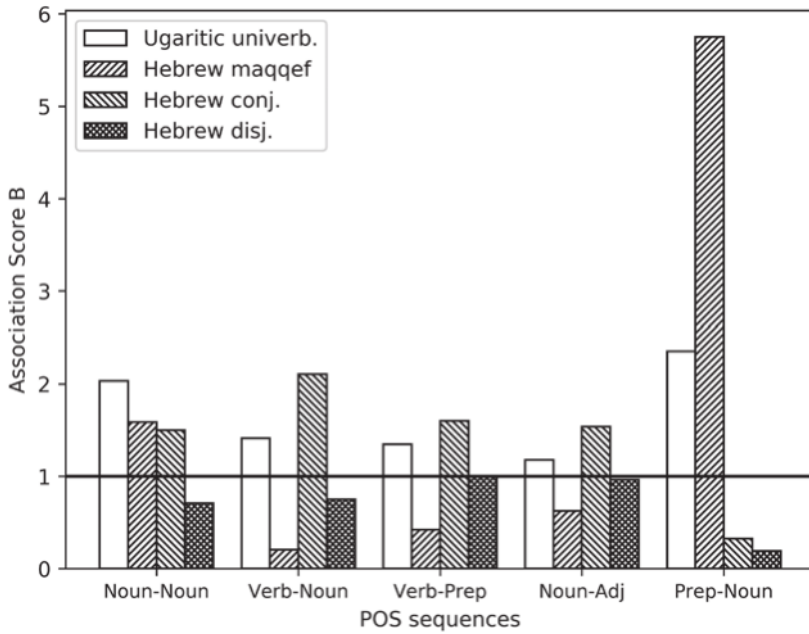


Figure 7.1: Syntagms positively associated with univertation in Ugaritic, along with Tiberian Hebrew counterparts (Association Score B)

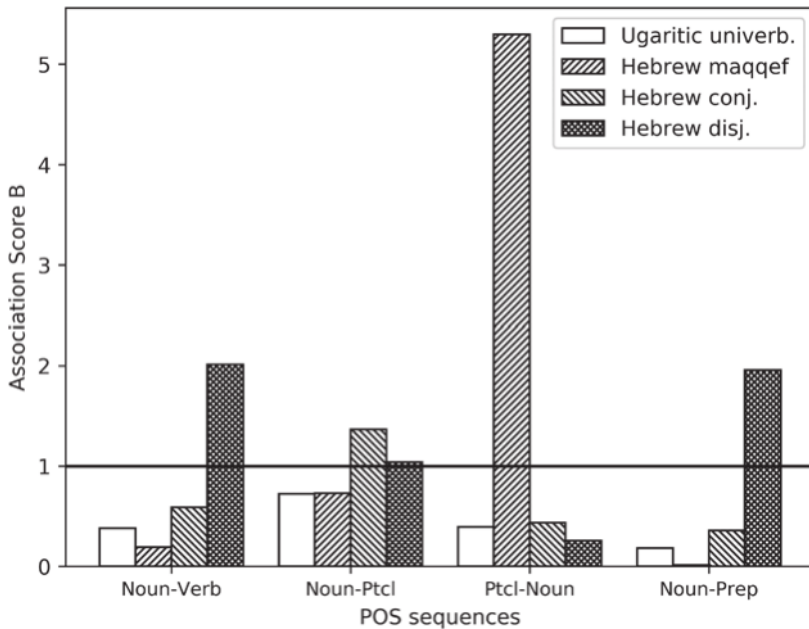


Figure 7.2: Syntagms negatively associated with univertation in Ugaritic, along with Tiberian Hebrew counterparts (Association Score B)

It is worth recalling that a negative association does not mean that the syntagm does not occur at all in the corpus. For example, Verb + Noun sequences joined by *maqfef* are attested in Tiberian Hebrew despite the strongly negative correlation, e.g.:

(239) Gen 2:24

יָצַח אִישׁ אֶת־אָבִיו ←
 (y'zb=yš_{ωφ}) (t=by=w_{ωφ})
 leaves=man OBJ=father-his
 'a man shall leave his father' (KJV)

Similarly, there are examples of Verb + Prep syntagms joined by *maqfef*:

(240) Job 22:24

וְשִׂית־עַל־עָפָר בָּצָר ←
 (w=šyt=ʿl=pr_ω) bšr_{ωφ}
 and=[set_v=[on=dust_{pp}]] [gold_{OBJP}]_{VP}]
 'Then shalt thou lay up gold as dust' (KJV)

The negative association means that the incidence of such syntagms joined by *maqfef* is lower than would be expected based on the occurrence of the syntagm in the corpus as a whole.

The positive association of Verb + X syntagms with conjunctive accentuation, on the other hand, means that not only are such sequences much more frequently joined by conjunctive accents than *maqfef*, but that the association is greater than their frequency in the corpus as a whole would suggest.

(241) Gen 10:24

וְאַרְפַּכְשָׁד יָלַד אֶת־שָׁלַח ←
 (w=rpkšd_{ωφ}) (yld_ω t=šlh_{ωφ})
 and=[PN_{SUBJP}] [begat OBJ=PN_{VP}]
 'And Arphachshad begat Salah' (KJV)

• **Verb + PrepP** cf. (227):

(242) Gen 7:9

בָּאוּ אֵל־נֹחַ ←
 (b^λw_ω) (t=nh_{ωφ})
 [they_came_v] [to=PN_{PP}]_{VP}]
 'they came to Noah'

The fact that Ugaritic patterns with Hebrew conjunctive accentuation over against *maqfef* in Verb + X sequences carries one of two possible implications:

- Ugaritic univertation represents units joined at the level of the prosodic phrase;
- In the Ugaritic language, Verb + X sequences frequently formed single prosodic words, whereas this is not the case in Tiberian Hebrew.

From the present vantage point, it is hard to choose between these two possibilities. However, in Chapter 8 evidence will be provided that implies that points in the direction of the second possibility.

7.6.4. Summary

The foregoing analysis has shown that univertation in Ugaritic has syntagmatic affinities with both *maqfef* and conjunctive accentuation. This is to say that, from the perspective of Tiberian Hebrew, univertation in Ugaritic has affinities with both prosodic wordhood and prosodic phrasehood: some syntagms that are univertated in Ugaritic are often linked at the level of the prosodic phrase in Tiberian Hebrew, while others have a stronger affinity with prosodic word-level association. The distributional location of Ugaritic univertation between *maqfef* and conjunctive accentuation in Tiberian Hebrew can alternatively be visualised using MultiDimensional scaling, to which I turn in the next section (§7.7).

7.7. Visualising morphosyntactic collocation of linking features with MDS

7.7.1. Introduction

The measures described in §7.5 can be used to visualise the morphosyntactic collocation of linking features in Tiberian Hebrew and Ugaritic. Since the relationship between the variables (in this case subcorpora, *e.g.* Genesis, Exodus) is calculated in terms of many dimensions (in this case, POS sequences, *e.g.* Noun–Noun), it is impossible to plot the exact position of each subcorpus. It is therefore necessary to reduce the number of dimensions. A helpful tool for visualising the distributions of multivariate data is MultiDimensional Scaling (MDS).⁵

The data processing steps are outlined at §7.7.1.1. The findings in respect of Tiberian Hebrew and Ugaritic are then presented at §7.7.3.2.

7.7.1.1. Data processing steps for obtaining MDS plots

All data was processed in Python. The data processing steps were as follows:

- **Step 1: Produce a pandas DataFrame (<https://pandas.pydata.org>) *f* with:**
 - **Columns** labelled for POS sequences *w*, *e.g.* Noun–Noun, Verb–Noun, Noun–Verb;

⁵ (For an overview, see *e.g.* Mead (1992); see also https://en.wikipedia.org/wiki/Multidimensional_scaling, accessed 23/08/2021.

- **Rows** labelled for a subcorpora *c* (e.g. Bible book) and a particular linking feature *j* (e.g. disjunctive accent, conjunctive accent, *maqqef*), i.e. Genesis–Disjunctive, Genesis–Conjunctive, Genesis–*Maqqef*, Exodus–Disjunctive etc.
- **Values** Here two measures may be used:
 - ▶ **a) Proportion of Occurrences (Eq. 2)** of a given POS sequence under a particular linking feature in a given subcorpus: e.g. a value of 0.7 for Noun–Noun sequences joined by *maqqef* in Genesis would mean that Noun–Noun sequences represent 70% of POS sequences joined by *maqqef* in Genesis.
 - ▶ **b) Association Score (Eq. 9)** of a given POS sequence (e.g. Noun–Noun) under a particular linking feature (e.g. *maqqef*) in a given subcorpus (e.g. Genesis): The higher the value, the stronger the association between the POS sequence and the linking feature in the subcorpus; a value above 1 for a given POS sequence under a given linking feature in a given subcorpus means that the Proportion of Occurrences of that POS sequence under the linking feature in the subcorpus is greater than the proportion of instances of the POS sequence occurring in the population as a whole.
- **Step 2: Produce a Scaled DataFrame $s(f)$** from the values obtained in Step 1: The values in a given row *p* of DataFrame *f* are divided by the maximum value in the row, corresponding to the highest Proportion of Occurrence/Association Score of any POS sequence under a given linking feature in a subcorpus, to obtain a value between 0 and 1. A value of 1 for a POS sequence *w* joined by a linking feature *j* in subcorpus *c*, $w_{j,c}$ means that that the sequence *w* has the highest Proportion of Occurrence/Association Score under linking feature *j* in subcorpus *c*.
- **Step 3: Produce a distance matrix $d(s(f))$** giving the Euclidean distance between each subcorpus for each linking feature in terms of the scaled values produced in Step 2. Because the distances are calculated on a Scaled DataFrame, the comparison is relative to the maximum Proportion of Occurrence/Association Score of any POS sequence in a given subcorpus. Euclidean distances were calculated according to the following formula, for given rows of $d(s(f))$ *p* and *q*:⁶

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_i - q_i)^2 + \dots + (p_n - q_n)^2} \quad \text{Eq. 12}$$

- **Step 4: Produce a 2D MultiDimensional Scaling (MDS) plot** of the distance matrix $d(s(f))$. MDS plots were produced using the *EcoPy* package (<https://ecopy.readthedocs.io/en/latest/>), and plotted with the *Matplotlib* library (<https://matplotlib.org/>). The method:⁷

⁶ See https://en.wikipedia.org/wiki/Euclidean_distance, accessed 23/08/2021.

⁷ <https://ecopy.readthedocs.io/en/latest/ordination.html>, accessed 23/08/2021.

Takes a square-symmetric distance matrix with no negative values as input. After finding the solution that provide the lowest stress, *ecopy.MDS* scales the fitted distances to have a maximum equal to the maximum observed distance. Afterwards, it uses PCA to rotate the object (site) scores so that variance is maximized along the x-axis.

The method takes a transform parameter. For the MDS given in the present study the value for this parameter was absolute. With this parameter, the method ‘Conducts absolute MDS. Distances between points in ordination space should be as close as possible to observed distances’ (<https://ecopy.readthedocs.io/en/latest/ordination.html>, accessed 23/08/2021).

As already mentioned (§7.7.1), by its nature MDS involves the reduction in the number of dimensions of multivariate data. In representing a multi-dimensional reality in two dimensions there will always be a number of possibilities. How good a representation of the multi-dimensional original a given 2D representation may be expressed in terms of ‘stress’. In general, a stress value of 0.2 or above is regarded as suboptimal; if the stress figure is above this threshold, the plot should ideally be redone in higher dimensions.⁸

7.7.2. Overview

In this section the morphosyntactic distributions of Ugaritic graphematic univertation and Tiberian Hebrew accentuation are compared directly using Association Score B (§7.5.7). The variable is the POS type that comes before and after the small vertical wedge, or lack of it, *e.g.* Noun + Noun, or Verb + Noun etc. Recall that a figure above 1 indicates that univertation is positively associated with the combination, and a figure below 1 indicates that univertation is negatively associated with the combination. Owing to constraints of time, the Ugaritic dataset used for this part of the investigation was restricted to the Ba¹ epic (KTU 1–6).

7.7.3. Proportion of Occurrences

7.7.3.1. *Maqqef* and conjunctive accents have distinctive distributions in Tiberian Hebrew

In terms of their morphosyntactic collocations, the distributions of the two Tiberian Hebrew accent types under consideration are distinct. This can be visualised in the Multidimensional Scaling (MDS) scatter plot at Figure 7.3.

7.7.3.2. Comparing Ugaritic and Tiberian Hebrew

We are now in a position to compare the distributions of Ugaritic univertation and Hebrew accentuation. A MDS plot is given at Figure 7.4.

The reader will note that the population of Ugaritic POS sequences, irrespective of univertation, is located among the Hebrew conjunctive accents, albeit closer to the Hebrew population POS sequences than Ugaritic univertation. This is a matter

⁸ See <http://environmentalcomputing.net/multidimensional-scaling/>, accessed 23/08/2021.

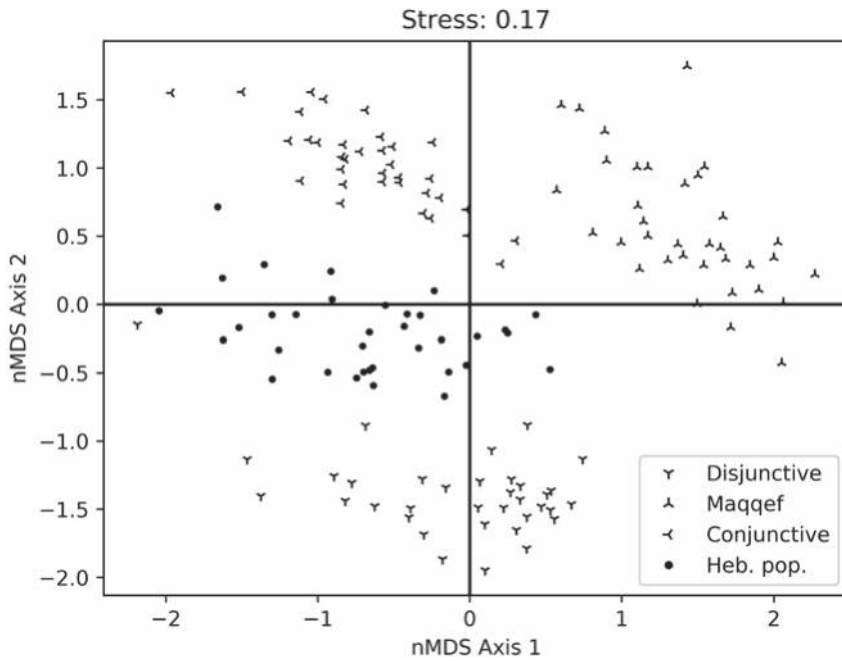


Figure 7.3: MDS plot of Hebrew accent distributions (Proportion of Occurrences)

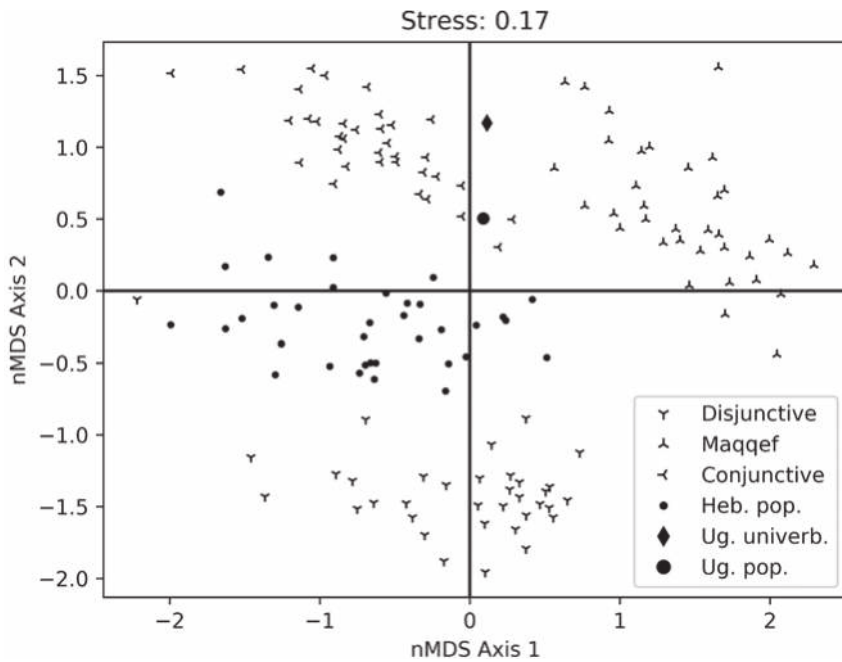


Figure 7.4: MDS plot of Hebrew accent distributions and Ugaritic univertation (Proportion of Occurrences)

of interest in itself, and deserves its own investigation. (It is not particularly surprising that the population of POS sequences in different, albeit related, languages do not fully overlap.)

It is, however, the position of Ugaritic univertation vis-à-vis Hebrew *maqfef* and conjunctive accentuation that is of primary interest for present purposes. From the MDS plot at Figure 7.4 it can be seen that the distribution of univertation in Ugaritic is somewhere between that of *maqfef* and conjunctive accents in Tiberian Hebrew, although perhaps closer to the latter than the former. This shows in a different way the same result that we found at §7.6.

7.7.4. Association Score B

At §7.5.7 I noted that the Proportion of Occurrence measure takes no account in itself of the rate of occurrence of a given syntagm in the corpus as a whole. In the analysis at §7.7.3 this may be taken into account visually by the fact that the population distribution is plotted on the same MDS. However, in Association Score B (§7.5.7) we have a measure that does take this factor into account. Plotting the distributions of the Hebrew accents and Ugaritic univertation in these terms gives the MDS at Figure 7.5. From this the same result is obtained, namely, that Ugaritic univertation patterns between Hebrew *maqfef* and conjunctive accentuation. (The reader should

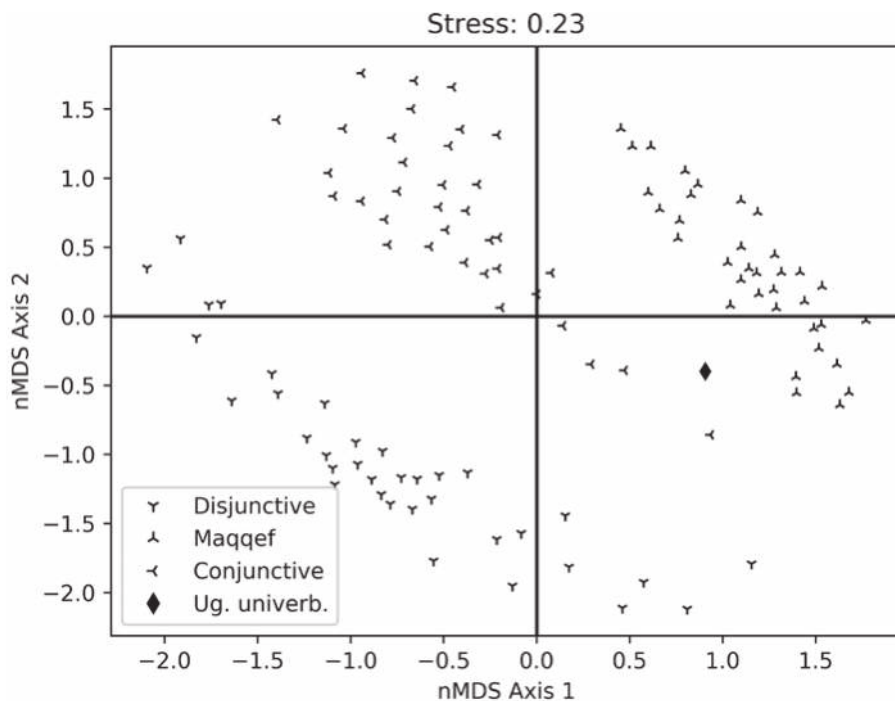


Figure 7.5: MDS plot of Hebrew accent distributions and Ugaritic univertation (Association Score B)

note, however, that in this case the stress figure is greater than 0.2, cf. §7.7.1.1. Accordingly, ideally the plot should be redone in 3D.)

7.8. Conclusion

In this chapter quantitative data concerning the distribution of word division and univertation in Ugaritic and Hebrew accentuation, principally *maqṣef* and conjunctive accents, have been compared, in order to establish whether univertation in Ugaritic shows a distribution closer to that of prosodic wordhood or that of prosodic phrasehood.

In §7.3 we saw that univertation in Ugaritic occurs much less frequently than either *maqṣef* or conjunctive phrasing in Tiberian Hebrew. Then in §7.4 phrase lengths of both *maqṣef* and conjunctive phrases in Tiberian Hebrew were compared with univertated phrases in Ugaritic, where the distribution of the latter was found to be closer to that of *maqṣef* than that of conjunctive phrases. Finally, in §7.5–§7.7 I found that the collocational data shows a clear alignment of Ugaritic univertation and prosodic units. In some syntagms Ugaritic univertation was found to pattern with *maqṣef*, notably Noun + Noun and Prep + Noun sequences, whilst Verb + X syntagms Ugaritic univertation patterns with Hebrew conjunctive accentuation.

In sum, then, while overall the distribution of univertation favours a closer relationship with *maqṣef* than with conjunctive phrases in Tiberian Hebrew, there is also evidence that points in the other direction. To answer the question of the linguistic level targeted by univertation in Ugaritic, therefore, it will be necessary to bring syntactic evidence to the fore. This will be done in the next chapter.

Chapter 8

Semantics of word division and univertation in the ‘Majority’ orthography: prosodic word or prosodic phrase?

8.1. Introduction

In the foregoing chapters we have observed that:

- Graphematic univertation in Ugaritic alphabetic cuneiform targets linguistically salient units. This is to say that the distribution is not random, although there is evidence of scribal error. However, the unit so demarcated is not necessarily a full syntactic phrase, such as a whole NP or VP;
- The kinds of unit targeted by graphematic univertation can be paralleled by *maqṣef* and conjunctive phrases in Tiberian Hebrew.

The fact that the distribution of graphematically univertated sequences in Ugaritic alphabetic cuneiform parallels that of both conjunctive and *maqṣef* phrases in Tiberian Hebrew is strongly suggestive of these sequences in Ugaritic representing ACTUAL PROSODIC PHRASES or ACTUAL PROSODIC WORDS. It is worth addressing directly, then, the question of the level of prosody to which Ugaritic univertation pertains, whether to the prosodic word or the prosodic phrase. The written Ugaritic material, unfortunately, does not provide direct access to information on accent, or indeed to any sandhi phenomena that may or may not have existed in Ugaritic. However, it is possible to address this question indirectly. Already at §5.9 we suggested that a principle of word division according to prosodic words would help account for the differential treatment of function morphemes of varying graphematic/prosodic weights (cf. §1.7.3.1). However, in light of the use of a prosodic phrase level word division strategy in Phoenician (Chapter 4), we should seek to identify explicitly whether word division in Ugaritic targets prosodic words or prosodic phrases.

This can be done by considering the syntactic distribution of graphematic words. Across languages alignment of the prosodic structure with that of syntax occurs at the level of the prosodic phrase, rather than that of the prosodic word (cf. §1.5.1; Truckenbrodt 2007, 435–437; §1.5). This means that, if word division in Ugaritic targets the prosodic phrase, we would expect to see alignment between syntactic structure and graphematic word divisions, whereas if it targets the prosodic word, we would expect to see syntactic boundaries to be ignored by the graphematic structure.

‘He made his way to the cave of ’El’ (for trans. cf. del Olmo Lete & Sanmartín 2015, 296)

(246) KTU³ 1.5:II:11

→ 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧

*bh*_{t<ω>} *l*= *bn*_{<ω>} *ilm=mt*_{<λ>}

[hail [INTERJ= [son DN_{NP}]=[DN_{VOC}]_{NP-VP}]

‘Hail, O son of ’El, Môt’ (for interpretation cf. del Olmo Lete & Sanmartín 2015, 482; Pardee 2003, 266)

(247) KTU³ 1.14:II:21

→ 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧

*w=ʔ*_{<ω>} *l=zr*_{<ω>} *mḡdl*_{<ω>}

and=go_up to=top tower

‘and go up/he went up to the top of the tower’

Finally, there is at least one example of a VP bisected by word division:

(248) KTU³ 1.1:III:11

→ 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧

*m-y=twth*_{<ω>} *iš*_{[d-k<ω>}

towards=me=let_hasten steps-your

‘Towards me let your steps hasten’ (trans. del Olmo Lete & Sanmartín 2015, 929)

Note that in this case, not only is the VP bisected, but the PP is written together with the verb: if word division reflected prosodic phrases, we might expect to find a small vertical wedge at the right edge of the maximal projection of the prepositional phrase, *i.e.* just before the verb. This is indeed the more commonly found configuration, *e.g.*:²

(249) KTU³ 1.2:I:24

→ 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧

*b=hm*_{<ω>} *ygʿr=bʔ*_{<ω>}

on=them reproach.PREF=DN

‘**Baʔ reproached** them’ (trans. after del Olmo Lete & Sanmartín 2015, 287)

² Cf. also KTU 1.2:IV:3: 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 𐎧𐎢𐎠𐎧 *w=bym*_{<ω>} *mnḥ=l=ābd* ‘and in DN calm was not lacking’.

It is also the more commonly found configuration in Tiberian Hebrew, e.g.:³

(250) Deut 14:2

וַיִּבְחַר יְהוָה ←

(w=b-k_{ω φ}) (bħr_ω yhwħ_{ω φ})

and=[[in-you.SG_{PP}] chose_{VP}] [DN_{SUBJP}]

‘and DN has chosen you’

There are, however, two parallels in Tiberian Hebrew of the Ugaritic (248), one of which is found at Ezek 45.3:⁴

(251) Ezek 45:3

וּבֵית־יְהוָה הַמִּקְדָּשׁ ←

(w=b=w=yhyh_ω h-mqdš_{ω φ})

and=[[in=it_{PP}]=shall_be_{VP}] [the-holy_place_{DP}]

‘and in it shall be the holy place’

Examples of this kind can be explained if one assumes bottom-up prosodic phrase construction, whereby once a prosodic word has been formed, prosodic phrasing is blind to any syntactic boundaries occurring within the prosodic word.

It is perhaps in this light that other examples of univerbation across syntactic phrase boundaries should be seen, such as the following, first mentioned above (234) (§6.6):

(252) KTU³ 1.2:IV:11

→ ⚡ ⚡ ⚡ || ⚡ ⚡ ⚡ ⚡ ⚡ ⚡ ⚡ ⚡ ⚡ ⚡ ⚡ ⚡

ktr=šmdm_{⟨ω⟩} ynħt_{⟨ω⟩}

[DN_{SUBJP}]=[[double_mace_{NP}] brought_{VP}]

‘Ktr brought a double mace’ (trans. per del Olmo Lete & Sanmartín 2015, 620)

Although I have not been able to find a direct Tiberian Hebrew parallel of this example, the univerbation is compatible with prosodic wordhood in principle, if prosodic words are built before prosodic phrases.

On the basis of the syntactic evidence of graphematic univerbation, therefore, it is possible to conclude that the Ugaritic graphematic word targets the level of the

³ Parallels: Deut 24:13; Isa 14:11, 24:17, 60:2; Ezek 10:13; Jer 48:43; 2Sam 15:4; 1Kgs 11:2; 1Chr 18:8.

⁴ The other is at 1Sam 9:2: וַיְהִי בְנֵי וַיִּבְחַר יְהוָה w=lw=hyh bn ‘And he had a son’ (KJV). There are also at least three examples of a preposed PP joined to a following verb with a conjunctive accent, at Lev 13:28, Psa 31:15 and Psa 68:30.

prosodic word rather than that of the prosodic phrase. This result is not only consistent with the quantitative analysis of the morphosyntactic context conducted in §7.5, but it is also consistent with the basic patterns of word division in Ugaritic (§5.3), which are for the most part isomorphic with the word division practice of the consonantal text of Tiberian Hebrew.

Further support for this proposal will come in Part III, where I conclude that graphematic words in the Tiberian Hebrew consonantal text represent MINIMAL PROSODIC WORDS. This is to say that the ORL of word division in both cases is the same, namely, the prosodic word. Where Ugaritic differs from consonantal Tiberian Hebrew is in the fact that the graphematic word corresponds, in an important minority of instances, with ACTUAL PROSODIC WORDS in Tiberian Hebrew, viz. units joined with *maqfef*.

8.3. Consistency of the representation of ACTUAL PROSODIC WORDHOOD in Ugaritic

It remains to ask how consistently the Ugaritian scribes represent ACTUAL PROSODIC WORDS, and whether also in Ugaritic, as in later Northwest Semitic, there was a tendency to abstract away from ACTUAL PROSODIC WORDS, so as to represent MINIMAL PROSODIC WORDS. In support of such a view we could note that univertation in Ugaritic is much less frequent than the representation of prosodic words by *maqfef* in Tiberian Hebrew (§7.3). It is also possible to find minimal pairs and near-minimal pairs of univertated sequences in Ugaritic where word division follows the basic principles, without any univertation (§5.8). The same is true of Tiberian Hebrew, where near identical syntagms can have different prosodic representations. Compare the following:

(253) Deut 33:7

שְׁמַע יְהוָה קוֹל יְהוּדָה ←
 (*šm*'_ω *yhwh*_{ωφ}) (*qwl*_ω *yhwdh*_{ωφ})
 [hear.IMPV_v [DN_{voc}]] [voice Judah_{NP}]_{VP}]
 'Hear, O LORD, the voice of Judah' (KJV)

(254) Psa 27:7

שְׁמַע-יְהוָה קוֹלִי אֶקְרָא ←
 (*šm*'=*yhwh*_ω *qwl*=*y*_ω *'qr'*_{ωφ})
 [hear.IMPV_v]=[DN_{voc}]] [voice-my I_{cry}]_{VP}]
 'Hear, O LORD, when I cry with my voice' (KJV)

Without a further source of evidence on Ugaritic prosody, it is impossible to say for certain whether Ugaritian scribes represent ACTUAL PROSODIC WORDS consistently.

However, the existence in Tiberian Hebrew of examples such as (253) and (254) shows that the fact that univerbation is apparently inconsistent is not in itself a reason to dismiss the possibility, and may in fact be a reason to endorse it.

8.4. Univerbation at clause boundaries

The fly in the ointment to the proposal that univerbation represents prosodic words in Ugaritic is the occasional tendency for the phenomenon to occur at clause boundaries. At (235) §6.6 I gave the following example:

(255) KTU³ 1.3:III:14–15

→ <=> >> >> >> >> >> >> >> >> >> >>
 >> >> >> >> >> >> >> >> >> >> >> >>

*qryy*_{<ω>} *b=ārs*_{<λ>} *m̄lhmt=št*_{<ω>} *b=‘prm*_{<ω>} *ddym*_{<λ>}
 [meet.IMP in=land war_s]=[put in=steppe harmony_s]

‘Meet war in the land, put harmony in the steppe’ (trans. del Olmo Lete & Sanmartín 2015, 264, 704)

Such univerbations can also include conjunctions joining clauses:⁵

(256) KTU³ 1.4:III:31–33

→ >> >> >> >> >> >> >> >> >> >> >>
 >> >> >> >> >> >> >> >> >> >> >>
 >> >> >> >> >> >> >> >> >> >> >>

*hm*_{<ω>} *ǧztm*_{<λ>} *bny*_{<ω>} *bnwt=w=t‘n*_{<λ>}
 PTCL win_over creator creatures=and=answered

*b̄tlt*_{<ω>} ‘nt’_{<ω>}
 virgin DN

‘Have you won over the creator of creatures? And the virgin ‘Anat answered ...’ (trans. del Olmo Lete & Sanmartín 2015, 169, 325)

We should note, however, that a number of examples occur between the cola of bi- or tricola. Thus in the previous example univerbation occurs at the boundary of two cola, while in the following example it occurs at the boundary of the second and third cola:⁶

⁵ The following examples were found: KTU 1.2:I:15–16; 1.2:IV:12; 1.3:II:7; 1.3:III:15; 1.4:IV:36; 1.5:II:13; 1.14:VI:22; 1.16:IV:3.

⁶ Cf. the parallels at KTU 1.3:IV:4 (second in a tricolon) and KTU 1.4:IV:36 (second in a tricolon).

(yšplw_{ω φ})be_low_s]

‘Let every valley be lifted up, and every mountain and hill be made low’ (trans. NAS)

If the examples are real,⁸ given the complete lack of parallels in Tiberian Hebrew, they must reflect a feature of Ugaritic verse without parallel in the Biblical tradition. This, of course, is not problematic: just because so many features of Ugaritic verse have parallels in Biblical poetry does not mean that there should be a one-to-one mapping between the two.

While there are no parallels on the Tiberian Hebrew side, there are, of course, parallels from later alphabetic inscriptions. At §3.4.6 we showed that univertation across clause boundaries in KAI 24 could be exploited for metrical effect. Note too the phenomenon of enjambement in Greek epic poetry. Greek epic verse is metrical, with the length of lines governed by the rules of the hexameter verse. However, it frequently happens that the syntax is not commensurate with the rhythm of the verse, so that the elements of a clause can ‘spill over’ on to the next line, as in the first two lines of the *Iliad*:⁹

(260) *Iliad* 1.1–2 (text Munro & Allen 1920)

→ μῆνιν ἄειδε θεὰ Πηληϊάδεω Ἀχιλῆος
οὐλομένην, ἣ μυρῖ Ἄχαιοις ἄλγε’ ἔθηκε

<i>mênin</i>	<i>Áeide</i>	<i>theà</i>	<i>Pēlēiádeō</i>
wrath.	sing.	goddess.	son_of_Peleus.
ACC	IMP	VOC	GEN

<i>Akhilêos</i>		<i>ouloménēn</i>	<i>hē</i>	<i>muri'</i>
Achilles.		ruined.ACC	REL	ten_thousand.ACC
GEN				

<i>Akhaiōis</i>	<i>álge'</i>	<i>éthēke</i>
Achaeans.DAT	sufferings.ACC	placed

‘Of the wrath of Achilles, son of Peleus, the wretched, sing, goddess, that brought myriad sufferings upon the Achaeans ...’

In this example, *ouloménēn* belongs syntactically with the elements of line 1, yet appears at the start of line 2. Assuming that the line end induces pause in the prosody,

⁸ Since it could of course be that these instances are errors on the part of the writer. Such a possibility is, however, not only a counsel of despair, but also difficult to assess.

⁹ On the effect of mismatches between verse and syntactic structure, see Devine & Stephens (1994, 410).

this arrangement induces a misalignment between prosody and syntax, whereby the intra-clausal pause after Ἀχιλῆος *Akhilēos*, but before οὐλομένην *ouloménēn* is greater than the inter-clausal pause between οὐλομένην *ouloménēn* and ἦ *hē*. In terms of prosody, therefore, οὐλομένην *ouloménēn* and ἦ *hē* are prosodically linked in a way that does not align with the syntax. This prosodic linking would have many of the hallmarks of prosodic wordhood in terms of prosodic connection. Consequently, in a system where graphematic word division corresponded to prosodic wordhood, one might choose to indicate this relationship by univertation.

It should be reiterated that scholarly consensus is that Ugaritic poetry does not have a metrical basis, or at least, that such cannot be proven on the basis of the evidence of the Ugaritic texts as we have them (cf. §5.2.4; Horwitz 1971, 89; Pardee 1981, 115; Wansbrough 1983, 221–222). Yet it seems likely that word division practices have important implications for the analysis of Ugaritic poetry in the future.

8.5. Adoption of the ‘Majority’ orthography outside of literary contexts

So far we have considered the ‘Majority’ orthography as it is manifested in literary tablets. The reason for this is that these tablets contain a (for Ugaritic) large corpus of similar material, from which it is possible to make an assessment of the orthography. However, we have noted, especially in the previous section, the possible presence of univertation in a context specific to verse. Before concluding this chapter, therefore, it is worth asking to what extent the features of the orthography we have observed for literary compositions also hold for non-literary works.

At §5.6 we observed that the orthography is by no means limited to literary compositions, and can be found in a wide range of other text types. The next example gives the tablet 3.12 in full to illustrate the use of the orthography in a legal/administrative context:

(261) KTU³ 3.12

→ 1 IIII , IIIIII IIII II 𐎗 𐎗 𐎗 IIII 𐎗
 2 IIII IIII 𐎗 𐎗 𐎗 𐎗 𐎗 𐎗 𐎗 𐎗 𐎗 𐎗
 3 IIII IIII 𐎗 𐎗 𐎗 𐎗 𐎗 𐎗 𐎗 𐎗
 4 IIII 𐎗 𐎗 𐎗 𐎗 𐎗 𐎗 𐎗 IIII 𐎗
 5 IIII IIII 𐎗 𐎗 𐎗 𐎗 𐎗 𐎗 𐎗
 6 𐎗 IIII 𐎗 𐎗 𐎗 𐎗 𐎗 𐎗 𐎗 IIII
 7 𐎗 IIII 𐎗 IIII 𐎗 𐎗 𐎗
 8 𐎗 𐎗 𐎗 𐎗 IIII 𐎗 𐎗 IIII 𐎗 𐎗 IIII 𐎗
 9 𐎗 𐎗 IIII 𐎗 𐎗 IIII 𐎗 IIII 𐎗 IIII
 10 IIII 𐎗 IIII 𐎗 𐎗 𐎗 𐎗 𐎗 𐎗 𐎗 IIII 𐎗
 11 𐎗 IIII IIII 𐎗 IIII 𐎗 IIII
 12 𐎗 𐎗 𐎗 𐎗 𐎗 𐎗 𐎗 IIII 𐎗 𐎗

13

14

15

¹ $l_{\langle\omega\rangle}$ $yihd=\dot{\mathfrak{s}}\mathfrak{t}q\mathfrak{s}lm_{\langle\lambda\rangle}$ ² $b=\dot{u}nt_{\langle\omega\rangle}$ $km_{\langle\omega\rangle}$ $\mathfrak{s}p\mathfrak{s}_{\langle\lambda\rangle}$
not recruit.PASS=PN for=service like sun

³ $d=brt_{\langle\omega\rangle}$ $kmt_{\langle\omega\rangle \langle\lambda\rangle}$ ⁴ $br_{\langle\omega\rangle}$ $\mathfrak{s}tq\mathfrak{s}lm_{\langle\lambda\rangle}$
which=is_free so is_free PN

⁵ $b=\dot{u}nt_{\langle\omega\rangle}$ $\mathfrak{d}=\mathfrak{c}lm_{\langle\lambda\rangle}$ ⁶ $mi\mathfrak{s}mn_{\langle\omega\rangle}$ $nqmd_{\langle\lambda\rangle}$
from=service for=ever seal PN

⁷ $mlk=\dot{u}grt_{\langle\lambda\rangle}$ ⁸ $nqmd_{\langle\omega\rangle}$ $mlk_{\langle\omega\rangle}$ $\dot{u}grt_{\langle\lambda\rangle}$ ⁹ $ktb_{\langle\omega\rangle}$
king=TN PN king TN wrote

$spr=hnd_{\langle\lambda\rangle}$ ¹⁰ $d=tbrrt_{\langle\omega\rangle}$ $\mathfrak{s}tq\mathfrak{s}lm_{\langle\lambda\rangle}$ ¹¹ $\mathfrak{b}d=h_{\langle\omega\rangle}$
decree=this of=exemption PN servant-his

$hnd_{\langle\lambda\rangle}$ ¹² $w=mnkm_{\langle\omega\rangle}$ $l=yqh_{\langle\lambda\rangle}$ ¹³ $spr_{\langle\omega\rangle}$ $mlk_{\langle\omega\rangle}$
this and=no_one not=take decree king

$hnd_{\langle\lambda\rangle}$ ¹⁴ $b=yd_{\langle\omega\rangle}$ $\mathfrak{s}tq\mathfrak{s}lm_{\langle\lambda\rangle}$ ¹⁵ $\mathfrak{d}=\mathfrak{c}lm_{\langle\omega\rangle}$
this from=hand PN for=ever

' $\mathfrak{S}tq\mathfrak{s}lm$ is not recruited for service. Like the sun which is free, so is $\mathfrak{S}tq\mathfrak{s}lm$ free from service in perpetuity. Seal of Niqmadu king of Ugarit. Niqmadu king of Ugarit wrote this decree of exemption (for) $\mathfrak{S}tq\mathfrak{s}lm$ his servant. And let no one take this royal decree from the hand of $\mathfrak{S}tq\mathfrak{s}lm$ in perpetuity.'

Many of the features we have observed in the literary compositions can be seen in this tablet, including:

- Monoconsonantal suffix particles are written together with the preceding morpheme. Thus, $\mathfrak{b}d-h$ (11);
- Monoconsonantal prefix particles are generally written together with the following morpheme, but not absolutely always. The exception in this text is l (1);¹⁰
- Morphemes consisting of two or more consonants may be univerted with one another, e.g. $yihd=\mathfrak{s}tq\mathfrak{s}lm$ (1), $\mathfrak{d}=\mathfrak{c}lm$ (5, 15), $mlk=\dot{u}grt$ (7), $spr=hnd$ (9);
- There is no requirement for a given sequence to be written consistently as a single graphematic word. Thus, $mlk=\dot{u}grt$ (7) vs. $mlk_{\langle\omega\rangle} \dot{u}grt$ (8); cf. $spr=hnd$ (9) vs. $\mathfrak{b}d-h_{\langle\omega\rangle} hnd_{\langle\lambda\rangle}$ (11) and $mlk_{\langle\omega\rangle} hnd_{\langle\lambda\rangle}$ (13);

¹⁰ For interpretation, cf. del Olmo Lete & Sanmartín (2015, 37).

- Where univerbation occurs, there is a tendency for it to happen in Verb + X sequences, thus $l_{\langle\omega\rangle} y\dot{h}d=\dot{s}tq\dot{s}lm$ 'PN is not recruited' (1), in Prep + Noun sequences, thus $d=lm$ 'forever' (5, 15) and in Noun + Noun construct chains, thus $mlk=\dot{u}grt$ 'king of TN' (7);
- Word division may separate constituents of XP^{MAX} , for example within NPs, e.g. $mlk_{\langle\omega\rangle} \dot{u}grt_{\langle\lambda\rangle}$ 'king of Ugarit' (8) and $spr_{\langle\omega\rangle} mlk_{\langle\omega\rangle} hnd_{\langle\lambda\rangle}$ 'this royal decree' (13).¹¹

The overall pattern of word division and univerbation is therefore consistent with that which we see in the literary texts. Furthermore, since there is no necessary right alignment with XP^{MAX} , the domain of word division/univerbation is again consistent with ACTUAL PROSODIC WORDHOOD.

¹¹ For interpretation, cf. del Olmo Lete & Sanmartín (2015, 339).

Chapter 9

Separation of prefix clitics

9.1. Introduction

At §5.3 above, I noted that one of the strongest tendencies in the Ugaritic orthography of word division is that monoconsonantal prefixes are unverbated with the following morpheme(s), *e.g.*:

(262) KTU³ 1.2:IV:12

→       

ygrš_{<w>} grš=ym=grš=ym_{<w>}

l=ksi-h_{<λ>}

DN₁.VOC drive_away.PTPL=DN₂=drive_away.IMP=DN₂

from=throne-his

‘Ygrš, who drives away Yam, drive away Yam from his throne’

In a number of instances in the Ugaritic alphabetic cuneiform corpus, however, we do find a word divider placed after these monoconsonantal prefixes, *e.g.*:

(263) KTU³ 1.1:III:4

→    

w_{<w>} rgm_{<w>} l=kt[r_{<w>}

and say.IMP to=DN

‘And say to DN...’

This phenomenon occurs in a number of contexts:

- Occasionally in literary texts, per (263)
- Occasionally in non-literary texts adopting the ‘Majority’ orthography;
- As a rule in non-literary texts adopting the ‘Minority’ orthography.

In the sections that follow each of these contexts is discussed in turn.

9.2. Literary texts

9.2.1. Complete graphematic separation

Table 9.1 gives the distribution of word division and word non-division after the three prepositional and two clausal monoconsonantal prepositions occurring with a word divider both before and after.¹ From the table the following observations may be made:

- The incidence of a postpositive word divider is low: at most 7.5% in the case of **III** *d*-;
- Clausal particles are considerably more likely to be followed by a word divider than prepositional particles.

Table 9.2 shows that the monoconsonantal prepositions **II** *b*-, **III** *l*- and **▶** *k*- are considerably less likely to be followed by a word divider than the clausal particles **▶** *w*- and **III** *d*-: while approximately 5% of tokens of the latter are followed by word division, only 1% of preposition tokens behave in this way.

Table 9.1: Postpositive word division and non-division in monoconsonantal particles (KTU 1-23)

	<i>l</i>		<i>w</i>		<i>k</i>		<i>b</i>		<i>d</i>	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
No word division after	398	99.0	413	94.94	108	97.3	364	98.64	55	91.67
Word division after	4	1.0	22	5.06	3	2.7	5	1.36	5	8.33
Total	402		435		111		369		60	

9.2.2. Prefix particle chains

Word division is found after prepositional particles not only when the prefix is a completely separate graphematic word, but also when it is part of a clitic chain with a clausal particle. It is often the case that a prepositional particle immediately follows a clausal particle, *e.g.*:

Table 9.2: Distribution of postpositive word division and non-division of monoconsonantal prefix particles grouped by syntactic type (KTU 1-23)

	Clausal		Prepositional	
	Freq.	%	Freq.	%
No word division after	488	94.76	762	98.83
Word division after	27	5.24	9	1.17
Total	515		771	

¹ For this calculation, line division was taken to be equivalent to the small vertical wedge. The corpus of texts included was KTU 1-23, per UDB (Cunchillos, Vita & Zamora 2003).

(264) KTU³ 1.2:IV:5

→ III → II , III = III , III III → III , → III , < = → , < = < → III

l=āṣ_{<w>} ypl_{<w>} ūl-n(-)y_{<w>} w=l ṣpr
 to=ground fell military_forces=our/my and=to dust

ṣm-ḥ-y

strength=our/my

‘our/my forces fell to the ground, to the dust our/my strength’ (trans. with ref. to del Olmo Lete & Sanmartín 2015, 50, 171)

Table 9.3: Word division after III l- according to whether or not it is preceded by → w- or III d- (KTU 1-23)

	No PTCL before		w or d before	
	Freq.	%	Freq.	%
No postpos. word divider	398	99.0	10	45.45
Postpos. word divider	4	1.0	12	54.55
Total	402		22	

Table 9.4: Word division after II b- according to whether or not it is preceded by → w- or III d- (KTU 1-23)

	No PTCL before		w or d before	
	Freq.	%	Freq.	%
No postpos. word divider	364	98.64	8	80.0
Postpos. word divider	5	1.36	2	20.0
Total	369		10	

It is reasonable to suppose that the presence of a clausal clitic immediately prior to a prepositional clitic might have an effect on the presence of a word divider after it. Table 9.3 shows that III l- is much more likely to be followed by a word divider if it is preceded by → w- or III d-. Table 9.4 shows a similar effect in the case of II b-.

One final fact is worth highlighting: there are no instances in KTU 1-23 of prefix particle chains where word division occurs between the two particles. This is not only significant for understanding the nature of word division in the ‘Majority’ orthography, but also

stands in marked contrast to word division in the Ugaritic ‘Minority’ orthography, where particles occurring in chains are separated from one another (§9.4.2.1).

Earlier I pointed out that one of the objections raised against seeing a relationship between graphematic word demarcation and word stress was the fact that chains of clitics may be graphematically demarcated (§5.2). However, the evidence presented here for the treatment of clitic chains in the ‘Majority’ orthography suggests that prefix clitic chains could project prosodic words in their own right in Ugaritic. That two or more clitics might together project a prosodic word is paralleled cross-linguistically, not least in Ancient Greek. There we find that

prepositives and prepositive chains may be written as independent graphematic words without any lexical host (§13.5.2; for further inscriptional examples see Devine & Stephens 1994, 329). At the prosodic level there is evidence that combinations of clitics in Greek can form rhythmically autonomous units (Devine & Stephens 1994, 219–223). I therefore see no reason why the graphematic demarcation of clitic chains should be used as an argument against the prosodic wordhood of graphematic words in the Ugaritic ‘Majority’ orthography.

9.2.3. Accounting for the graphematic separation of prefix particles

While relatively rare in literary texts, the graphematic separation of prefix particles occurs frequently enough, especially in the case of $\text{P}\rightarrow w$ - and $\text{M}\rightarrow d$ -, that it deserves an explanation beyond the postulation of scribal error.

At least in the case of clitic chains followed by word division, a prosodic explanation seems reasonable. At §1.4.2.1 I showed that it is possible for two clitics to comprise a single graphematic word. In the Greek case discussed there, we provided the example of a proclitic and an enclitic forming a single prosodic word. If the clitic chains discussed at §9.2.3 here represent prosodic words, this could be interpreted by postulating that the proclitic $\text{P}\rightarrow w$ - provides the following prepositional particle with an accent, enabling the two to stand as a single prosodic word.

The case of monoconsonantal particles standing as independent graphematic words is trickier to account for. If these are monomoraic, we would not expect them to be capable of standing as independent prosodic words (§1.4.2.4); if word division represents the demarcation of prosodic words, we would not expect to find these particles standing as independent graphematic words. The verse form may be partly responsible. It is well known that in Greek epics certain short vowels are artificially lengthened so as to fit the scansion of the verse (see *e.g.* Hoekstra 1978). The explanation could also be related to contrastive focus. As we will see in the next section, there is a tantalising example from a non-literary text where a lone graphematically independent $\text{P}\rightarrow w$ - appears to be associated with change of topic and/or subject (cf. also §9.4.4.3). Further work is needed before it is possible to be certain. From the evidence presented here, however, the graphic separation of prefix clitics is too frequent a phenomenon in literary texts to be ignored.

9.3. Non-literary texts adopting the ‘Majority’ orthography

As with the literary texts (§9.2), so in letters we occasionally find the graphematic separation of prefix clitics. Consider the following example:

(265) KTU³ 2.86:4–13

→
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13

<i>ilm</i>	<i>tgr=k</i>	<i>tšlm=k</i>	<i>l=p'n</i>	<i>b'ly</i>
gods	protect.3PL=you.SG	give_wellbeing=you.SG	to=feet	lord-my
<i>šb'd</i>	<i>w=šb'd</i>	<i>mrḥqtm</i>	<i>qlty</i>	<i>w=hn=n<y></i>
seven_times	and=seven_times	from_afar	fall.1SG	and=here=me
<i>'m</i>	<i>'bd-k</i>	<i>mid</i>	<i>šlm</i>	<i>w</i>
with	servant-your.2SG	very	well-being	and
<i>b'ly</i>	<i>šlm-h</i>			
lord-my	wellbeing-his			

'may the gods guard you, may they keep you well. At the feet of my master seven times and seven times (from) afar do I fall. Here with your servant it is very well. As for my master, (news of) his well-being ...' (trans. Pardee 2003, 112; see also Dietrich & Loretz 2009, 132–133)

At lines 6 and 7 the prefix clitics \equiv *l*- and \blacktriangleright *w*- are written together with the following morphemes. This is the practice almost everywhere in this 25-line text, except in line 12, where in the sequence *w · b'ly* 'and my lord' \blacktriangleright *w*- is followed by a word divider.

While it is difficult in general to find a one-size-fits-all explanation for the graphematic separation of \blacktriangleright *w*- (§9.2.3), in this case it is tempting to suggest one. Specifically, as Pardee's translation indicates, \blacktriangleright *w*- in line 12 begins a new section of the letter, in which the topic shifts from the well-being of the letter's author, to that of his lord. This change of topic coincides with the graphematic separation of \blacktriangleright *w*-. If word division in this letter does indeed correspond to prosody, the placement of a word divider after \blacktriangleright *w*- should indicate that \blacktriangleright *w*- is its own prosodic word, with its own accent. That \blacktriangleright *w*- might receive an accent of its own in a context where the topic is shifting would make considerable sense from a prosodic point of view,

<i>tlmyn</i> _{<λ>}	<i>ʿbd-k</i> _{<λ>}	<i>l</i> _{<ω>}	<i>pʿn</i> _{<λ>}	<i>ʾadt-y</i> _{<ω>}
PN	servant-your	to	feet	lady-my
<i>šbʿid</i> _{<ω>}	<i>w</i> _{<ω>}	<i>šbʿid</i> _{<ω>}	<i>mrḥqt=m</i> _{<ω>}	<i>qlt</i> _{<λ>}
seven_times	And	seven_times	distance=PTCL	fall.1SG
<i>ʿm</i> _{<ω>}	<i>ʾadt-y</i> _{<λ>}	<i>mnm</i> _{<ω>}	<i>šlm</i> _{<λ>}	<i>rgm</i> _{<ω>}
with	lady-my	whatever	well_being	message
<i>tttḅ</i> _{<λ>}	<i>l</i> _{<ω>}	<i>ʿbdh</i> _{<ω>}		
send.IMPF	to	servant-her		

‘Speak to the queen, my lady; word of Talmiyānu, your servant. At the feet of my lady seven times and seven times from afar I have fallen. Whatever well-being (there is) with my lady, may she send back word to her servant’ (trans. after Huehnergard 2012, 192–193)

This letter is very short, consisting of a mere 15 lines, and 23 graphematic words. Yet in this short span, there are no fewer than three instances of the prefix preposition **𐤀** *l-* followed by word division, at lines 1, 6 and 15. Furthermore, **𐤁** *w-* is followed by a word divider at line 9. This degree of word division after monoconsonantal particles is wholly unexpected, given its rarity in the literary texts (cf. Chapter 6).

Finally, note that in both instances of construct NPs, the elements are separated from one another, in one instance by means of a word divider, and in the other by means of line division (which has the function of word division in this text):

- *tḥm*_{<ω>} *tlmyn*_{<λ>} ‘message of PN’ (line 4)
- *pʿn*_{<λ>} *ʾady*_{<λ>} ‘feet of my lady’ (lines 6–7)

Nor is this text without parallel. Other non-literary documents appear to adopt the same or similar word division strategy.²

The goal of the chapter is to identify the ORL of word division in these documents. I argue that, unlike any other Northwest Semitic writing system, word division in the ‘Minority’ orthography targets the pre-phonological morphosyntactic level of linguistic representation.

9.4.2. Morphosyntax of word division

The orthography of word division in a number of non-literary documents from Ugarit clearly has a very different character from that of the ‘Majority’ orthography. While in the ‘Majority’ orthography, *es* may be prefixes or suffixes, in the ‘Minority’ orthography, *es* are all suffixes, either suffix pronouns, or suffix clitics. In what follows, I first consider in detail the treatment of particles that in the ‘Majority’ orthography

² Parallels, where all/almost all clitics are written separately, include: KTU 2.97, 98, 108.

are written as graphematic prefixes, principally monoconsonantal clausal and prepositional particles.

9.4.2.1. Prefix particles

Monoconsonantal clausal and prepositional particles are almost always written as separate words in the ‘Minority’ orthography, that is, separated from morphemes both before and after by a word divider (for exceptions, see §9.4.4). This practice is clearly in evidence in (266) above. We saw that the graphematic separation of prefix particles in the ‘Majority’ orthography is attested, albeit rarely (§9.2). It is not, therefore, the fact that this graphematic separation occurs in the documentary texts that sets the orthography apart, but rather the relative proportions of separated and non-separated instances. Thus, while in KTU 1–23, the proportion of separated prefix particles is around the 5% mark, in texts adopting the ‘Minority’ orthography, the proportion is a lot higher. In (266) it is 100%.

Another feature of clitic separation that distinguishes the ‘Minority’ orthography from the ‘Majority’ orthography is the treatment of clitic chains. In KTU³ 1–23, the clitics in such chains are never written separately from one another, although the chain as a whole may be separated graphematically from its surrounding morphemes (§9.2.3). By contrast, prefix clitics in the ‘Minority’ orthography are separated from one another, *e.g.*³

(267) KTU³ 2.23:21–22 (cf. Huehnergard 2012, 35)

→ III, II → r, r → T → r, r → III, II →
 III, T II → T, r

*l*_{<w>} *pn*_{<w>} *amn*_{<w>} *w*_{<w>} *l*_{<w>}
 to face DN and to

*pn*_{<w>} <λ> *il*_{<w>} *mšrm*_{<w>}
 face gods Egypt

‘before ʾAmun and before the gods of Egypt’ (del Olmo Lete & Sanmartín 2015, 68)

(268) KTU³ 4.168:6–8

→ r, II, II →
 T III → r, T III II <V>
 II → r, III II T

*w*_{<w>} *b*_{<w>} *bt*_{<λ>} *mlk*_{<w>} *mlbš*_{<λ>}
 and in house king cloak

³ Parallels: KTU³ 2.23:1 *w*_{<w>} *k*_{<w>} *rgm*_{<w>} *špš*_{<λ>} ‘And thus says the Sun’; KTU³ 4.145:6–7 *w*_{<w>} *l*_{<w>} *tt*_{<w>} *mrkbtm*_{<λ>} *inn*_{<w>} *uṭpt*_{<λ>} ‘And two chariots lack quivers’.

*ytn*_{<ω>} *l-hm*_{<λ>}
 give.PASS to-them

‘And in the house of the king a cloak will be given to them’ (trans. in part from del Olmo Lete & Sanmartín 2015, 976)

(269) KTU³ 4.367:1

→ [𐤓]𐤔 𐤔 𐤕 𐤔 𐤕 𐤔 𐤕 𐤔 𐤕 𐤔 𐤕 𐤔 𐤕 𐤔 𐤕 𐤔 𐤕 𐤔 𐤕 𐤔 𐤕 𐤔 𐤕

[*s*]pr_{<ω>} *bnš*_{<ω>} *mlk*_{<ω>} *d*_{<ω>} *b*_{<ω>}
 list personnel.PL.CSTR king who in

*tbq*_{<λ>}

TN

‘List of the king’s personnel who (are) in *Tbq*’

(270) KTU³ 4.338:3

→ 𐤔 𐤕 𐤔 𐤕 𐤔 𐤕 𐤔 𐤕 𐤔 𐤕 𐤔 𐤕 𐤔 𐤕 𐤔 𐤕 𐤔 𐤕 𐤔 𐤕

*w*_{<ω>} *b*_{<ω>} *spr*_{<ω>} *l*_{<ω>} *št*_{<λ>}
 and on list not place.PASS

‘[A list of people who have entered the king’s palace] and are not placed on the list’

The only exception to the graphematic independence of this particle class is where they are immediately followed by a suffix pronoun or enclitic. In these cases, the suffix pronoun is written together with the particle in question, *e.g.*:

(271) KTU³ 2.13:13

→ 𐤔 𐤕 𐤔 𐤕 𐤔 𐤕 𐤔 𐤕 𐤔 𐤕 𐤔 𐤕 𐤔 𐤕 𐤔 𐤕 𐤔 𐤕 𐤔 𐤕

*w*_{<ω>} *rgm*_{<ω>} *ttb*_{<ω>} *l-y*_{<λ>}
 and word send_back to-me
 ‘and send back word to me’

(272) KTU³ 4.132:4

→ 𐤔 𐤕 𐤔 𐤕 𐤔 𐤕 𐤔 𐤕 𐤔 𐤕 𐤔 𐤕 𐤔 𐤕 𐤔 𐤕 𐤔 𐤕 𐤔 𐤕

*ktn*_{<ω>} *d*_{<ω>} *šr*_{<ω>} *phm*_{<ω>} *b-h*_{<ω>}
 tunic of Tyre ruby on-it

‘A tunic from Tyre (with) ruby on it’

9.4.2.2. Suffix pronouns

While prefix particles are generally written as separate graphematic words in the documentary texts in question, this cannot be said for suffix particles. First, exactly as in the ‘Majority’ orthography, in the ‘Minority’ orthography light suffix pronouns are always written together with the immediately preceding morpheme (see also Robertson 1994, 277). The following is a typical example:

(273) KTU³ 2.13:1–4

→ III ṽ ṽ III ▷
 III ṽ III ṽ III ▷
 ṽ ṽ ṽ ṽ III ▷
 III ṽ III ▷

$l_{\langle\omega\rangle}$ $mlkt_{\langle\lambda\rangle}$ $um-y_{\langle\omega\rangle}$ $rgm_{\langle\lambda\rangle}$ $t\dot{h}m_{\langle\omega\rangle}$
 to queen mother-my speak message

$mlk_{\langle\lambda\rangle}$ $bn-k_{\langle\omega\rangle \langle\lambda\rangle}$
 king son-your

‘To the queen, my mother, speak! A message of the king, your son’

As covered at §9.4.2.1, this includes instances where the suffix pronoun combines with a prepositional prefix.

There are very occasional exceptions to this rule (Tropper 2012, 69), e.g. $bn_{\langle\omega\rangle} h$ (KTU 1.117:4) and $l_{\langle\omega\rangle} y$ (KTU 1.117:5). However, tellingly, KTU 1.117 is not a text that uses the ‘Minority’ orthography, given syntagms such as $k=ilm$ (line 5) and $w=t'n$ (line 11).

Heavy pronominal suffixes may either be written independently or dependently in the ‘Minority’ orthography. The following examples illustrate graphematic independence:⁴

(274) KTU³ 2.33:26⁵; cf. (Cunchillos, Vita & Zamora 2003, 600)

→ III ṽ
 III ṽ III ṽ III ▷

$\dot{m}_{\langle\lambda\rangle}$ $l_{\langle\omega\rangle}$ $ytn_{\langle\omega\rangle}$ $hm_{\langle\omega\rangle}$ $mlk_{\langle\omega\rangle}$
 why not give them king

⁴ Parallels: $\text{▷ ṽ III} \text{▷ III} \text{▷} ksp_{\langle\omega\rangle} hn$ ‘their price’ (KTU³ 4.132:3).

⁵ For the interpretation, see del Olmo Lete & Sanmartín (2015, 334–335), Sivan (2001, 54) and Gordon (1998, §12.1). Cf. also textual note at Cunchillos, Vita & Zamora (2003, 600). Ahl (1973, 432) translates: ‘Why doesn’t he give (them/word what to do?) or does he rule against me’, with hm giving the second part of a two-part question (Ahl 1973, 434–435).

(278) KTU³ 4.182:56

→ 𐎗 𐎗𐎗𐎗 𐎗𐎗𐎗 𐎗𐎗𐎗 𐎗𐎗𐎗 𐎗𐎗𐎗 𐎗𐎗𐎗

bnš_{<ω>} mlk_{<ω>} yb^l-hm_{<ω>}

man king makes-them

‘The man of the king will make them’ (del Olmo Lete & Sanmartín 2015, 334)

Furthermore, there appear to be no examples of a monoconsonantal preposition separated from a following heavy suffix pronoun. Even in the strictest versions of the ‘Minority’ orthography, heavy suffixes are written together with monoconsonantal prepositions, *e.g.*:⁸

(279) KTU³ 4.168:6–8 [= (268)]

→ 𐎗𐎗𐎗 𐎗𐎗𐎗 𐎗𐎗𐎗

𐎗𐎗𐎗 𐎗𐎗𐎗 𐎗𐎗𐎗 𐎗𐎗𐎗

𐎗𐎗𐎗 𐎗𐎗𐎗 𐎗𐎗𐎗

w_{<ω>} b_{<ω>} bt_{<λ>} mlk_{<ω>} mlbš_{<λ>}
 and in house king cloak

ytn_{<ω>} l-hm_{<λ>}
 give.PASS to-them

‘And in the house of the king a cloak will be given to them’ (trans. in part from del Olmo Lete & Sanmartín 2015, 976)

9.4.2.3. Suffix clitics

OVERVIEW

Ugaritic is furnished with several enclitic suffix particles, many of which have uncertain function and/or semantics (Pardee 2008, 27; Bordreuil & Pardee 2009, 61–62; Huehnergard 2012, 78–79). The main productive particles are:⁹

- 𐎗 -m
- 𐎗𐎗 -n
- 𐎗𐎗 -y

In contrast to prefix clitics, which are regularly separated in the ‘Minority’ orthography, the suffix clitics -m, -n and -y are not.

⁸ Parallel: 2:38: w_{<ω>} ttb_{<ω>} ank_{<ω>} l-hm ‘and I sent back to them’ (Linder 1970, 44), although this tablet has an example of 𐎗𐎗𐎗 -hm not separated from a noun as well.

⁹ There are a number of other particles, for a list of which see Bordreuil & Pardee (2009, 61–62). However, these are not productive in the sense that they may not ‘in theory be attached to any other word’ (Pardee 2003–2004b, 385).

We see the same topicalising function in *špš=n* (line 21) in the next example:

(281) KTU³ 2.39:17–21

→ 17 18 19 20 21

*w*_{<ω>} *lht*_{<ω>} *ákl*_{<ω>} *ky*_{<λ>} *likt*_{<ω>}
and tablets food that sent.1SG

*m*_{<ω>} *špš*_{<λ>} *b'l-k*_{<ω>} *ky*_{<ω>} *ákl*_{<λ>}
to sun lord-your that food

*b*_{<ω>} *hwt-k*_{<ω>} *inn*_{<λ>} *špš=n*_{<ω>} *t'bd*_{<λ>}
in country-your there_is_not **sun=PTCL** ruin.IMPF.F.3SG

‘And the tablet concerning grain that you sent to the ‘Sun’, your Lord: [you have written] that there is no grain in your country. [Know that] **the ‘Sun’** is being ruined’ (trans. after del Olmo Lete & Sanmartín 2015, 417, 483, 825; see also Dietrich & Loretz 2009, 131)

The important point for our purposes is that both of these texts generally exhibit morphosyntactic word division, but suffix $\rightarrow -n$ is always univerted with the foregoing morpheme. Thus in (281) compare $\rightarrow w-$ and $\rightarrow b-$ in lines 17 and 20 respectively, with $\rightarrow -n$ in line 21. Similarly, compare $\rightarrow -n$ in line 10 of (280) with $\rightarrow l-$ in the first line of KTU 2.38:

(282) KTU³ 2.38:1

→

*l*_{<ω>} *mlk*_{<ω>} *ugrt*_{<λ>}
to king Ugarit

‘To the king of Ugarit’

-M

The function of the suffix clitic $\rightarrow -m$ is not certain, but it too appears to have a focusing or emphasising role (Huehnergard 2012, 79).¹⁴ Some scholars have been reluctant to propose any unifying semantic or functional description. Bordreuil &

¹⁴ For surveys of enclitic $-m$ in Semitic languages generally, including Hebrew, see Hummel (1957) and del Olmo Lete (2008). $\rightarrow -m$ may collocate with any part of speech (Pope 1951, 123; Bordreuil & Pardee 2009, 61) in a wide range of permutations (see the list at Bordreuil & Pardee 2009, 61–62).

Pardee (2009, 61–62) offer no suggestion, while del Olmo Lete (2008, 53–54) concludes that its role is to highlight ‘semantic and syntactic functions already embe[dd]ed in the morpho-syntaxis of the discourse’. Underlying the difficulty is the likelihood that more than one particle is represented by the spelling with *m* (Tropper 2012, 825).

Other scholars have been willing to propose functions for the particle (Watson 1992, 251–252; see also Huehnergard 2012, 78–79), boiling down to two primary functions, namely, ‘a focussing particle and an adverbial ending’ (Watson 1992, 251–252). Huehnergard (2012, 79) similarly sees two primary functions in 𐤓 -*m*. On the one hand the particle ‘generally adds focus or emphasis to the word to which it is attached’. Furthermore, ‘[s]ome adverbial forms normally occur with this particle, where it has lost any emphasizing nuance’. Huehnergard evidently entertains the possibility that the two uses of 𐤓 -*m* have the same origin. Watson (1992, 251), by contrast, considers it as an open question whether they are in fact the same particle.

The emphasising use is often found in poetic bicola (Huehnergard 2012, 79, 85–87). 𐤓 -*m* may occur either in the first colon or the second. In the following example, 𐤓 -*m* is affixed to the item that is most salient for the bicolon as a whole, in this case, that the subject went to the very top of the tower:

(283) KTU³ 1.14:IV:2–4

→ 𐤓 𐤓 < 𐤓 𐤓
 𐤓 𐤓 < 𐤓 𐤓 𐤓 𐤓 𐤓 𐤓 𐤓 𐤓 𐤓 𐤓
 𐤓 𐤓 𐤓 𐤓 𐤓 𐤓 𐤓 𐤓 𐤓 𐤓

$w=^{\text{c}}ly_{\langle\lambda\rangle}$ $\dot{l}=zr_{\langle\omega\rangle}$ $mgdl_{\langle\omega\rangle}$ $rkb_{\langle\lambda\rangle}$
 and=he_went_up to=roof tower he_climbed

$\dot{t}km=m_{\langle\omega\rangle}$ $hmt_{\langle\omega\rangle}$
 shoulder=PTCL wall

‘and he went up to the roof of the tower, he climbed the very shoulders of the wall’ (for trans. see Huehnergard 2012, 86)

While in this case an argument can be made for 𐤓 -*m* having an emphatic function, in other cases the particle may simply be used for the purpose of *variatio* (per Gzella 2007a, 140), e.g.:

(284) KTU 1.14:I:31–32 (example quoted at Huehnergard 2012, 87)

→ 𐤓 𐤓 𐤓 𐤓 𐤓 𐤓 𐤓 𐤓 𐤓 𐤓
 𐤓 𐤓 𐤓 𐤓 𐤓 𐤓 𐤓 𐤓 𐤓 𐤓

$\dot{b}m_{\langle\omega\rangle}$ $bky-h_{\langle\omega\rangle}$ $w=yšn_{\langle\lambda\rangle}$ $\dot{b}=dm^{\text{c}}-h_{\langle\omega\rangle}$
 in weeping-his and=he_slept in=shedding-his

šb'îd_{<λ>} mrhqt=m_{<λ>} qlt_{<λ>}
 seven_times distance=PTCL I_fall

'At the feet of my lady seven times and seven times from afar I have fallen' (trans. after Huehnergard 2012, 192–193)

As with \Rightarrow -n, the important point for our purposes is that, where this particle occurs in texts that otherwise write prefix clitics separately, Υ -m is unverbated with the preceding morpheme(s). Thus, at (266), line 10, we saw the adverbial Υ -m written together with the preceding morpheme in *mrhqt=m* 'from afar'. Similarly, in (285), while the prefix clitic particles \Rightarrow w- and |||| l- are written separately, Υ -m is unverbated with the infinitive *yd*⁶.

-Y

The primary function of the particle |||| -y is to indicate direct speech (Tropper 1994a; 2012, 833; Huehnergard 2012, 79). As with \Rightarrow -n and Υ -m, where this particle occurs in documents that employ morphosyntactic word division, |||| -y is written together with the preceding morpheme(s).¹⁷

(287) KTU³ 2.33:30–32

→³⁰ |||| \Rightarrow Υ |||| \Rightarrow Υ |||| \Rightarrow Υ |||| \Rightarrow Υ |||| \Rightarrow
³¹ |||| \Rightarrow Υ |||| \Rightarrow Υ |||| \Rightarrow Υ |||| \Rightarrow Υ |||| \Rightarrow
³² |||| \Rightarrow Υ |||| \Rightarrow Υ |||| \Rightarrow Υ |||| \Rightarrow Υ |||| \Rightarrow

*ht*_{<ω>} *hm*_{<ω>} *yrgm*_{<ω>} *mlk*_{<λ>} *b'î-y*_{<ω>}
 now if command.3SG king lord-my

*tmgy=y*_{<ω>} *hn*_{<λ>} *ālpm*_{<ω>} *śśwm*_{<ω>} *hnd*_{<λ>}
 arrive.3PL=PTCL here two_thousand horses these

'Now, if the King my Lord commands: "These two thousand horses should come here"': (trans. with ref. to Tropper 1994a, 476 and del Olmo Lete & Sanmartín 2015, 931–932)

Elsewhere, the word division orthography of the tablet is strongly morphosyntactic, as may be seen from the following:

¹⁷ For the example, see Tropper (1994a, 476). Tropper lists the example under 'Sichere Belege' (p. 475), but for other interpretations, see del Olmo Lete & Sanmartín (2015, 932).

(288) KTU³ 2.33:20, 22

→ 𐎗𐎠 𐎗𐎠 𐎗𐎠 𐎗𐎠 𐎗𐎠 𐎗𐎠 𐎗𐎠 𐎗𐎠 𐎗𐎠 𐎗𐎠

...

22 𐎗𐎠 𐎗𐎠 𐎗𐎠 𐎗𐎠 𐎗𐎠 𐎗𐎠 𐎗𐎠 𐎗𐎠

w _{<ω>}	āp _{<ω>}	mlk _{<ω>}	ūd̄r̄ _{<λ>}	...
And	also	king	messenger	...

w _{<ω>}	mlk _{<ω>}	b ^l -y _{<ω>}
And	king	lord-my

‘and also the king, the messenger ... and the King, my Lord ...’ (trans. after del Olmo Lete & Sanmartín 2015, 35, 204)

9.4.3. Morphosyntactic word division in non-literary documents

9.4.3.1. Accounting for the word division strategy of the ‘Minority’ orthography

The rate of separation of prefix clitics in some letters and administrative texts is so high in comparison with that which we see in literary texts, that it seems different principles are being followed (see also Tropper 2012, 69, §21.412a). This is likewise the case for the fact that the elements of clitic chains are written separately from one another, an orthography of word division that is not attested in KTU 1–23.

To a reader from a modern European or North American background, the fact that word division in these tablets conforms to what we would expect from our own writing traditions, as opposed to what we see everywhere else in Northwest Semitic, is striking. Since word division in modern European languages (in general) targets morphosyntax rather than prosody, a morphosyntactic explanation seems *a priori* attractive. Such an explanation would certainly account for the graphematic separation of prefix clitics, as well as the splitting up of clitic chains.

An alternative explanation is that the ‘Minority’ orthography, like the ‘Majority’ orthography, targets the prosodic layer of linguistic representation. The prosody in question, however, would be different, such that prefix particles would no longer be cliticised. While this would provide a consistent account of word division in the relevant documents, this explanation, too, has its drawbacks. In particular, such an explanation would presume the decliticisation of prefix particles, when, to the author’s knowledge, this is not attested anywhere else in Northwest Semitic. For this to be the case, the prosody of the language would have to be significantly reconfigured. This of course is not out of the question, but it is hard to imagine how one might find external evidence to support such a reconfiguration. This hypothesis will therefore not be pursued here.

The morphosyntactic explanation is not, however, without its own difficulties. These are now addressed in turn.

9.4.3.2. Differential treatment of prefix clitics and suffix clitics

It seems at first sight inconsistent for prefix clitics and suffix clitics to be treated differently by the orthography, with the former regularly written separately, and the latter univerted. However, this can be explained with reference to function (cf. Vajda 2005, §1.4.1.2), by observing that focalising, emphatic and direct speech-marking functions of 𐤀 -*m* and 𐤁 -*y* belong to the level of discourse, while the prefix clitics perform, in the terms of Vajda (2005), phrasal functions. This is to say that prefix clitics relate elements to one another within the world of the discourse, whereas suffix clitics relate entities in the discourse to elements outside it, *e.g.* the speaker/writer/hearer/reader. The graphematic separation of prefix clitics alongside the univertation of suffix clitics could therefore remain compatible with a consistent application of word division if word division targets only the referential and phrasal linguistic levels, while ignoring the discourse layer.

The fact that adverbial 𐤀 -*m* is not written separately can also be argued to be consistent with a morphosyntactic word division strategy. Insofar as 𐤀 -*m* is affix-like (§1.4.3), a word division strategy that targets the linguistic level after the combination of morphological affixes, but before the combination of clitics, would expect to univertate it with its hosts (for a similar argument in respect of suffix pronouns, see §9.4).

These conclusions may also apply in the case of 𐤁 -*n*. However, the existence of a determinative function for this particle is contested. Against the existence of such a function, see Gzella (2007b, 552). If it exists, however, it is possible to analyse the alleged examples under the heading of topicalisation. According to one prominent understanding, its core ‘the core of definiteness depends on the existence of a referent in the common ground known by the speaker and the hearer’ (Aguilar-Guevara, Pozas Loyo & Vázquez-Rojas Maldonado 2019, iii). Accordingly, a topicalising or focusing particle might function in some contexts like a determiner, if the purpose of the topicalisation/focus were to highlight to the reader that the participant marked is in some way known to the hearer/reader. For an analysis of the particle -*ni* in Yokot’an Maya along these lines, see Pico (2019). The alleged apodotic function of 𐤁 -*n* can be observed in ritual texts (Hoftijzer 1982; Tropper 2000, 824, also citing Tropper 1994b; Bordreuil & Pardee 2009, 62). However, this too can be understood under the rubrik of ‘topicalisation’ marker (Tropper 2000, 824): ‘[Das -*n apodoseos*] betont als eine Art Topikalisierungsmarker das erste Wort der Apodosis und markiert so den Anfang der Apodosis.’ Finally, the use of 𐤁 -*n* to generate alloforms of the prepositions, *e.g.* 𐤁𐤁𐤁 *ln* = 𐤁𐤁 *l*-‘to’. However, alloforms of prepositions with 𐤁 -*n*, although originally formed with productive 𐤁 -*n*, do not necessarily demonstrate the functionality of 𐤁 -*n*. Cf. Pardee (2003–2004b, 386) reviewing Tropper (2000): ‘An attempt should be made to distinguish between the productive use of the enclitic particle and frozen forms of other particles which arose at some time in the past by affixation of the etymologically identical particle.’

9.4.3.3. *Differential treatment of prefix clitics and suffix pronouns*

More problematic is the differential treatment of prefix clitics and suffix pronouns. Since pronouns belong either to the phrasal or referential layers of representation, according to the particular usage in context, they might be expected to be written separately if prefix clitics are so treated. As we have seen (§9.4.2.2), however, this is never the case with light suffix pronouns, nor is it reliably the case for their heavy counterparts. Furthermore, it is conspicuous that when preceded by a prefix clitic, suffix pronouns are not written separately (§9.4.2.2). By contrast, a prosodic explanation would not be vulnerable on this point, since one could simply assume that there existed a phonological rule for some speakers whereby prefix particles were routinely decliticised.

9.4.3.4. *Morphological status of suffix pronouns: comparison with Tiberian Hebrew*

The face value implication of the fact that suffix pronouns are not graphematically independent is that they have a different morphosyntactic status from prefix clausal and phrasal clitics, *i.e.* that they are affixes. In favour of this interpretation is that in Tiberian Hebrew the morphological binding of pronouns is different from that of prefix clitics. When a pronoun is suffixed to a noun, the resulting combination is predictable on purely phonological grounds. Consider the following pair of examples:

(289) $b-$ + $bayit$ = $bbayit$

(290) $b-$ + $yhūdāh$ = $bīhūdāh$

Before most consonants in most contexts, the result is simply the prefix preposition with *wa* followed by the noun in question, along with any spirantisation following from sandhi, per (289). By contrast, before ׃ , *wa* of the preposition combines to generate *e.g.* $bī$, as in (290) (for examples and further discussion, see van der Merwe, Naudé & Kroeze 2017, 325–326).

On prepositions, however, the form of the (singular) suffix pronoun is not predictable, synchronically at least, on the basis of the phonology alone. This may be seen from a consideration of Table 9.5. Compare especially of the second person singular masculine and feminine forms. At the morphological level, therefore, these elements are fused. The same situation pertains *mutatis mutandis* for suffix pronouns on verbal forms (van der Merwe, Naudé & Kroeze 2017, 91–98, esp. 96).

The distribution of word division in the Ugaritic ‘Minority’ orthography would be consistent with word division targeting pre-phonological combination if to the writers of the texts, the combinations of

Table 9.5: *Forms of the preposition ל- combining with singular suffix pronouns in Tiberian Hebrew*

לִי	<i>lī</i>	to me
לְךָ	<i>lkā</i>	to you (m.sg.)
לְךָ	<i>lāk</i>	to you (f.sg.)

preposition/verb + suffix pronoun were morphosyntactically fused in the same way. Some evidence for this in Ugaritic may be found from the great variety of accusative pronominal forms found suffixed to verbs terminating in *-n* (Bordreuil & Pardee 2009, 40). This is shown by the existence of a form such as *ylmn* /yallumannu/ ‘he struck him’ (RS 24.258:8) alongside the form *tbrknn* /tabarrikannanu/ ‘you should bless him’ (RS 2.[004] i 23’). In *ylmn* /yalluman-nu/ < /yalluman-hu/, we find the result of the combination of the verbal termination /-an/ and the suffix pronoun /hu/. By contrast, in *tbrknn* we find /-annu/ reanalysed as a suffix pronoun in its own right, which is then added to the verbal termination /-anna-/ to produce /tabarrikanna=nnu/ (for the mechanics of this, and these particular examples, see Bordreuil & Pardee 2009, 40–41).

MORPHOLOGICAL STATUS OF SUFFIX PRONOUNS: INTERNAL SYNTACTIC EVIDENCE

Ugaritic suffix pronouns are collocationally restricted, in that they must co-occur with nominals. Furthermore, they are not capable of modifying phrases. They therefore have the characteristics of morphological affixes to a greater degree than monoconsonantal prefix prepositions (§1.4.3).

In the following example, the pronoun 𐎎 *-y* ‘my’ is repeated on the two coordinate nouns 𐎎𐎎𐎎 *mid* ‘plenty’ and 𐎎𐎎𐎎𐎎𐎎 *ǰbn* ‘well-being’:¹⁸

(291) KTU³ 2.46.9–11

→ 𐎎 𐎎 , 𐎎𐎎𐎎 𐎎𐎎 𐎎𐎎 𐎎𐎎𐎎 𐎎𐎎 ,
 𐎎𐎎𐎎𐎎𐎎𐎎 , 𐎎𐎎 𐎎𐎎 𐎎𐎎 , < 𐎎 𐎎
 𐎎 𐎎𐎎𐎎 𐎎𐎎 , 𐎎𐎎 𐎎𐎎𐎎 𐎎𐎎

*ky*_{<ω>} *lik*_{<ω>} *bn-y*_{<λ>} *lht*_{<ω>} *ākī*_{<ω>}
 since sent son-my message grain

*‘m-y*_{<λ>} *mid-y*_{<ω>} *w=ǰbn-y*_{<λ>}
 with-me plenty-my and=well_being-my

‘Since my son sent the message about grain, with me (is) my plenty and my well-being’ (cf. del Olmo Lete & Sanmartín 2015, 506)

The affixal status of suffix pronouns is also indicated in the next example, where the monoconsonantal preposition 𐎎 *l-* governs the pair of pronouns 𐎎 *-y* ‘me’ and 𐎎𐎎 *-k* ‘you.SG’:

¹⁸ Parallel: KTU 2.82:2 *l mlkt ūmy ādty* ‘to the queen, my mother and my Lady’.

(292) KTU³ 2.45:23–24¹⁹

→ 𐎠 𐎢𐎣 𐎠𐎡𐎢𐎣
 → 𐎢𐎣 𐎠𐎡𐎢𐎣 𐎠𐎡𐎢𐎣 𐎠𐎡𐎢𐎣 𐎠𐎡𐎢𐎣 𐎠𐎡𐎢𐎣

*bly-m*_{<λ>} *ālp-m*_{<ω>} *aršt*_{<ω>} *l-k*_{<ω>} *w*_{<ω>}
 worn_out-PL ox-PL I_asked_for for-you and

*l-y*_{<λ>}
 for-me

‘The oxen I asked for for you and for me are worn out’ (trans. after Ahl 1973, 452)

Note that the behaviour of suffix pronouns differs from construct chain nominals. In this case more than one noun in the genitive can modify a single noun in the nominative, *e.g.*:

(293) KTU³ 1.3:III:22–23

→ 𐎠𐎡𐎢𐎣 𐎠𐎡𐎢𐎣
 < 𐎠𐎡𐎢𐎣 𐎠𐎡𐎢𐎣 𐎠𐎡𐎢𐎣 𐎠𐎡𐎢𐎣 𐎠𐎡𐎢𐎣

*rgm*_{<λ>} *š*_{<ω>} *w*_{<ω>} *lhšt*_{<ω>} *ābn*_{<ω>}
 matter wood and whisper stone

‘It is a matter of wood and a chatter of stone’ (trans. after del Olmo Lete & Sanmartín 2015, 493)

By contrast, prefix prepositions are phrasal in scope, and are therefore syntactically freer than suffix pronouns, *e.g.*:

(294) KTU³ 1.108:3–4

→ 𐎠𐎡𐎢𐎣 𐎠𐎡𐎢𐎣 𐎠𐎡𐎢𐎣 𐎠𐎡𐎢𐎣 𐎠𐎡𐎢𐎣 𐎠𐎡𐎢𐎣
 𐎠𐎡𐎢𐎣 𐎠𐎡𐎢𐎣 𐎠𐎡𐎢𐎣 𐎠𐎡𐎢𐎣 𐎠𐎡𐎢𐎣 𐎠𐎡𐎢𐎣

*il=tpz*_{<ω>} *b=hdr’y*_{<ω>} *d=yšr*_{<ω>} *w=ydmr*_{<λ>}
 god=judge.PTCP in=TN REL=sing.PASS/IMPERS and=praise.PASS/IMPERS

*b=knr*_{<ω>} *w=tlb*_{<ω>} *b=tp*_{<ω>} *w=mšltm*_{<ω>}
 at=lyre and=flute at=drum and=cymbals

‘The god who judges in TN, to whom they sing and praise to [the sound of] the lyre and flute, to [the sound of] the drum and cymbals’ (trans. with reference to del Olmo Lete & Sanmartín 2015, 330, 446, 579)

¹⁹ For the interpretation, see Ahl (1973, 452).

As the phrase structure analysis of the PP *b=knr w=tlb* ‘to the lyre and the flute’ at Figure 9.1 shows, the scope of the preposition *b=* is the phrase *knr w=tlb* ‘lyre and flute’.

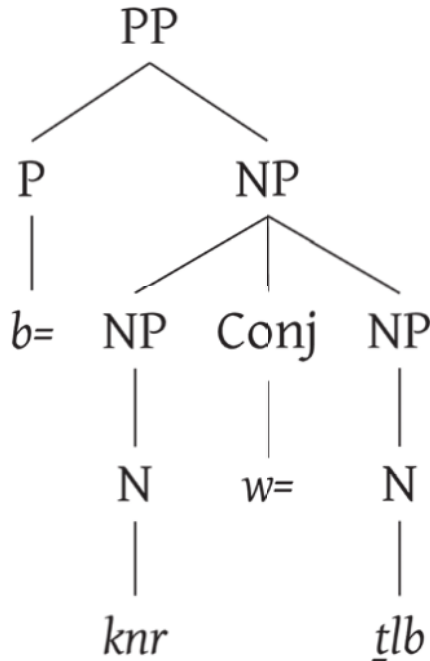


Figure 9.1: Phrase structure analysis of PP *b=knr w=tlb* (KTU³ 1.108:4)

The preposition 𐤁 *l=* can behave in a similar way, as in the following sequence of phrases, where it governs two coordinate NPs:

(295) KTU³ 1.19:IV:5–6²⁰

→ $\text{𐤁} \text{𐤇} \text{𐤁}$
 $\text{𐤁} < \text{𐤁} \text{𐤇} \text{𐤁} > \text{𐤁} < \text{𐤁} \text{𐤇} \text{𐤁} >$

*l=ht*_(λ) *w=lmh*_(ω) *l=nt*_(ω) *p=dr*_(ω)
 for=now and=forever.ADV for=now and=generation(s)

*dr*_(λ)

generation(s)

‘for now and forever, for now and for all generations [*lit.* generation(s) of generation(s)]’

²⁰ For the interpretation, see del Olmo Lete & Sanmartín (2015, 474). For the issues surrounding plural or singular interpretation of ‘generation(s)’, cf. del Olmo Lete & Sanmartín (2015, 277).

It is, however, possible simply to repeat the preposition:

(296) KTU³ 1.10:III:30²¹

→ >>> < ||| v, & >> > & >> >>
 & > & >> >> >> & || >>>

w=t^l_{<(w)} bkm_{<(w)} b=árr^o_{<(λ)} bm_{<(w)} árr_{<(w)}
 and=she_climbed next.ADV to=TN to TN

w=b=šp^on^o_{<(λ)}
 and=to=TN

‘and next she went up to Árr, to Árr and to špn’

It was noted earlier (§1.4.3) that a morphological affix may have scope over a coordinate construction if that coordinate construction is lexicalised. In such a case, therefore, the preposition would still govern a noun, rather than a phrase, and so could be viewed as morphological. In (295) it is in principle possible that both *ht w=lmh* ‘now and forever’ and ‘*nt p=dr dr* ‘now and for all generations’ are lexicalised: one could certainly imagine that they might be frequently used turns of phrase. However, the expressions occur in this form only in this text:²² the former also at III:48, 55 (in both cases with the conjunction ≡ *p-* instead of >> *w-*),²³ the latter at III:48.

The case for lexicalisation as an explanation for *b=* governing the coordinate NPs in (294) is still weaker: *knr* occurs in conjunction with *tlb* only here, and likewise for the coordination of *tp* and *mšltm*, despite each of these items occurring independently elsewhere.²⁴ On the basis of this evidence, therefore, we may conclude that prefix prepositions are syntactically independent, that is, syntactic words.

On the basis of these considerations, therefore, we can conclude that monoconsonantal prepositional prefixes in Ugaritic have properties both of morphological and free syntactic elements.

VARIABLE TREATMENT OF HEAVY SUFFIX PRONOUNS

It remains to account for the inconsistency in the word division of heavy suffix pronouns in this orthography (§9.4.2.2). If the ORL of word division is consistent, we would expect to find these treated in a consistent fashion. In Tiberian Hebrew the affixal nature of the heavy suffix pronouns is indicated by the fact that the form of

²¹ For the interpretation, cf. del Olmo Lete & Sanmartín (2015, 102, 217).

²² Based on a search of UDB. A case can, however, be made for the lexicalisation of the construct phrase *dr dr* (see del Olmo Lete & Sanmartín 2015, 277).

²³ At III:48, a word divider intervenes between *lm* and *h*, i.e. *lm_{<(w)} h*.

²⁴ e.g.: 1.101:16 (*knr*); 1.113:3 (*tlb*); 1.16:1:41 (*tp*); 1.3:1:19, 1.19:IV:26–27 (*mšltm*).

Table 9.6: Forms of the preposition -לִּ- combining with plural suffix pronouns in Tiberian Hebrew

לָנוּ	<i>lānū</i>	to us
לָכֶם	<i>lākem</i>	to you (m.pl.)
לָכֶן	<i>lāken</i>	to you (f.pl.)
לָהֶם	<i>lāhem</i>	to them (m.pl.)
לָהֶן	<i>lāhen</i>	to them (f.pl.)

the prepositional element of the combination of -לִּ- with plural pronominal suffixes is consistently לָ *lā*, as opposed to the form -לִּ- *l-*, the form found before nouns (cf. Table 9.6). By contrast, in later Northwest Semitic inscriptions heavy suffix pronouns appear to be graphematically separated, suggesting that they were not treated as affixes (Chapter 12). The Ugaritic

‘Minority’ orthography differs from both of these in that heavy suffix pronouns are on different occasions written either as independent graphematic words or as graphematic suffixes.

It is impossible to say for certain why suffix pronouns are treated in this ambivalent fashion. However, one possibility is that these pronouns are univerbated with their hosts by analogy with the singular pronouns.

9.4.3.5. Conclusion

The present section has sought to account for the word division strategy of the ‘Minority’ orthography. While at first sight a strategy of separating morphemes on the basis of morphosyntax appears to be employed, such an analysis must be able to account for two phenomena that are *a priori* problematic, namely the non-separation of suffix clitics and suffix pronouns. It was found that suffix clitics are not grammatical in the same way that prefix clitics are. Specifically, while prefix clitics relate entities to one another within the world of the discourse, suffix clitics for the most part relate elements in the discourse to entities outside it. These therefore target different linguistic levels. In these terms, therefore, the word division orthography is consistent, but only targets the pre-phonological morphosyntactic layer, and ignores discourse particles.




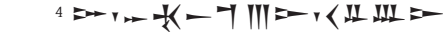
Suffix pronouns, by contrast, although grammatical insofar as they relate elements of the discourse to one another, just like prefix clitics, are, from a morphosyntactic perspective, more like affixes than free syntactic morphemes, while prefix clitics are syntactically freer. The word division orthography is therefore consistent insofar as it regularly separates freer morphemes, but univerbates affix-like morphemes.

9.4.4. Inconsistency in the use of the ‘Minority’ orthography

9.4.4.1. Occasional aberrance from morphosyntactic word division

Tablets such as KTU 2.12 (266) are consistent in their treatment of prefix clitics. However, in other tablets graphematic separation of these morphemes word division is not implemented consistently. This aberration can be occasional, as in the first line of KTU 2.11:



(297) KTU³ 2.11:1-4

→ ¹ 
² 
³ 
⁴ 

l=um-y_{<w>} *ádt-ny*_{<λ>} *rgm*_{<λ>} *tĥm*_{<w>} *tlmyn*_{<w>}
 to=mother-my lady-our say.IMP message PN

*w*_{<w>} *áĥtmlk*_{<λ>} *ʿbd-k*_{<λ>}
 and PN servant-your

‘To my mother, our lady, say: message of Talmiyānu and ʾAĥātumilki, your servants’
 (trans. Pardee 2003, 90)

After the first line in this 18-line letter, prefix particles are always separated from the following morpheme by a postpositive word divider, cf. the word divider after  *w*- in line 4. However, in the first line  *l*- is written together with the following morpheme in *l=umy* ‘to my mother’.

9.4.4.2. Hybrid word division orthography

While some non-literary documents deviate only to a small degree from ‘Minority’ orthography, others show a much greater internal variation. KTU 2.88 is a case in point. Sequences such as the following might suggest the adoption of the ‘Minority’ orthography:

(298) KTU³ 2.88:2

→ 

*l*_{<w>} *úrtn*_{<w>} *rgm*_{<λ>}
 to PN say.IMP

‘Say to ʾUrtēnu’ (trans. with ref. to Dietrich & Loretz 2009, 148–150)

By contrast, the next line appears to adopt the ‘Majority’ orthography, *e.g.*:

(299) KTU³ 2.88:3

→ 

*hlly*_{<w>} *ánk*_{<w>} *b=ym*_{<λ>}
 behold I on=day

‘Behold I am on the day’ (trans. with ref. to Dietrich & Loretz 2009, 148–150)

9.4.4.3. Accounting for inconsistent word division orthographies

The existence especially of hybrid orthographies such as that exemplified in the previous section. One possibility is that instances such as these represent instances of the ‘Majority’ orthography. On such a view, the frequent graphematic separation of prefix clitics would be understood as a reflection of prosody (cf. §9.3). Along these lines it may be significant that many of the instances of graphematically separated \blacktriangleright w- in this text involve a change in subject and/or topic. Compare, for example, the two instances of the sequence *w āt* ‘and you’, at lines 11 and 21 respectively:

(300) KTU³ 2.88:10–12

→ ¹⁰ \blacktriangleright 𐤌𐤍 \blacktriangleleft
¹¹ \blacktriangleright \blacktriangleright \blacktriangleright \blacktriangleright \blacktriangleright 𐤌𐤍 \blacktriangleright 𐤌𐤍 \blacktriangleright
¹² \blacktriangleleft \blacktriangleright \blacktriangleright \blacktriangleright \blacktriangleright \blacktriangleleft \blacktriangleright \blacktriangleright

$w=d_{\langle\lambda\rangle}^c$ $w=\dot{a}t_{\langle\omega\rangle}$ $klkl-k_{\langle\lambda\rangle}$ $\check{s}kn_{\langle\omega\rangle}$
 and=know.IMP and=you.SG everything-your establish.IMP

$l=\check{s}m-k_{\langle\lambda\rangle}$
 to=name-your

‘... And know [this]! And you, establish all that is yours in your (own) name’ (trans. with ref. to Dietrich & Loretz 2009, 149)

(301) KTU³ 2.88:16–17, 21–22

→ ¹⁶ \blacktriangleright \blacktriangleright 𐤌𐤍 \blacktriangleright \blacktriangleright \blacktriangleright 𐤌𐤍 []
¹⁷ 𐤌𐤍 \blacktriangleright 𐤌𐤍 \blacktriangleright \blacktriangleright \blacktriangleright
 ...
²¹ \blacktriangleright \blacktriangleright \blacktriangleright \blacktriangleright \blacktriangleright 𐤌𐤍 \blacktriangleright \blacktriangleright \blacktriangleright 𐤌𐤍
²² 𐤌𐤍 𐤌𐤍 𐤌𐤍 \blacktriangleright \blacktriangleright \blacktriangleright \blacktriangleright \blacktriangleright

w $bt_{\langle\omega\rangle}$ $\dot{a}h\dot{d}[]_{\langle\lambda\rangle}$ $d_{\langle\omega\rangle}$ $\dot{a}drt[]_{\langle\lambda\rangle}$
 and house particular of PN

... $w_{\langle\omega\rangle}$ $\dot{a}t_{\langle\omega\rangle}$ $b=p-k_{\langle\omega\rangle}$ $\dot{a}l_{\langle\lambda\rangle}$
 ... and you.SG in=mouth-your.SG NEG

$y\check{s}i=mnk_{\langle\omega\rangle}$
 come_out=anything

‘... And as for the particular house of *Ādrt* ... and as for you, let nothing come out of your mouth’ (trans. with ref. to Dietrich & Loretz 2009, 149)

We saw the same topic-changing function for \blacktriangleright w- followed by the word divider earlier (§9.3). If this possibility is borne out by further research, the ‘inconsistency’

in word division that we see in many of tablets can instead be seen in terms of the consistent application of a prosodic approach to word division.

9.4.5. Co-existence of two word division orthographies

In closing, it is interesting to consider the implications of the co-existence of the ‘Minority’ orthography and the ‘Majority’ orthography in what are ostensibly the same type of text. If the two orthographies are distinguished on the basis of the relevant linguistic level of representation, one morphosyntactic and the other prosodic, one need simply posit the existence of one or more ‘schools’ (see Horwitz 1971, 107–108, 112–113). On such a view, one school would favour the morphosyntactic separation of morphemes, the other favouring separation based on prosody. That there were different scribal schools at Ugarit, and that these took divergent views on word division, is supported by the fact that certain texts, *e.g.* KTU 1.65, 1.127, do not show any word division at all. This is in line with the description of Hawley, Pardee & Roche-Hawley (2016, 246), who have argued that ‘scribal education at Ugarit was not centralized (in the royal palace for example)’, but rather that ‘schools formed around individual scribes in a domestic setting, where these very scribes, often acting at high levels of the royal administration, not only taught but carried out their duties’. It is therefore not difficult to imagine that particular ‘master scribes’ may have favoured one or another approach, with their disciples adopting their habits. (On the training of scribes in Ancient Israel for comparison, cf. Koller 2021.)

In this context, it is also worth drawing attention to how unique word division in the ‘Minority’ orthography really is in the context of Northwest Semitic writing systems. As is made clear in the rest of the present study, word division in all other Northwest Semitic writing systems is fundamentally prosodic in denotation. In the history of writing, to the knowledge of the author, word division on the basis of morphosyntax, rather than prosody, does not emerge until much later, and is, of course, now the dominant form of word division in European languages. That (a school of) scribes in 13th-century BCE Ugarit were experimenting with separating morphemes on this basis is therefore of great significance for the history of word division in the world’s writing systems.

9.4.6. Conclusion

This chapter has addressed the ORL of word division in the ‘Minority’ orthography, used in a variety of epistolary and administrative texts. I have argued that word division in these texts, in contrast to what we see elsewhere in Ugaritic and Northwest Semitic, targets the pre-phonological morphosyntactic layer of linguistic representation. This is to say that freer syntactic morphemes, notably clausal proclitics and monoconsonantal prepositions, are regularly separated from their neighbours. By contrast, morphological affixes, *viz.* singular suffix pronouns, along with suffix particles with discourse functions, are written together with the foregoing morpheme. The ORL of the word division strategy can therefore be understood as consistent.

This word division strategy is unique in Northwest Semitic, and remarkable in the history of writing systems, representing, as it does, the first instance of the morphosyntactic approach to word division that we find in the writing systems of many European languages today.

9.5. Conclusion

In Part II of the study I have sought to account for word division practices in the Ugaritic material, encompassing both literary and non-literary texts. Starting from the hypothesis that graphematic words in Ugaritic alphabetic cuneiform correspond to prosodic units – specifically, ACTUAL PROSODIC WORDS – I set out to compare the distribution of these units with that of prosodic word and prosodic phrase-level units cross-linguistically. The primary target of comparison was Tiberian Hebrew, on the grounds that it is the language most closely related to Ugaritic for which we have plentiful ancient evidence – in the form of Tiberian vocalisation – for the constituency of prosodic word level units.

After surveying the kinds of morphosyntactic units represented as single graphematic words in Ugaritic alphabetic cuneiform (Chapter 6), a quantitative comparison of incidence, phrase length and morphosyntactic collocation was made (Chapter 7). It was found that the distribution of graphematic words in Ugaritic has affinities to both prosodic words and prosodic phrases in Tiberian Hebrew.

However, since especially the morphosyntactic collocation showed affinities both ways, it was necessary to take syntax into account. Syntax is relevant because there is acknowledged to be a cross-linguistic relationship between prosodic and syntactic structures at the prosodic phrase level and above (§1.5.1). Specifically, left or right edges of prosodic phrases align with the maximal projections of syntactic XPs. In Chapter 8, therefore, the relationship between Ugaritic graphematic words and syntactic structure was analysed. I started by showing for Tiberian Hebrew that prosodic phrases do in principle demonstrate edge alignment with syntactic XPs, whereas this is only a feature of prosodic words where prosodic words consist of whole prosodic phrases. From this I went on to argue, since graphematic words frequently do not show $X^{P^{MAX}}$ edge alignment, that graphematic words in Ugaritic alphabetic cuneiform correspond to prosodic words rather than prosodic phrases.

At §8.4 the issue of graphematic univertation at clause/colon boundaries was addressed. Since this pattern of univertation is *prima facie* at odds with word division corresponding to ACTUAL PROSODIC WORDS, these were an important set of exceptions to consider. I suggested that univertation at clause/colon boundaries might have *comparanda* both in univertation at clause boundaries in Phoenician inscriptions, discussed in Part I, as well as in enjambement in Ancient Greek epics. While the units so univertated would not correspond to prosodic words in normal speech, the poetic effect of the syntax from one line running on to the next might have an auditory

value akin to that of prosodic wordhood, accounting for the lack of word division in these instances.

In §8.5 I showed that the major features of word division and univerbation that we saw in literary compositions may also be found in non-literary documents, demonstrating that the use of word division in the former is not purely a feature of the verse form, but has a wider validity for understanding the orthography of Ugaritic texts in general.

Finally, in Chapter 9 I addressed the graphematic separation of monoconsonantal prefix particles. Although the general tendency is for these to be written together with the following morpheme, there are important exceptions, in both literary and non-literary documents. These were explored in turn, with particular attention paid to the 'Minority' orthography, where contrary to what is found elsewhere in Ugaritic and Northwest Semitic, monoconsonantal prefix particles are regularly written as independent graphematic words.

In sum, I have attempted to demonstrate that prosodic and morphosyntactic explanations are able to provide compelling accounts of word division practices in Ugaritic and that these provide greater explanatory power than those offered to date. There are, to be sure, areas where further work is needed, especially in the occasional graphematic separation of monoconsonantal prefix particles in the 'Majority' orthography, as well as in univerbation across clause boundaries. Nevertheless, it is hoped that the foregoing argumentation is able to provide a solid foundation for future research.

PART III

Hebrew and Moabite

Chapter 10

Word division in the consonantal text of the Hebrew Bible

10.1. Introduction

10.1.1. Problem of word division in Tiberian Hebrew

In the Northwest Semitic writing systems discussed up to this point a key issue has been the morphosyntactic inconsistency of word division. This is to say that graphematic word boundaries frequently do not correspond to the edges of what one might call ‘dictionary words’ or ‘morphosyntactic words’. By contrast, the type of morphosyntactic inconsistency observed has generally been found to be consistent with demarcation according to prosodic words.

The word division orthography of Tiberian Hebrew consonantal text poses quite a different problem: word division here is remarkably consistent. The difficulty is that this consistency does not correspond to consistency at a morphosyntactic level. This is because items with parallel morphosyntactic statuses, such as conjunctions or prepositions, are treated differently, seemingly, on a lexeme-by-lexeme basis. Consider the following example:

(302) 2Sam 17:11

← כַּחֹל אֲשֶׁר-עַל-הַיָּם לְרֹב

$k=hl_{(w)}$ $\text{ʔ}sr_{(w)}$ $l_{(w)}$ $h-ym_{(w)}$ $l=rb_{(w)}$

as=sand which on The-sea to=abundance

‘as the sand that is by the sea for multitude’ (KJV)

The example contains three prepositions: $k-$ ‘as, like’, $l-$ ‘to, for’ and $עַל$ ‘on’. Two of these, $k-$ and $l-$, are univertated with the following morphemes, $חֹל$ hl ‘sand’ and $רֹב$ rb ‘abundance’, respectively. By contrast, $עַל$ l is graphematically independent.

Such a distribution is not in itself unparalleled in Northwest Semitic inscriptions: $l-$ and $k-$ are regularly to be found written together with the following morpheme in both Ugaritic and Phoenician. Furthermore, the preposition l is well attested as an

independent graphematic word in both writing systems. What differentiates the Tiberian Hebrew consonantal text from its Northwest Semitic epigraphic forebears is that morphemes of more than one consonant are *always* written as independent words.

The question to address is the following: on what grounds are monoconsonantal prefixes, such as -כ־ *k-* and -ל־ *l-*, always graphematically *dependent*, while their multiconsonantal counterparts are always graphematically *independent*? The following domains of explanation are in principle available:

- The word division orthography is conventionalised on a lexeme-by-lexeme basis, *i.e.* -ל־ *l-* is always written with the following morpheme, simply because that is the way things are done, and על־ *ʿ* is always written as an independent word, for the same reason;
- Word division is applied on purely graphematic grounds: single consonant units are unverbated, but longer units are not;
- Word division is morphosyntactic, along the lines that we saw in the Ugaritic ‘Minority’ orthography;
- Word division is prosodic.

The present part sets out to harness the Tiberian tradition of Biblical Hebrew to assess what layer of the language – whether prosodic, morphosyntactic, semantic or graphematic – is targeted by word division in Tiberian Hebrew. Key to the inquiry is the treatment of morphemes that would in many writing systems be written as separate words, notably MŠH WKLB מִשֶׁה וְכֹלֵב . These I term ‘graphematic affixes’. I will ask what properties this group of morphemes share that mean that they are written as part of the following morpheme, rather than as separate graphematic units. The answer to this question will dictate what linguistic domain is targeted by graphematic word division. I argue for the claim that graphematic words in the consonantal text of the Hebrew Bible correspond to MINIMAL PROSODIC WORDS, that is, the smallest units that can exist as prosodic words in their own right, irrespective of context.

10.1.2. Outline

This argument is made over two chapters. The goal of the present chapter is first to provide an analysis of the morphosyntactic status of graphematic affixes in Tiberian Hebrew (§10.2). On the basis of this analysis, the chapter sets out to demonstrate that neither morphosyntactic nor graphematic explanations are sufficient to account for the unverbation of graphematic affixes (§10.3 and §10.4).

In Chapter 11, I set out to demonstrate that a purely prosodic explanation of word division is able to account for the distribution of graphematic affixes in Tiberian Hebrew. This is done first by analysing the prosodic properties of graphematic affixes (§11.1). Morphemes are divided into five types according to their phonological shape as specified in the lexicon: C-, CV-, CVC- and CVX-prefix morphemes, and CVC-suffix

morphemes. It is found that in principle only CVC-prefix morphemes may stand as independent prosodic words. Thereafter Drescher's existing account based on a combination of prosody and morphosyntax is considered (§11.2), but found not to match the full range of behaviours observed. The remainder of the chapter is devoted to providing a phonological explanation whereby graphematic word division in Tiberian Hebrew targets minimal prosodic words.

Finally, in Chapter 12 I provide evidence for the considerable antiquity of the word division strategy of the consonantal Masoretic Text, by showing that in almost all respects the same approach can be observed in the Siloam Tunnel inscription (8th century BCE) and the Mesha^c inscriptions (9th century BCE).

10.2. Morphosyntactic status of graphematic affixes in Tiberian Hebrew

10.2.1. Introduction

Graphematic affixes in Hebrew may be categorised according to the following three morphosyntactic types:

- Definite article
- Suffix pronouns
- Prepositions
- Clausal prefixes

10.2.2. Definite article

The definite article -הַ *ha-* in Tiberian Hebrew is always written together with the following morpheme (on the definite article in Semitic, cf. Tropper 2001). Thus in the following example, -הַ *ha-* is prefixed to both שָׁמַיִם *šmym* 'heavens' and אֶרֶץ *ʾrṣ* 'land':

(303) Gen 1:1

← בְּרֵאשִׁית בְּרָא אֱלֹהִים אֶת הַשָּׁמַיִם וְאֶת הָאָרֶץ:

b=ršyt br' ʾlhym ʾt h-šmym

in=beginning created God OBJ the-heavens

w=t h-ʾrṣ

and=OBJ the-earth

'In the beginning God created the heaven and the earth' (KJV)

Syntactically the Biblical Hebrew definite article is notable for the fact that it is not necessarily affixed to the phrasal head. In particular, in construct genitive constructions it is the dependent NP that is marked as definite, e.g.:

(304) Gen 24:3 (cf. van der Merwe, Naudé & Kroeze 2017, 224)

אֱלֹהֵי הָאָרֶץ וְאֱלֹהֵי הַשָּׁמַיִם ←
 *lhy h-šmym w=ʔlhy h-ʔrṣ
 God the-heavens and=God the-earth
 ‘the God of heaven and the God of earth’

Wintner (2000) shows that the definite article in Modern Hebrew is much closer to an affix than to a free syntactic element. There is not the space to conduct a similarly detailed analysis of the definite article in Biblical Hebrew. Suffice it to say that evidence in this direction is provided in (304): it is usually not possible for a construction of this kind, of the shape ‘A-X and A-Y’ in the form ‘A-X and Y’ (van der Merwe, Naudé & Kroeze 2017, 224), *i.e.*:

(305)

*lhy=h-šmym w=ʔrṣ
 God=the-skies and=earth

This is to say that it is the whole construct phrase element that is marked as definite as one morphosyntactic word; the article does not in general have scope over a coordinated phrase. The contrast with English in this regard is noteworthy: *the*, as the previous example demonstrates, is able to head a phrase containing coordinated constituents.

10.2.3. Suffix pronouns

Suffix pronouns are always graphematically univerted with the previous morpheme in Tiberian Hebrew. This is true even in the case of the heavy suffixes of the second- and third-person plural, whether suffixed to nouns or verbs, *e.g.*:

(306) Deut 12:6

וְהֵבֵאתֶם שְׂמֶה עֲלֵיכֶם וּזְבָחֵיכֶם ←
 w=hbʔtm šmh ʔty-km w=zbḥy-km
 and=you_shall_bring there burnt_offerings-your and=sacrifices-your
 ‘And thither ye shall bring your burnt offerings, and your sacrifices ...’ (KJV)

(307) Psa 18:15

:וַיִּהְיֶם ←
 w=yhm-m
 and=he_routed-them
 ‘and (he) discomfited them’ (KJV)

In this respect Tiberian Hebrew differs from early Northwest Semitic inscriptions, to be discussed in the next chapter, where heavy suffixes are found written as separate words (cf. Segert 1961, 236), *e.g.*:

(308) KAI 181.18

וַיִּשְׁחַבְתֶּם לְפָנַי קִמּוֹשׁ ←

w=ʔšhb hm lḥny kmš

and=I_dragged them before DN

‘And I dragged them (*i.e.* the articles of Yahweh) before Kemosh’

Pronouns by their nature play the same morphosyntactic role as nominals. This does not in itself mean that they are not morphological elements. After all, their position in the NP or VP is fixed, being placed as they are directly after the governing verb or preposition, or immediately after the noun on which they are dependent. By contrast, the word order of full nominals is not rigidly fixed in this way. The distribution of pronominal suffixes is therefore affixal, and so morphological rather than syntactic.¹

10.2.4. Prefix prepositions

The monoconsonantal prepositions בִּ- *b-* ‘in’, לִ- *l-* ‘to’ and כִּ- *k-* ‘as, like’ are obligatorily graphematic prefixes:²

(309) Gen 25:16

וְאֵלֶּה שְׁמֹתֵם בְּחֻצְרֵיהֶם וּבְטַרְיָתָם ←

w=ʔlh šmt-m b=ḥsry-hm w=b=tyrt-m

and=these names-their by=settlements=their by=battlements-their

‘And these are their names, by their settlements, and by their battlements’ (after KJV)

These prepositions have by-forms – לְמוֹ *lmō*, בְּמוֹ *bmō* and כְּמוֹ *kmō* – which are written as independent words, *e.g.*:

(310) Job 27:14

אִם יִרְבּוּ בָנָיו לְמוֹתָרְבָּ ←

ʔm=yrbw bn-yw lmw=ḥrb

if=multiply children-his for=sword

‘If his children be multiplied, [it is] for the sword’ (KJV)

¹ Compare the affixal nature of suffix pronouns in Ugaritic (§9.4.3.4).

² On the place of these prepositions within the history of Semitic, see del Olmo Lete (2011).

(311) Isa 44:19

← הַצִּיּוֹ שָׂרַפְתִּי בְמוֹ-אֵשׁ

hšy-w śrpty bmw=ʔš

half-it burned.PRF.ACT.1SG in=fire

‘I have burned part of it in the fire’ (KJV)

The graphematic prefix prepositions are phrasal in scope. This is to say that, while they cannot head clauses, they are able to head phrases, including those containing coordinated elements:³

(312) 1Sam 15:22

← הַחֶפֶץ לַיהוָה בְּעֹלֹת וּזְבָחִים

h=hps l=yhwh b=ʔwt w=zbhym

Q=pleasure to=DN in=burnt_offerings and=sacrifices

‘Does the LORD have [as great] delight in burnt offerings and sacrifices ... ?’ (after KJV)

(313) Gen 48:5

← אֶפְרַיִם וּמְנַשֶּׁה כְּרֵאֲבִין וְשִׁמְעוֹן יִהְיוּ-לִי:

ʔprym w=mnšh k=rʔwbn w=šmʕwn yhyw

PN and=PN as=PN and=PN will_be

l-y

to-me

‘Ephraim and Manasseh, even as Reuben and Simeon, shall be mine’ (ERV)

These examples can be paralleled with structurally identical examples multiconsonantal prepositions:

(314) 1Sam 14:21

← עִם-יִשְׂרָאֵל אֲשֶׁר עִם-שָׂאֻל וַיִּזְנְתֶּן:

ʕm=ysrʔl ʔsr ʕm=šʔwl w=ywntn

with=Israel who with=Saul and=Jonathan

‘with the Israelites that [were] with Saul and Jonathan’ (KJV)

³ Compare the situation in Modern Hebrew, where it has been argued that light prepositions, such as *לְ*, do not head their own PPs (Botwinik-Rotem 2004; Botwinik-Rotem & Terzi 2008, 409).

These data show that prefix prepositions are not morphological affixes, but are rather syntactic in nature. It should be said, however, that such examples are relatively rare, at least in the case of monoconsonantal prepositions. Much more commonly the prefix preposition is repeated on each governed element:

(315) Gen 13:17

← קוֹם הַתְּהַלֵּךְ בְּאַרְצוֹ לְאַרְצָהּ וּלְרֵחְבָּהּ

qwm hthlk b=ʾrṣ l=ʾrk-h w=l=rḥb-h

rise walk in=land to=length-its and=to=width-its

‘Arise, walk through the land in the length of it and in the breadth of it’ (KJV)

(316) Gen 19:16

← וַיַּחֲזִקוּ הָאֲנָשִׁים בְּיָדוֹ וּבִידֵי אִשְׁתּוֹ

w=yḥzqw h-ʾnšym b=yd-w w=b=yd ʾšt-w

and=seized the- with=hand-his and=with=hand wife-his
men

‘So the men seized him and his wife ... by the hand’ (KJV)

(317) Gen 48:20

← יִשְׁמַנְךָ אֱלֹהִים כְּאֶפְרַיִם וְכַמְנַשֶּׁה

yśm-k ʾlhym k=ʾprym w=k=mnšh

will_make-you.SG God like=PN and=like=PN

‘God make you as Ephraim and as Manasseh’ (after KJV)

10.2.5. Preposition מִן *min* ‘from’

Unlike the other graphematic prefix prepositions, מִן *min* has the graphematic prefix allomorphs מִי- *mi-* and מֵ- *mē-*. מִן *min* consists of two consonants, and so is written as an independent word. However, both מִי- *mi-* and מֵ- *mē-* are written together with the following morpheme, and presuppose the assimilation of the final /-n/. The main conditioning factor for these two variants appears to be the consonant following מִן *min*. Before ה *h*, in practice mostly the definite article, or א ², מִן *min* is used; before other gutturals we find מֵ- *mē-*; otherwise מִי- *mi-* (with the following consonant doubled) is used. Compare the following examples:

(318) 2Sam 20:6

← מִן־אֲבִשָׁלוֹם

mn ʾbšlwm

than Absalom

‘than [did] Absalom’ (KJV)

(319) 2Sam 20:12

מִן־הַמִּסְלָה ←

mn h-mslh

from the-highway

(320) Gen 2:10

מִיֶּדֶן ←

m=^cdn

from=Eden

(321) Gen 4:4

מִבְּכֹרֹת ←

m=bkrwt

from=firstlings

'of the firstlings' (KJV)

The important point for the immediate discussion, however, is that in each of these cases, the morphosyntactic and semantic roles of מִן *min* are identical, yet the morpheme has different treatments in terms of graphematic wordhood.

10.2.6. Clausal prefixes

Clausal prefixes are the most syntactic of the prefixes: their lack of selectivity as to the word class of the graphematic host, which may be *inter alia* a preposition, a finite verb or an adjective, and their clausal scope (see examples immediately below). -וּ *w-*, as a conjunction, simply joins together two sentence level constituents. -שֶׁ *še-*, by contrast, as a relativiser, introduces a subordinate clause, as in the following examples:

(322) Judg 7:12

בְּחֹל שַׁעַל־שַׁפַּת הַיָּם ←

k=hwl š=^lśpt h-ym

as=sand REL=on=edge the-sea

'as the sand (which is) by the seaside' (KJV)

(323) Ezra 8:20

וּמִן־הַנְּתִינִים שָׁנָתָן דָּוִד ←

w=mn=h-ntynym š=ntn dwyd

and=the-PN REL=gave PN

'and of the Nethinims, whom David ... appointed ...' (KJV)

10.2.7. Interrogative pronoun -מַּי ma- ‘what?’

מַּי *mah* has two graphematic allomorphs, written respectively as a graphematic proclitic -מַּי *ma-*, and the graphematically independent מַּי *meh* / מַּי *māh*.⁴ The independent word is by far the most common form found, *e.g.*:⁵

(324) Isa 1:5

עַל־מַּי תִּכּוּ עוֹד ←
 ʿl mh tkw ʿwd
 on what strike.PRF.PASS.2PL again
 ‘Why should ye be stricken any more?’ (KJV)

(325) Isa 5:4

מַּי-לַעֲשׂוֹת עוֹד לְכַרְמִי ←
 mh l=ʿswt ʿwd l=krm-y
 what to=do more for=vineyard-my
 ‘What could have been done more to my vineyard ... ?’ (KJV)

Graphematic proclitic -מַּי *ma-* is much rarer (cf. also Isa 3:15; examples cited in Drescher 2009, 98):

(326) Exod 4:2 (Ketiv)

וַיֹּאמֶר אֵלָיו יְהוָה מַזֶּה בְּיָדְךָ ←
 w=yʿmr ʿl-yw yhwh m=zh b=yd-k
 and=spoke to-him DN what=this in=hand-your
 ‘And the LORD said unto him, What [is] that in thy hand?’ (KJV)

The graphematically independent allomorph may, however, be found in the same syntactic contexts as dependent -מַּי *ma-* (Josh 22:24 is parallel to Isa 3:15):

(327) Exod 13:14

מַּי־זֶה ←
 mh=zʾt
 what=this
 ‘What [is] this?’ (KJV)

⁴ See Kantor (2017, 286 and n. 356) for the conditions under which מַּי *meh* is found, as well as for discussion in the context of the diachrony of the various other Hebrew reading traditions. For מַּי *māh* see §11.1.6 and §11.4 below.

⁵ Evidence for the antiquity of this spelling can be found in the rendering in *e.g.* 1QIsa^a.

As with מִן *min*, therefore, the distribution of the graphematic word and non-graphematic word graphematic variants cannot be motivated on morphosyntactic grounds.

10.3. Morphosyntactic status of graphematic affixes

A word division orthography that targets a consistent morphosyntactic ORL should separate morphemes of the same type in the same way. This is what we see in the Ugaritic ‘Minority’ orthography, where syntactically freer forms are separated, while more affix-like elements are univerbated.

This is not what we see in the case of the consonantal text of the Hebrew Bible. Granted, the definite article and suffix pronouns are affix-like, and might therefore reasonably be expected to univerbate with a graphematic host. However, the critical evidence against a morphosyntactic explanation lies in the treatment of prepositions and clausal prefixes. Clausal particles are free syntactic morphemes *par excellence*: they are restricted very little in terms of collocation, and freely coordinate entire phrases and clauses. Yet וְ *w-* and שֶׁ- *še-* are always univerbated with the following morphemes.

The prefix prepositions לְ *l-*, בְּ *b-* and כְּ *k-* are more collocationally restricted than the clausal prefixes וְ *w-* and שֶׁ- *še-*. Yet they are still more syntactically free than suffix pronouns in that they may govern coordinate phrases. Furthermore, their syntactic distribution is identical to multi-consonant prepositions. Yet the latter are never univerbated with a graphematic host. The fact that the allomorphs of לְ *l-*, בְּ *b-* and כְּ *k-*, namely לְמוֹ *lmō*, בְּמוֹ *bmō* and כְּמוֹ *kmō*, are not univerbated provides further evidence that graphematic status is not a function of morphosyntactic status. The same variation may be observed in the case of the allomorphs of מִן *min*, as well as the clausal pronoun מַה *mah*.

If morphosyntactic status does not offer the key to graphematic wordhood, perhaps a purely graphematic explanation might serve. It is this question that is addressed in the next section.

10.4. Graphematic status of graphematic affixes

It has often been observed that graphematic wordhood correlates with weight: while monoconsonantal morphemes (as written) are graphematically dependent, morphemes consisting of two or more consonants are without exception graphematically independent. However, it is not immediately clear whether this constraint is phonological or graphematic.

Let us assume that the constraint is graphematic. This is to say, let us now assume that the minimal weight constraint states that a graphematic word minimally consist of two full letters. We therefore posit a minimal graphematic word structure along the lines of Figure 10.1.

Such an assumption also provides an account of the graphematic wordhood of the morphemes explained under the former scenario:

- MŠH WKLB מִשָּׁה וְכָלֵב are expected to be graphematic prefixes, since they are written with only one full letter;
- Morphemes consisting of more than one consonant, including bi- and tri-consonantal prepositions, are expected to be written as independent words, since in their written form they will be written with at least two full letters;
- Finally, morphemes like מִי *mī*, i.e. those phonologically of the shape CV, but which are written with two full letters, are expected to be written as independent words, for the very reason that they are written with two full letters.

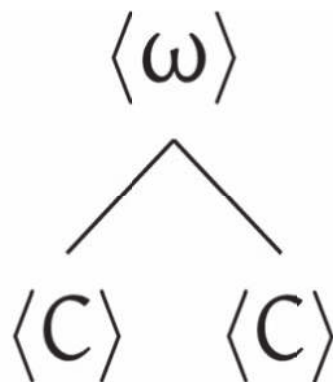


Figure 10.1: Minimal graphematic word template

So far, so good: a graphematic account accords with the observed phenomena. Indeed, a graphematic explanation can account for the distribution of מַה *mah* and מֵ- *mē-*, which a phonological explanation was not capable of doing:

- מַה *mah* is written with two full letters, and so is expected to be written as an independent graphematic word;
- By contrast, מֵ- *ma-* is written with only one full letter, and so is expected to be written as a prefix;
- מֵ- *mē-* is written with only a single consonant, and so is expected to be written as a prefix.

However, a graphematic explanation cannot account for the suffixal behaviour of : these are minimal graphematic words, just as they are minimal prosodic words, and yet they are written as suffixes.

A further weakness of the graphematic approach is that it is only explanatory to a limited extent: it stipulates that any morpheme, except suffix pronouns, spelled with two full letters will be written as an independent graphematic word, which is the observed behaviour.

At this point it is worth pointing out that, from a broader Semitic perspective, there is no *a priori* reason why Tiberian Hebrew should have adopted the orthography it has. While it may be captured synchronically by graphematic rules, these in themselves do not explain why two consonants rather than one should constitute a graphematic word. It is therefore worth asking the question, from a diachronic perspective, where such a stipulation might have arisen.

It is plausible to suppose that the underlying motivation for a minimal graphematic word of two letters might ultimately be prosodic. After all, a morpheme consisting

of fewer than two consonants is necessarily of limited prosodic extent, and cross-linguistically such entities tend to be prosodically weak. It is therefore time to consider the prosodic characteristics of graphematic affixes. It is this possibility that is addressed in Chapter 11.

10.5. Conclusion

While in the orthography of the consonantal Masoretic Text morphological affixes are written together with neighbouring lexical words – notably the definite article and suffix pronouns – this is not sufficient to account for their status as graphematic affixes. This is because many other graphematic affixes are phrasal or clausal – *i.e.* syntactic – in scope. Furthermore, both מִן *min* and -מַ *ma-*, as well as the monoconsonantal prepositions -כִּ *k-*, -לִ *l-* and -בִּ *b-*, have graphematic allomorphs which are entirely equivalent in terms of the morphosyntactic and semantic function. Word division in Tiberian Hebrew can, therefore, be said to be neither morphosyntactic nor semantic in nature. Nor does a purely graphematic account explain the univerbation of heavy suffix pronouns, or account for where a minimal graphematic weight of two graphemes might have arisen. It is therefore to prosody that I now turn for explanations of these data.

Chapter 11

Word division in the consonantal Masoretic Text: Minimal prosodic words

11.1. Introduction

In the previous chapter it was established that word division in the consonantal text of the Hebrew Bible is neither graphematic nor morphosyntactic in nature. The present chapter explores the possibility of a prosodic explanation, in particular, that graphematic words bear a relationship to minimal prosodic words in the consonantal Masoretic Text. The first step in this is to establish the minimal characteristics of prosodic words in Tiberian Hebrew. In these terms, the prosodic characteristics of affixes in Tiberian Hebrew can be shown to be a function of their status vis-à-vis MINIMAL SYLLABICITY and FOOT-BINARITY. Five types of affix can be identified, distinguished in terms of the shape of the lexical representation of their morphemes:

- C-prefix morphemes, *i.e.* those consisting in their synchronic lexical representation of a single consonant only, *e.g.* -כּ k-, -לּ l- and -בּ b-;
- CV-prefix morphemes, *i.e.* those consisting in their synchronic lexical representation of a consonant and a vowel: -הּ h^a-;
- CVC-prefix morphemes, including many prepositions, but also nominal morphemes: אֶת 'et (*nota objecti*), אִתּ 'et 'with', עַד 'ad 'up to, until', עַל 'al 'to, upon', עִם 'im 'with', אִם 'im 'if', כּל kol 'all', בֶּן bēn 'son', בַּת bat 'daughter' and לֹא lō 'not';
- CVX-prefix morphemes: *i.e.* morphemes with an unspecified second consonant, which is copied from the following morpheme at the phonetic level: -הּ haX-, -שׁ šeX-, -מּ miX- and -(ה)מּ ma(h)X-;
- CVC-suffix morphemes, *i.e.* the heavy pronominal suffixes, *e.g.* מֵהֶם -hem 'them.M'.

I discuss each lexical representation type in turn (§11.1.1 to §11.1.5), as well as the particular characteristics of מַה mah (§11.1.6). The discussion is carried out in terms of minimal prosodic wordhood as described in the Introduction (§1.4.2.5).

Once the criteria for minimal prosodic wordhood have been established, I discuss the one existing account of graphematic wordhood in Tiberian Hebrew of which I am aware, that of Drescher (2009) (§11.2). Drescher combines prosodic and morphosyntactic

criteria to arrive at necessary and sufficient constraints on graphematic wordhood. This account, however, turns out not to be capable of accounting for all the data. In the chapter's final sections (§11.3 onwards), therefore, I offer an account purely in terms of prosody, whereby the graphematic word in Tiberian Hebrew corresponds to the MINIMAL PROSODIC WORD.

11.1.1. C-prefix morphemes

The morphemes in this category are monoconsonantal clausal and prepositional prefixes, *e.g.*:¹

(328) Gen 1:2

וְהָאָרֶץ ←
 [{vɔ.hɔ:.ʕɔ:.ʔɛsʕ}]
 'And the land'

(329) Gen 1:1

בְּרֵאשִׁית ←
 [{ba.ʔe:.ʕi:.iθ.}]
 'In (the) beginning'

(330) 2Kgs 1:1

לִיּוֹאֵשׁ ←
 [{li.jo:.ʕɔʃ.}]
 'to Joash'

Per the principles outlined in the previous section, the *shwa* vowel can be modelled as a structurally short vowel, *i.e.* one that occurs short in all environments (Khan 1987; 2020). These morphemes can therefore be modelled as having no vowel position in their lexical representations per Figure 11.1. Since the minimal prosodic word is of the shape CVC or CVV, the prefixation of a C-morpheme will generate the phonological structure CCVC, which is not a permitted phonetic structure. The morpheme must, therefore, be incorporated into the prosodic structure of a neighbouring morpheme. This is often achieved by inserting an epenthetic vowel after the morpheme, usually [a], generating the phonetic bi-syllabic structure [Cv.CVC].

¹ Phonetic transcriptions follow the scheme in Khan (2020). Prosodic feet (= phonological syllables) are indicated in parentheses '(...)' ; phonetic syllables are marked terminally with a period '.' ; prosodic words are marked between curly braces '{...}' ; phonological transcriptions are given between forward slashes '/.../' ; phonetic transcriptions are given between square brackets '[...]'

For example, בְּרֵאשִׁית *bršyt* ‘in beginning’ (Gen 1:1) at (303) above comprises the two morphemes בִּ- *b-* ‘in’ and רֵאשִׁית *ršyt* ‘beginning’. רֵאשִׁית *ršyt* can be said to have the following syllabic structure, of two feet each comprising one phonological syllable:

(331) /{(rē)(šīt)}/

When the morpheme *b* is preposed, it is adjoined to the morpheme to its right. Since *b* has no vowel of its own, it cannot generate its own syllable, and is therefore incorporated into the following syllable, maintaining the structure of two feet (=phonological syllables) (see also Figure 11.2):

(332) /b+{(rē)(šīt)}/

(333) /{(brē)(šīt)}/

However, on transposition to the phonetic layer, this shape violates the permitted phonetic syllable structure by introducing a syllable of the shape CCV. Each foot is therefore broken up at the phonetic level into two syllables, of the shapes [CV.CV:] and [CV:VC.] respectively, satisfying the minimal phonetic syllabicity constraints (compare Figure 11.2 with Figure 11.3):

(334) [{{(ba.rē:.)}({fi:iθ.})}]

Note that there is no lexical rule that states that these morphemes cannot be accented. Rather, the accentuation follows simply from the application of stress rules at the phonetic level. Therefore, if the syllable containing one of these morphemes is in a stressable position, it can be stressed even when the syllable’s vowel is *shwa* or *ḥaṭef* (cf. Khan 2020, 486–496):



Figure 11.1: C-prefix lexical template

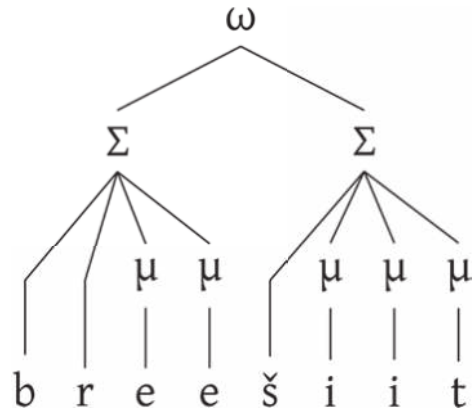


Figure 11.2: Analysis of /brēšīt/

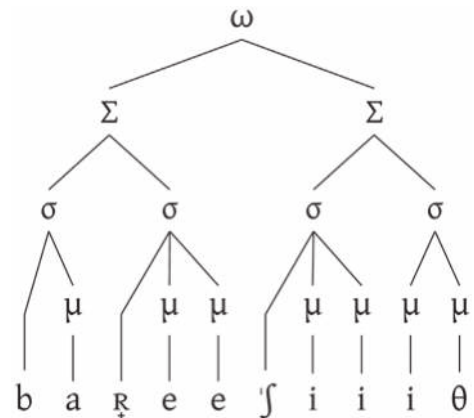


Figure 11.3: Analysis of [ba.rē:fi:iθ]

(335) Num 5:22

בְּמִעֲיָךְ ←

[(,ba.me:)(ʿa:jiχ.)]

in=stomach-your

‘in your stomach’

Note that a separate historical process of pre-tonic lengthening in Hebrew accounts for examples such as the following (cf. Suchard 2019, ch. 4):

(336) Exod 30:4

בְּהֵמָהּ:

[{(bo:)(ʿhe.em.)(mo:.)}] < *bā' - < *ba' -

‘with them’

Since the Tiberian Hebrew accent is calculated by counting backwards from the final syllable of the word, it is in principle not possible to tell, at least on accentual grounds, between a FREE CLITIC and INTERNAL CLITIC analysis of Tiberian Hebrew PREFIX + HOST units (for a similar situation in Ancient Greek, see Klavans 2019[1995]). Furthermore, since in Tiberian Hebrew junctural phenomena have scope of the prosodic phrase (φ) (Dresher 1994), the fact that these are found at the boundary of the prefix and host is also not diagnostic.

C-prefix morphemes never carry a disjunctive accent.

11.1.2. CV-prefix morpheme: -הַ h^a

CV-prefix morphemes differ from C-prefix morphemes in having a specified lexical vowel, which is invariably present. The interrogative morpheme -הַ h^a- carries such a vowel, which is usually realised as *ḥaṭef pataḥ*, e.g.:

(337) Nah 3:8

הֲתִיטֵבִי

[ha,θe:tʰa.'vi:]

‘Art thou better ... ?’ (KJV)

As with -שֶׁ- še-, -הַ h^a- never stands as an independent prosodic word, demonstrated by the fact that it never occurs with a disjunctive accent. However, -הַ h^a- differs from שֶׁ- še- in that the lexical vowel is short, rather than of unspecified length. -הַ h^a- does not therefore satisfy the criteria for a minimal prosodic foot, and must in principle be incorporated into the first foot of the following morpheme:

(338) /ha + {(tē)(ṭbī)}/ → /{(ha.tē)(ṭbī)}/ → [{{(ha,θe:)}(tʰa.'vi:)}]

These facts suggest a lexical representation along the lines of Figure 11.4. As in the case of $\text{-}הָֿ$ $še$ -, however, since the vowel length is not specified, $\text{-}הָֿ$ h^a - is not a valid prosodic word at the morphophonemic level, and so never carries a disjunctive accent, cf. the discussion on lexical *ḥaṭef* vowels in Khan (2020, 429–438).

When, however, $\text{-}הָֿ$ h^a - occurs before another CV syllable, an illicit CVCV structure is generated at the phonetic level (Khan 2020, 436). Under these circumstances $\text{-}הָֿ$ h^a - must be coerced into a permitted structure. The most straightforward way to achieve this is simply to incorporate the following consonant into a closed syllable:

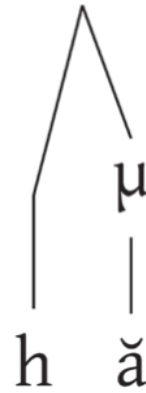


Figure 11.4: Analysis of $\text{-}הָֿ$ - h^a -

(339) Gen 34:31 (cf. Khan 2020, 385)

הַכּוֹזֵן

/ha+k+{(zō)(nat)}/

Q=like=harlot

This strategy is favoured when the first consonant is a morpheme in its own right, as in the case of C-prefix morphemes (see previous example). Many of the examples adduced in Khan (2020, 385–386) are of this kind.

However, where the first consonant of the following foot is not a morpheme in its own right, other strategies can be attempted. One approach is to double the first consonant of the following morpheme to produce a licit CVC.CV structure. The morphosyntactic structure is preserved, since the final consonant in the first syllable in such cases does not belong to the following morpheme as such:

(340) 1Sam 10:24 (cf. Khan 2020, 534)

הֲרִאִיתָם

[{{(haṛ.)}(ṛi.ʕi:)}(ʰθɛ:ɛm)}]

‘Have you seen?’

Another approach that similarly preserves the morphosyntactic structure is to coerce $\text{-}הָֿ$ h^a - into its own foot, and lengthen the vowel to meet the bimoraic constraint of canonical syllabicity (on the morphological motivation of such a move, cf. Khan 2020, 386). Consider the following example:

(341) Jer 8:22 (for vocalisation and discussion see Khan 2020, 436)

הַצִּירִי ←
 [{{(ha:.)}(sʰo:.)}({ʔri:})}]
 Q=balm

Under these circumstances, -ִירִי h^a - may carry secondary stress, marked with *gaʿya*:

(342) Gen 27:38

הַבְּרָכָה ←
 /hǎ+{(bra)(kat)}/
 Q+ blessing

11.1.3. CVC-prefix morphemes

Prefix morphemes of the shape CVC include: אֶת $\text{ʔ}ēt$ (*nota objecti*), אִתּוֹ $\text{ʔ}ēt$ ‘with’, עַד $\text{ʔ}ad$ ‘up to, until’, עַל $\text{ʔ}al$ ‘to, upon’, עִמּוֹ $\text{ʔ}im$ ‘with’, אִם $\text{ʔ}im$ ‘if’, כֹּל $\text{ʔ}ol$ ‘all’, בֶּן $\text{b}ēn$ ‘son’, בַּת $\text{b}at$ ‘daughter’ and לֹא $\text{l}ō$ ‘not’. In contrast to C- and CV- prefix morphemes, these are characterised by their capacity to carry primary stress.

Since secondary stress can be marked by a variety of means, including by means of conjunctive accents (Price 2010, 3–4; Khan 2020, 458), it is more helpful to use a type of accent that is unique to primary accents, namely disjunctive accents. These denote that the prosodic word so accented is final in its prosodic phrase (Dresher 2009, 99). At least for a proclitic, this must indicate that the accent is primary and not secondary. Accordingly, the capacity to carry a disjunctive accent can be used as a determinant of minimal prosodic wordhood. Consider the following examples involving the prepositions עַד $\text{ʔ}ad$ ‘until’, עַל $\text{ʔ}al$ ‘upon’ and עִם $\text{ʔ}im$ ‘with’ respectively:

(343) Gen 8:5

עַד הַחֹדֶשׁ הָעֲשִׂירִי ←
 [{{(ʔaδ.)} {{(ha:.)}(hō:.)}(ðeʃ)(hɔ:)(ʔa.sí:.)}({ʔri:})}]
 ‘until the tenth month’

(344) Gen 8:4

עַל הַרֵי אֲרָרָט: ←
 [{{(ʔal.)} {{(hɔ:.)}({ʔre:.)}(ʔa.ʔɔ:.)}({ʔɔ:ɔtʰ.})}]
 ‘upon the mountains of Ararat’ (KJV)

(345) Gen 24:12

← עַם אֲדֹנָי אַבְרָהָם:
 [{{('ʕim.)} {(ʕa.ðo:.)('ni:.)}{{(ʕav.)(ʔɔ:.)('hɔ:.ɔm.)}}]
 'unto my master Abraham' (KJV)

The class also includes several morphemes that are most of the time joined to the following prosodic word by *maqef*, such as אֶל 'to', אַל 'not'. Consider, e.g.:²

(346) Josh 4:18

← אֶל הַחַרְבָּה
 [{{('ʕɛl.)} {(hɛ:.)('ħɔ:.)(ʔɔ:.)('vɔ:.)}}]
 'unto the dry land' (KJV)

(347) 1Sam 2:24³

← אֶל בָּנָי
 [{{('ʕal.)} {(bɔ:.)('nɔ:.ɔj.)}}]
 'Nay, my sons' (KJV)

In fact, examples with the disjunctive accent can be provided for each of the '[s]mall function words that can be cliticized to any word' listed by Dresher (2009, 101), in turn listing from Breuer (1982, 167).⁴

The fact that these morphemes may carry a disjunctive accent is consistent with their meeting the criteria for minimal prosodic wordhood, and, therefore, with a lexical representation includes a prosodic word node dominating the morpheme, per Figure 11.5.

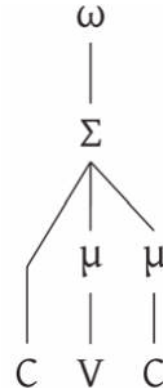


Figure 11.5: CVC morpheme template

11.1.4. CVX-prefix morphemes

Some prefix morphemes require in principle the doubling of the initial consonant of the following morpheme group, notably -הָ *ha-* 'the', מִן *min* 'from' and the relativiser -שֶׁ *še-*. The lexical template of morphemes of this kind can be generalised as per the tree in Figure 11.6. This is to say that the final mora is unspecified at the lexical level; its nature is determined at the phonological level in context, e.g. here in the case of -הָ *ha-*:

² See also אֵף *ʔp* 'also' (Isa 48:12), פֶּן *pen* 'lest' and בַּל *bal* 'not' (Psa 32:9).

³ This may be the only example involving אֶל *ʔl*.

⁴ Thus, for those not already exemplified, see: אֵת *ʔt* (*nota objecti*), Gen 1:16; אִם *ʔim*, Jer 12:17; כֹּל *kol*, Gen 8:19; בֵּן *bēn*, Gen 21:2; בַּת *bat*, Gen 36:39; עֵת *ʔt* 'time', Gen 24:11. In addition, Dresher does not mention לֹד *lōḏ*, e.g. Gen 2:25, or אִתּוֹ *ʔt* 'with', e.g. Lev 4:17.

(348) Gen 9:28

הַמַּבּוּל ←

/(haX) + {(mab)(būl)}/ → [{(ham.) (mab.) ('bu:.ul.)}]

'the flood'

Since the final mora is unspecified at the lexical level, the lexical representation of the morpheme does not satisfy the requirements of minimal prosodic wordhood. Accordingly, the prosodic foot representation has no superordinate prosodic word node dominating it. This account is consistent with the fact that $\text{-הַ} ha\text{-}$ is never found carrying a prosodic word's primary stress, *i.e.* it never stands at the end of a prosodic phrase, but must instead be incorporated in the prosodic word of the following morpheme(s).

Where the syllable containing the morpheme $\text{-הַ} ha\text{-}$ is of the shape CVC, $\text{-הַ} ha\text{-}$ does not in principle carry a secondary accent, since syllables of this shape do not in general carry secondary stress (§1.4.2.5).

The matter is different, however, where $\text{-הַ} ha\text{-}$ stands before gutturals: although at earlier stages of Hebrew guttural consonants could geminate, in Tiberian Hebrew this was not possible (Khan 2020, 280). Accordingly, where $\text{-הַ} ha\text{-}$ occurs in this environment, its vowel is lengthened in compensation. Because the resulting syllable is open, it can take secondary stress:

(349) Gen 1:27

הָאָדָם ←

/haX+ {(ʔa)(dam)}/ → [{(hɑ:.)(ʔɑ:.)('ðɑ:.ɑm.)}]

'the man'

(350) Gen 1:21

הַיָּיִת ←

/(haX) + {(ḥay)(yat)}/ → [{(,ha:.) (ḥaj.) ('jɑ:.)}]

'the living'

Also among CVX-prefix morphemes can be placed the relativiser $\text{-שֶׁ} še\text{-}$. Note the dagesh after $\text{-שֶׁ} še\text{-}$ in the following example.⁵

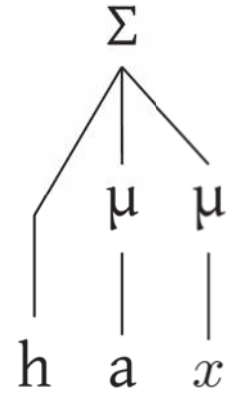


Figure 11.6: $\text{-הַ} ha\text{-}$ lexical template

⁵ I am grateful to Ivri Bunis and Geoffrey Khan for pointing this out to me.

(351) Qoh 2:26

לְאָדָם שְׂטוּב ←

l=ʾdm š=tb

to=person who=good

‘to a man that [is] good’ (KJV)

The vowel of this particle is almost invariably *segol*, implying a lexical representation $\check{s}eX$.⁶ However, like C-prefix morphemes, -שְׂט $\check{s}e$ - never carries a disjunctive accent. It may, however, carry secondary stress marked either by *ga‘ya* or by a conjunctive accent, *e.g.*:⁷

(352) Qoh 1:7

שְׁהֵנָּה לְיָם ←

[ʃɛ:.han.na.ħɔ:.li:.im.]

‘where the rivers’

Since the form of -שְׂט $\check{s}e$ - is CVX, it constitutes a prosodic foot in its own right; there is therefore no need for the introduction of an epenthetic vowel. Given a syllabic realisation of ‘the rivers’ of [han.na.ħɔ:.li:.im], -שְׂט $\check{s}e$ - can simply be introduced as an independent foot:

(353) /(\check{s}eX)/ + [(han.)(na.ħɔ:)(li:.im)]

Accentuation rules are then applied. The phonetic sequence [na.] might be expected to receive stress, since it is two syllables prior to the primary stress, but since -שְׂט $\check{s}e$ - is an open syllable in the same prosodic word, -שְׂט $\check{s}e$ - carries the secondary stress preferentially:

(354) [(ʃɛ:)(han.)(na.ħɔ:)(li:.im.)]

The fact that -שְׂט $\check{s}e$ - never carries a disjunctive accent suggests that its lexical representation is dominated only by a prosodic foot node, and not a prosodic word. The lexical schema is given at Figure 11.7.

⁶ The Masoretic Text gives -שְׂט $\check{s}e$ - on one occasion, at Qoh 3:18, : וְלִרְאוֹת שְׁהֵם בְּהִמָּה הִמָּה לָהֶם, which I have checked against L at https://archive.org/details/Leningrad_Codex. However, this should perhaps be seen as an instance of propretonic reduction (cf. Prince 1975, 201). On the development of -שְׂט $\check{s}e$ - in the history of Hebrew, see Givón (1991).

⁷ For the particular use of *mehuppak* instead of *ga‘ya* especially with -שְׂט $\check{s}e$ -, see Yeivin (1980, 196 # 241). For the tendency for -שְׂט $\check{s}e$ - to bear secondary stress, see Yeivin (1980, # 215, # 233, # 241). Note that *dehi* in *e.g.* Psa 135:2 is prepositive, and does not, therefore, mark the location of the primary accent (cf. Yeivin 1980, 268).

מִן *min* ‘from’ is a more complicated case. This has three allomorphs, מִן *min*, -מִי *mi-* and -מֵי *mē-*, whose distribution is largely conditioned by the phonological environment: מִן *min* occurs before ה *h* and שׁ *ʃ*, -מִי *mē-* before other gutturals, and -מֵי *mi-* before all other consonants. Consonant doubling is only seen in -מֵי *mi-*, e.g.:

(355) Gen 3:2

מִפְּרִי
 [(mip.)(pa.ʕi:.)]
 ‘from fruit’

Before gutturals other than ה *h* and שׁ *ʃ*, the vowel is lengthened instead of doubling, as was the case with -הֵי *ha-*. The long vowel in this case is realised as [e]:

(356) Gen 2:10

מֵ־עֵדֶן ←
 [(me:.)('ʕe:ðen.)]
 ‘from Eden’

Before ה *h* and שׁ *ʃ*, however, מִן *min* is realised as מִן *min*:

(357) Gen 2:9

מִן־הָאָדָמָה ←
 [(min.)(ħa:.)('ʕa.ða:.)('mɑ:.)]
 ‘from the ground’

Conversely, only the allomorph מִן *min* may carry a disjunctive accent, although this is very rare:⁸

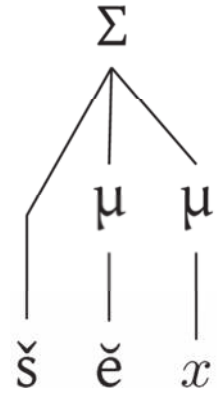


Figure 11.7 -שׁ šeX-lexical template

⁸ Exod 2:7 appears to be the only example of מִן *min* carrying any accent, either conjunctive or disjunctive, despite the approx. 7000 examples of the preposition occurring without prepositional suffix (on the basis of a search in BibleWorks 9.0). The phonological environment of מִן *min* in this example is, however, parallel to the majority of instances where מִן *min* is written as an independent orthographic word, namely before the definite article -הֵי *ha-*. This shows that, in principle, מִן *min* can function as an independent prosodic word in this environment in Tiberian Hebrew, even if it only very seldom does so.

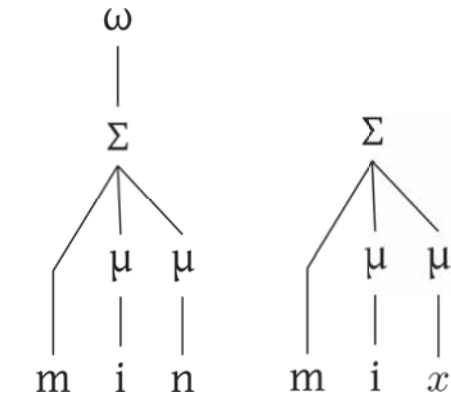
(358) Exod 2:7

מִן הַעֲבָרֹת ←

[{'mi:.in} {(hə:)(ʕiv.)(ʔij.)('jo:.ot.)}]

'of the Hebrew women' (KJV)

One way of accounting for the syllabification and accentuation of מִן *min* allomorphs is to assume a lexical representation per Figure 11.8 and Figure 11.9. This is to say that two allomorphs are listed in the lexicon: /min/ and /miX/. Only the first is dominated by a prosodic word node at the lexical level, since only this form meets the minimal prosodic word requirements. The correct allomorph is selected at the phonological level, and is then incorporated into the prosodic structure at the phonetic level:

Figure 11.8: מִן *min*
lexical templateFigure 11.9: -מִן *miX*-
lexical template

(359) /{miX + {(p̄ri)}/ → [{{(mip.)(pa.'ʔri:.)}]

(360) /{miX} + {(e:dn)}/ → [{{(me:.)('ʕe:δ̄en.)}]

(361) /{(min)} + {haX} + {(i:)(riy)(yōt)} → [{{(min.)} {(ha:.)}(ʕiv.)(ʔij.)('jo:.ot.)}]

Because only the allomorph מִן *min* is dominated by a prosodic word node, only this allomorph is capable of standing as an independent prosodic word with primary stress. By contrast, both -מִן *mi-* and -מִן *mē-* must be incorporated into the following prosodic word.

Of course, one could account for the same behaviour by proposing three lexical allomorphs, namely, /min/, /mē/ and /miX/ respectively, instead of two. However, modelling /mē/ as an instantiation of /miX/ has some advantages, in that it is more parsimonious at the lexical level, requiring only two allomorphic representations. Furthermore, it preserves the underlying unity of representation of /miX/ and /mē/, whereby the latter is a lengthened equivalent of the first only on account of the fact that later in the reading tradition gutturals could not be lengthened. Finally, there is evidence of some degree of free variation of /min/ ~ /miX/, as we saw in the previous section, whereas the same is not true of /mi/ ~ /mē/, which is strictly environmentally conditioned, the latter only occurring in the environment of the guttural.

11.1.5. CVC-suffix morphemes

Satisfying the moraic criteria for minimal prosodic wordhood is necessary but not sufficient for a morpheme to behave as an independent prosodic word. This is shown by the accentuation of heavy suffix pronouns for the second- and third-person plural, namely, כֶּם -*kem*, כֶּן -*ken*, הֶם -*hem* and הֵן -*hen*. Consider the following example:

(362) Gen 1:21

לְמִינֵיהֶם ←

[{{(la.mi:)(ne:)}('hε:εm.)}]

'according to their kinds'

The prosodic word is stressed on the final syllable. This implies that, for the purposes of accentuation, the suffix הֶם -hem is treated as the final syllable of the prosodic word. Yet הֶם -hem satisfies the moraic criteria for minimal prosodic wordhood, as it is bimoraic. Accordingly, despite fulfilling the criteria for minimal prosodic wordhood, the lexical representation of הֶם -hem must not be dominated by a prosodic word node, per Figure 11.10. This makes sense, if it is recalled that suffix pronouns, including the heavy pronouns, are affixes rather than clitics (§10.2.3).

At the phonological level, therefore, הֶם -hem is incorporated into the prosodic word of the preceding morpheme:

(363) /l + {{(mī)(nīm)} + (hēm)/ → /{(lmī)(nē)(hem)}/ → [{{(la.mi:)(ne:)}('hε:εm.)}]

Although the Tiberian tradition is clear that heavy suffixes are incorporated into the preceding prosodic word, there is evidence that this was not always the case. As we will see (§12.3.2) in the next chapter in early Northwest Semitic inscriptions suffix pronouns were written as separate orthographic words, which, we will argue, suggests that they were at least potential prosodic words in their own right at that stage of the language.

11.1.6. Interrogative pronoun mah 'what?'

The interrogative pronoun מַה mah represents an *a priori* liminal case between CVC and CVX analyses. In most instances, this morpheme is represented orthographically as מַה mah, but is joined to the following morphemes by *maqfef*, thereby denoting that מַה mah does not carry its own accent in that instance:

(364) Exod 13:14

מַה זֶה ←

[{{ma:.'zzo:θ.}}]

'What [is] this?' (KJV)

It is, however, perfectly possible for מַה mah to stand at the end of a prosodic phrase, marked by a disjunctive accent, as here with *atnah*:

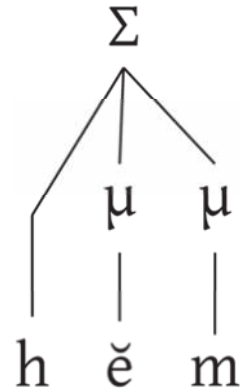


Figure 11.10: הֶם -hem lexical template

(365) Mal 2:14

עַל-מָה ←

[{\ʕal.'mɑ:.}]

‘Wherefore?’ (KJV)

מָה *mah* is, therefore, a potential prosodic word, suggesting a bimoraic lexical representation along the lines of Figure 11.11.

Complicating this picture, however, are two facts. First, the consonant of the following morpheme is in principle doubled, which we see in (364) in the doubling of [z] in זָאֵת *zōt* [this.F]. This doubling might suggest an analysis along the lines of that adopted for CVX-prefix morphemes above, whereby the doubling is an expression of an unspecified mora at the lexical level. However, synchronically, at least, the doubling in instances such as this was understood in the Tiberian Masoretic tradition as an instance of the phenomenon known as *deḥiq* (Khan 2020, 447–453).

Deḥiq is where ‘a long vowel in word-final position is shortened’ (Khan 2020, 443). This affects the vowels *qamaṣ* and *segol*, i.e. /ɔ/ and /ɛ/, typically when the following conditions are met (cf. Khan 2020, 443):

- The vowel is final in a word that either:
 - Has the stress on the penultimate syllable, and that stress is marked by a conjunctive accent; or
 - The word is joined to the following morpheme by *maqṣef*;

and:

- The following word is stressed either:
 - On its first syllable; or
 - On the first full vowel, after an initial *wa*, i.e. in its first prosodic foot.

Consider the following example:

(366) Deut 31:28 (see Khan 2020, 443–444 for vocalisation, translation and discussion)

וְאֶעֱיֶדָה בָּם ←

[{\vɑ.ʔɑ:.'ʔi:.ðɑ:.}{'bbɑ:.ɑm.}]

‘I shall cause to witness against them’

The doubling is motivated by lengthening the consonant in compensation for the shortening of the previous vowel (Khan 2020, 447). One of the exceptions to this is



Figure 11.11: מָה *mah* lexical template

where the following word begins with a guttural. *Ga'ya* is used to indicate that the vowel retains its expected length in such circumstances (Khan 2020, 446):

(367) 2Kgs 1:13 (cf. Khan 2020, 446)

← עֲבָדֶיךָ אֵלֶיךָ

[{ʕa.vɔ:.ðe:.χɔ:.'ʕe:.el.lɛ:.}]

'These servants of yours'

Assuming a representation of *מה* *mah* at some level as /*ṣ*/, its phonetic output certainly matches that of syllables under *deḥiq*, and the medieval tradition understood it as such:

The compression [of a long vowel] may occur in a word that does not have an accent but is a small word, as in *מה־תאמַר* 'whatever (your soul) says' (1 Sam. 20.4), *זֶה־בְנִי* 'This is my son' (1 Kings 3.23), *מה־בְּרִי* 'What, my son?' (Prov. 31.2). (*Hidāyat al-Qāri'* trans. Khan 2020, 452)

Accordingly, one could take the view that consonant doubling after *מה* *mah*, need not be accounted for at the lexical level, but rather as a phonetic process, and the analysis per Figure 11.11 is adequate.

The problem with this analysis, however, is that in cases where *מה* *mah* is supposedly affected by *deḥiq*, the normal conditions for *deḥiq* are not always met. Granted, they are met in (364), since *מה* *mah* is joined to the following morpheme by *maqṣef*, and the following word is accented on the first syllable. Consider, however, the following example:

(368) Gen 28:17

← מַה־נֹּרָא

[{ma'.nno:.'ɾɔ:}]

'How [this place] is feared!'

In (368), *מה* *mah* is joined to the following morpheme by *maqṣef*, but the latter is not stressed on the first full vowel, but on its second. Similar is the following example, where again the following morpheme is not stressed on the first full vowel, but two syllables later, on the penultimate:

(369) Gen 29:15

מַה־מִּשְׁכָּרְתְּךָ:

[{ma'.mmaʃ.kuɾ.'te.χɔ:}]

'What are your wages?'

Indeed, it seems in general that *deḥiq*-like behaviour in מַה *mah* occurs regardless of the accentuation of the morpheme following.

The consonant doubling does seem to be compensatory, however. This is suggested by the lack of shortening before gutturals: since gutturals cannot be doubled in Tiberian Hebrew, the vowel remains long. The long vowel is marked by *ga‘ya* in the next example:

(370) Gen 44:15

מַה־הַמַּעֲשֶׂה הַזֶּה ←
 [{mɔː.ham.maː.ʔa.ʕeː.}] {haz.ʕeː.}]
 ‘What is this deed?’

Evidence that the *deḥiq*-like behaviour of מַה *mah* might have a different origin from *deḥiq* elsewhere is suggested by the quality of the vowel: while in (366) above the vowel under *deḥiq* is *qamaṣ*, i.e. [ɔ], the quality of מַה *mah* under *deḥiq* is *pataḥ*, i.e. [a]. As Khan (2020, 447–448) points out, since Tiberian Hebrew /ɔ/ originates from *ā, the fact that מַה *mah* has *pataḥ*, i.e. [a], rather than [ɔ], means that the shortening in the case of מַה *mah* must have occurred before *ā became /ɔ/. This, however, is not the case with *deḥiq* in instances such as (366), where a short(ened) /ɔ/ is found. In sum, the *deḥiq*-like behaviour in the case of מַה *mah* must be older than *deḥiq* elsewhere.⁹

It is true that in the Tiberian tradition מַה *mah* was not read as one might expect for a prefix of the shape CVX, in that the vowel was likely pronounced half-long, i.e. /Vː/, rather than short (Khan 2020, 450). The evidence for this comes from the Karaite transcriptions of the Hebrew Bible into Arabic script, where the vowel of מַה *mah* is transcribed by the Arabic *mater lectionis* ʾ (ʾalif) (Khan 2020, 450):

(371) Gen 21:17

מַה־לָּךְ ← mh=l=k [what=to=you]
 ما ← mː=lʰh [what=to_you]
 ‘What (is the matter with) you ... ?’ (after NAS)

However, this rendering of the vowel of מַה *mah* is not universal in the Karaite tradition (Khan 2020, 451). Furthermore, in the Babylonian tradition, the vowel was pronounced short (Khan 2020, 451). Finally, Greek transcriptions of the Hebrew Bible in Origen’s Hexapla indicate a short vowel (Kantor 2017, 281; Khan 2020, 451–452):

⁹ Alternatively, in *deḥiq* generally, but not in the case of מַה *mah*, *ā did not shorten completely, and underwent the vowel shift (Khan 2020, 448). Either way, however, the process must have taken place differently in the case of *deḥiq* generally, and in the case of מַה *mah*.

(372) Psa 30:10

מַה־בְּצֵעַ ← *mh=bš^c* what=profit

→ μεββεσε *mebbese*

‘What profit ... ?’ (KJV)

It seems, therefore, that the half-long reading of the vowel of מַה *mah* in the Tiberian tradition is a secondary orthoepic phenomenon of conforming the pronunciation to the consonantal text.

The second complication with מַה *mah* is that, in at least two instances, we do not find orthographic מַה *mah* with *maqfep* in the *ketiv* form of the text, but -מַ- *ma-*, written as a C-prefix morpheme (cf. Drescher 2009, 98):

(373) Exod 4:2 (*Ketiv*)

מַזֶּה ←

⟨mzh⟩

‘What [is] that ... ?’ (KJV)

(374) Isa 3:15 (*Ketiv*)

מַלְכֵם ←

⟨mlkm⟩

‘What mean ye ... ?’ (KJV)

Nevertheless, in the Tiberian tradition, at least, -מַ- *ma-* in these instances is to be read in the same way as מַה *mah* with *maqfep*, as shown by their *qere* variants (cf. Drescher 2009, 98; Khan 2020, 450–451):

(375) Exod 4:2 (*Qere*)

מַה־זֶּה ←

[ma·'zzε:.]

‘What [is] that ... ?’ (KJV)

(376) Isa 3:15 (*Qere*)

מַה־לְכֵם ←

[ma·llε:·'χε:·εm.]

‘What mean ye ... ?’ (KJV)

These anomalies may admit of a diachronic, if not a synchronic, explanation. Although in Tiberian reading מַה *mah* and -מַ- *ma-* are phonologically /ma/, the

orthographic distinction may represent an older phonological distinction, between prosodically independent /mah/, and proclitic /maX/. This state of affairs is suggested by the Ugaritic parallel clitic ~ independent pair, $\text{𐎎 } m$ and $\text{𐎎 } \Xi mh$ ‘what?’, respectively.

The most common variant of the impersonal interrogative pronoun is $\text{𐎎 } \Xi$ /mahu/ (vocalisation per Huehnergard 2012), e.g.:¹⁰

(377) KTU³ 1.17:VI:35

→ $\text{𐎎 } \Xi \cdot \text{𐎎 } \leftarrow \text{𐎎}$

$mh_{\langle\omega\rangle}$ $yqh_{\langle\lambda\rangle}$

what attain

‘What can he attain?’

Since Ξh is not a *mater lectionis* in Ugaritic (cf. Pardee 1997, 133), it must represent a consonant rather than a vowel (Bordreuil & Pardee 2009, 42). If one assumes the same etymological starting point for Hebrew $\text{מה } mah$, the final $\text{ה } h$ is consonantal, rather than vocalic (Bordreuil & Pardee 2009, 42; Huehnergard 2012, 36).

There is, furthermore, evidence of a reduced variant of the pronoun, $\text{𐎎 } /ma(V)/$ in the following example:¹¹

(378) KTU³ 1.14:I:38–39

→ $\text{𐎎 } \rightarrow \text{𐎎}$

$\rightarrow \rightarrow \rightarrow \rightarrow \rightarrow \rightarrow \rightarrow \rightarrow \rightarrow$

$m\acute{a}t_{\langle\lambda\rangle}$ $krt_{\langle\omega\rangle}$ $k=ybky_{\langle\lambda\rangle}$

‘What’s (the matter with) Kirta, that he weeps?’

The analysis of $\text{𐎎 } \rightarrow \text{𐎎}$ $m\acute{a}t$ in this instance is debated. Huehnergard (2012, 128) lists two possibilities:

- /máa-tĩ/ < *maha-tĩ, i.e. INT.ACC + pronominal ending *tĩ
- /má-ʔattã/ < /mahu ʔattã/, i.e. INT + PRON.2SG¹²

Whichever of these analyses turns out to be correct, both presuppose a reduced form of /mahV/ → /má(V)/, i.e. /h/ > /ʔ/.¹³

¹⁰ See del Olmo Lete & Sanmartín (2015, 528) for translation and further examples.

¹¹ See del Olmo Lete & Sanmartín (2015, 528) for translation and one other example.

¹² The issue with this second possibility is that the second person pronoun does not sit easily with the 3SG form *y-bky* [3SG-weep] (Huehnergard 2012, 128). A third possibility is emendation to *mn* /mannu/ ‘Who is Kirta that he should weep?’ (Pardee 2008, 31; cf. Huehnergard 2012, 128). However, such an approach should only, in the view of the present author, be adopted as a last resort.

¹³ KTU³ supposes that ⟨h⟩ is missing here, as indicated in the text *m-<h>*.

If a similar reduction also took place in Hebrew, this form could have served a precursor to -מַח *maX-*, i.e.:¹⁴

(379) /mahV.C/ > /màV.C/ > /maC.C/

Such a development would account for the *dehiq*-like behaviour of מַח *mah*, whereby orthographic מַח *mah* represents the (original) prosodically independent /mah(V)/. This was later reduced to -מַח *maX-*, which is represented orthographically in a small minority of instances by -מַח *ma-*, as well as in the pointed text by means of the doubling of the following consonant and the *maqfef*. *Dehiq*-like behaviour in מַח *mah* would then amount to a *qere/ketiv* alternation, with the pointed text representing -מַח *maX-*, but the consonantal text in most cases representing (original) מַח *mah*.

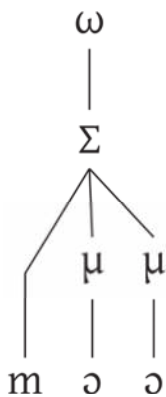


Figure 11.12: מַח *māh*
lexical template

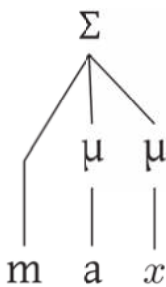


Figure 11.13: -מַח *maX-*
lexical template

On this account it is also possible to explain the vowel quality alternation in מַח *mah*: in environments where the following consonant were not doubled, the loss of final ח *h* would have motivated compensatory lengthening of /a/ to /ā/. This vowel would have then shifted with all other instances of /ā/ to /ō/. Since the shift from /ā/ to /ō/ took place early in Canaanite dialects, evidenced as it is in the Amarna letters, the orthographic form מַח *mah*, with final ח *h*, would have very ancient roots. It seems to have been preserved by a reanalysis, by which the final ⟨h⟩ was reinterpreted as a *mater lectionis* for /ō/.¹⁵

Nevertheless, the net result of the process would have been phonologically conditioned allomorphs מַח *māh* /mō/ (before gutturals and prosodic phrase boundaries) and -מַח *maX-* /maX/ elsewhere.¹⁶ From a synchronic perspective, therefore, Tiberian Hebrew can be seen to have two allomorphs of מַח *mah* represented at the lexical level, per Figure 11.12 and Figure 11.13, in an analogous way to מִן *min*. מַח *māh* is dominated by a prosodic word node, and is capable of standing as an independent prosodic word, while -מַח *maX-* is not so dominated, and is necessarily prosodically dependent on the following morpheme.

¹⁴ Cf. Kantor (2017, 286) who assumes that /ma/ with following gemination is the original form within the Hebrew reading tradition.

¹⁵ See also Kantor (2017, 286–288) for detailed discussion of the relative dating of changes to the vocalisation of מַח *mah* in the various Hebrew reading and transcription traditions.

¹⁶ For the conditions of the allomorph מַח *meh*, see Kantor (2017, 286).

11.1.7. Conclusion: Criteria for minimal prosodic wordhood

For a morpheme to be realised as a prosodic word in its own right, its lexical representation must satisfy foot-binarity, and both morae must be fully determined at the lexical level (*i.e.* one mora cannot be *X*). Now that we have established the nature of minimal prosodic wordhood, we are in a position to explore the relationship between the prosodic word and the graphematic word in Tiberian Hebrew. I turn first to the one existing account of graphematic wordhood in Tiberian Hebrew of which I am aware, that of Drescher (2009), who combines prosodic and morphosyntactic criteria to arrive at necessary and sufficient constraints on graphematic wordhood.

11.2. Combining prosody and morphosyntax (Drescher 1994; 2009)

Drescher (2009) argues that graphematic words in Tiberian Hebrew are ‘potential prosodic words’ (p. 98). To arrive at this conclusion, he combines prosodic and morphosyntactic criteria to provide a description of graphematic wordhood.

At a first pass, Drescher posits a two-consonant constraint on graphematic wordhood. He applies this specifically to prepositions (Drescher 2009, 96):

Prepositions that consist of only a single consonant (or consonant plus schwa, depending on whether the schwa is analyzed as inserted by rule or part of the underlying form)... are written as bound prefixes, with no space separating them from what follows.

Drescher (pp. 96, 98) takes this as evidence that one full syllable, *i.e.* CV, where V is not *wa*, is a sufficient but not necessary requirement for graphematic wordhood. I term this the MINIMAL SYLLABICITY constraint. This readily accounts for the graphematic prefixal properties of -ב- *b-*, -ל- *l-*, -כ- *k-* and -ו- *w-*, since, as we have seen (§11.1.1) these morphemes can be analysed at the morphophonemic level as consisting of a single element C. It also accounts for -ה- *h^a-*, not discussed by Drescher, which has underlying /ǎ/.

As noted, however, Drescher understands the minimal syllable constraint to be a sufficient, but not a necessary, condition of graphematic wordhood. The non-necessary character of the constraint is demonstrated, as Drescher points out (p. 97), by the fact that the allomorphs of מִן *min*, מִי- *mi-* and מֵ- *mē-*, are not written as independent graphematic words despite satisfying the constraint. The non-graphematic wordhood of -מֵ- *mē-* is notable in particular, since in its phonetic realisation there is no doubling of the following consonant (Drescher 2009, 97 and §11.1.4 above). -ה- *ha-* is similar in that before gutturals it is lengthened with no doubling of the first consonant of the following morpheme (Drescher 2009, 97–98 and §11.1.4 above).

By contrast, מֵה *mah* /*ma*/, with allomorphs מֵה *māh* /*mā*/ and מֵה *meh* /*me*/, is almost always written as an independent graphematic word, although, as discussed, in two instances it is written as the graphematic prefix -מֵ- *ma-* (Drescher 2009, 98–99 and §11.1.6 above).

The phonetic realisations [me:] ‘from’, [hɔ:] ‘the’ and [mɔ:] ‘what’ are all of the shape CV:, yet orthographically only the last of these has the status of an independent orthographic word.

In order to provide both sufficient and necessary criteria for graphematic wordhood, therefore, Drescher introduces a second, morphosyntactic, criterion, namely that for independent graphematic wordhood, ‘a morpheme must exhibit a certain syntactic-semantic independence’. On these grounds, Drescher asserts, ‘*ma* is a potential prosodic word, *ha-* is not’ (Drescher 2009, 98).

Drescher does not spell out exactly what he means by ‘a certain syntactic-semantic independence’. It is furthermore not clear how providing this additional constraint accords with his assertion that ‘orthographic words are potential prosodic words’ (p. 98), since on his framework, prosodic constraints are not sufficient for the full description of graphematic wordhood.

It is true, however, that the definite article -הַ *ha-* is syntactically tied to its host more closely than -מַ *ma-*, in that, as we have seen, -הַ *ha-* together with its host constitutes a syntactic phrase (§10.2.2), while מַה *mah* can serve as the predicate of a sentence:

(380) = (327) Exod 13:14

מַה־זֶּה ←

mh=z't

what=this

‘What [is] this?’ (KJV)

Although Drescher does not state it explicitly, this constraint is also able to account for the distribution of -מִי *mi-* / -מֵ *mē-*: although these allomorphs meet the criterion of phonological minimality, they fail on the criterion of syntactic-semantic independence, since they introduce a syntactic constituent, the prepositional phrase, in a way parallel to the article -הַ *ha-*, which introduces a determiner phrase.

The distribution of -שֶׁ *še-*, however, constitutes a challenge to Drescher’s hypothesis, since this morpheme satisfies both minimal syllabicity and morphosyntactic independence, yet is always written as a graphematic prefix. First, it satisfies minimal syllabicity, since it consists of a CV unit, where V is not *wa*. Second, it is at least as syntactically/semantically independent as מַה *mah*. Compare (380) with the Qoh 2:26 example quoted earlier, and repeated here for convenience:

(381) Qoh 2:26

לְאָדָם־שָׂטוּב ←

l='dm š=tb

to=person who=good

‘to a man that [is] good’ (KJV)

In both cases, the clitics, respectively **מה** *mah* and **-שׁ** *še-*, are in a subject-predicate relationship with the rest of their clause – in the case of **מה** *mah*, the predicate, and in the case of **-שׁ** *še-*, the subject. Both show the same level of ‘syntactic-semantic independence’.

It could be suggested that **-שׁ** *še-* is written together with the following word on syntactic grounds, as a mark of subordination, since **-שׁ** *še-*, unlike **מה** *mah*, introduces a subordinate clause. However, even where **מה** *mah* does introduce a subordinate clause, in indirect questions, it is still written as a separate orthographic word, e.g.:

(382) Gen 2:19

לְרִאֲוֹת מַה יִּקְרָא לָּו ←

$l=r^{\lambda}wt_{\omega}$ $mh=yqr^{\lambda}=l=w_{\omega}$

to=see what=he_call=to=it

‘to see what he would call it’ (after NAS)

Given, therefore, that **-שׁ** *še-* meets the condition of phonological minimality, and that its syntactic-semantic independence is at least at the level of that of **מה** *mah*, we would expect to find **-שׁ** *še-* written as a graphematically independent word,¹⁷ perhaps **שׁה**, given the spelling of the graphematically independent allomorph of **מה** *mah*, **מה** *meh*, e.g.:

(383) Gen 4:10

וַיֹּאמֶר מַה עָשִׂיתָ ←

$w=y^{\lambda}mr_{\omega}$ mh_{ω} $^{\lambda}sy_{\omega}$

and=he_said what=do.PRF.2SG

‘And he said, “What hast thou done?”’ (KJV)

A further issue is that Drescher’s second criterion provides no insight as to why, in two instances, **מה** *mah* is represented orthographically as **-מ** *ma-*: as remarked at §10.2.7 above, minimal pairs of the syntactic contexts of these can be found written with the graphematically independent form: why should this morpheme be capable of behaving in this way, when its morphosyntactic animate equivalent **מִי** *mī* cannot?

In the light of these observations, a combination of prosodic and morphosyntactic criteria per Drescher (2009) is not sufficient for providing a necessary set of conditions for graphematic wordhood in Tiberian Hebrew. It is therefore worth considering whether a purely graphematic explanation might suffice.

¹⁷ That this is a reasonable expectation is shown by **-שׁ** *še-* being the proclitic mostly frequently written as a separate word in the work of Israeli primary school children learning to spell (Ravid 2012, 112; citing Sandbank, Walden & Zeiler 1995).

11.3. Accounting for graphematic wordhood prosodically

In the absence of satisfactory semantic, morphosyntactic or graphematic accounts of graphematic wordhood in Tiberian Hebrew, or indeed, of a satisfactory combination of prosodic and syntactic constraints, it is worth pursuing the possibility of a purely prosodic explanation. On such an account, for a morpheme to have the status of an independent graphematic word, it must satisfy some minimal prosodic criteria, and, fulfilling these criteria would provide the necessary and sufficient conditions for independent graphematic wordhood. Such a set of criteria can in fact be found at the phonological level.

As we saw earlier, Dresher (2009) suggests that a necessary but not sufficient criterion for graphematic wordhood is that the morpheme meet the minimal shape CV, where V is not *shwa* (or *ḥaṭef*). Showing that this criterion is necessary but not sufficient, however, are various morphemes and allomorphs satisfying this criterion that are nevertheless never written as independent words in the Tiberian tradition, namely:

- -שֶׁ *še-*
- -הָ *ha-*
- -מֵ *mē-*

While these are all CV- or CVV- prefixes at the phonetic level, at the morphophonemic level they are either CV or CVX morphemes. These morphemes therefore share the property of not satisfying foot-binarity, meaning that their lexical representations are not dominated by a prosodic word node. Accordingly they are not introduced into the phonology as independent prosodic words, and are therefore incorporated into the following prosodic word. In other words, while Dresher's minimal syllabicity constraint is necessary but not sufficient, a constraint of phonological bimoraicity is both necessary and sufficient for accounting for the graphematic dependence of the above listed morphemes. -שֶׁ *še-* and -הָ *ha-* are bimoraic, *i.e.* CVX, but are not fully determined at the lexical level, since its second mora is determined by the first consonant of the following morpheme. Finally, -מֵ *mē-* is not a prosodic word, since this is introduced to the phonology via the lexical allomorph /miX/. By contrast, מִן *min* is a possible prosodic word, since this is introduced to the phonology as /min/, *i.e.* CVC, and is thus fully determined at the lexical level.

Minimal graphematic wordhood is therefore isomorphic with minimal prosodic wordhood. This is to say that minimal graphematic words can in principle stand as independent prosodic words, although they need not do so in a particular context. This is consistent with the observation made above that only morphemes of the phonological shapes CVC or CVV can be marked with a disjunctive accent, since only a minimal prosodic word becomes final in a prosodic phrase.

11.4. מַה *mah* 'What?'

It remains to explain the orthography of מַה *mah*: as we saw at §11.1.6, this is best analysed as having two allomorphs, CVV (*i.e.* [mɑː]) and CVX (*i.e.* /maX/), analogous to

the allomorph distribution in the case of מִן *min*. We might expect, therefore, to have two graphematic allomorphs: ⟨m-⟩ for /maX/ and ⟨mh⟩ for [mɑ:]. While ⟨mh⟩ is indeed used to represent [mɑ:], in most cases of /maX/ there is a mismatch between the consonantal and pointed texts: while the consonantal text uses ⟨mh⟩, implying [mɑ:], the pointed text indicates /maX/ by placing *dagesh* on the following consonant. As we saw, however, on rare occasions the consonantal text is consistent with /maX/.

As I argued earlier, the mismatch between consonantal and pointed texts on the representation of מַח *mah* can be seen as a *qere/ktiv*-like distinction. The distinction would have a historical orthographic basis, whereby ⟨mh⟩ represents the original shape of the morpheme. Synchronically, however, ⟨h⟩ would be understood as a *mater lectionis*; /maX/ would then be written ⟨mh⟩ because the morpheme as a whole can stand as an independent prosodic word, and would be viewed as a minimal prosodic word.

If מַח *mah* is written as an independent prosodic word on the grounds that the morpheme can in principle stand as an independent prosodic word, there is no reason in principle why מִן *min* could not also be so treated. This is to say that we might expect to find cases where מִן *min* is written, where, according to the phonological environment, we might expect -מִ *mi-*. We do in fact find such cases:

(384) 2Sam 22:14

יְרַעַם מִן־שָׁמַיִם יְהוָה

yr'm mn šmym yhwh

thundered from heaven DN

'The LORD thundered from heaven' (KJV)

It is not immediately clear how instances like this should be read, whether as /miX/ or as /min/. That the first consonant of the following morpheme, in this case שׁ *š*, does not receive *dagesh* should perhaps be taken to suggest that this sequence should be read [min.ʃɑ:.mɑ:.jim.]. However, it was noted earlier that both מִן *min* and -מִ *mi-* can occur in identical contexts within the space of two verses, as at 1Chr 12:15 and 1Chr 12:17. This shows that, at least in Chronicles, there was at some stage of the tradition somewhat free variation between the /miX/ and /min/ in consonantal contexts, and, as such, reading ⟨mn⟩ as /miX/ is not out of the question.

11.5. לֹא *lōʾ*

The isomorphy of graphematic wordhood and minimal prosodic wordhood provides a way of understanding the otherwise somewhat peculiar orthography of the negative particle לֹא *lōʾ*. This particle is always spelled in Hebrew with a final א׳, despite this א׳ not being a feature of the word etymologically (Zevit 1980, 22). Thus the negative particle in Ugaritic is 𐎎 *l-* /lā/, without ׳ (Pardee 2008, 26). The final א׳ in Hebrew orthography is therefore difficult to account for.

It is helpful to highlight, however, that לֹא $l\bar{o}$ is capable of carrying a disjunctive accent, and, therefore, of concluding a prosodic phrase:

(385) Gen 2:25

← וְלֹא יִתְבַּשְׁמוּ:

($w=l^p$) ($ytbššw$)

and=not be_ashamed.PREF.PASS

‘and they were not ashamed’ (KJV)

In such circumstances, despite a glottal stop not being present in the lexical representation of the word, it is reasonable to assume that a glottal stop may have been pronounced, especially in an environment where the following prosodic phrase started with a vowel. The writing of the glottal stop therefore arises from the possibility of this word terminating a prosodic phrase, that is, of being prosodically independent. Synchronically, the writing of a final glottal stop for the negative particle conveys the fact that, unlike לֹא l , לֹא $l\bar{o}$ is a prosodically independent word, capable of coming final in the prosodic phrase, and, concomitantly, of carrying its own disjunctive accent.

11.6. Minimal domains for stress assignment and sandhi

An important consequence follows from the correspondence between graphematic wordhood and minimal prosodic wordhood. Since the prosodic word is the domain for stress assignment (Dresher 1994, 9), and since each prosodic word carries only one primary stress (Dresher 1994, 9), delimiting minimal prosodic words is tantamount to delimiting *minimal domains for stress assignment*.

Furthermore, external sandhi phenomena in Tiberian Hebrew are a function of accent (Aronoff 1985, 68). Specifically, the prosodic phrase is the domain of the application of the post-vocalic spirantisation of stops after (Dresher 1994, 10).¹⁸ The following example constitutes a single prosodic phrase, and consequently the בּ b of בְּכַפִּי is spirantised:

(386) Judg 12:3

← וְאֵשֶׁתִּי מָה נָפְשִׁי בְּכַפִּי

$w=šymh$ $npš-y$ $b=kp-y$

and=I_put life=my in=hand-my

‘and I put my life in my hand’

Aronoff (1985, 68) gives the following pair of examples involving spirantisation of post-vocalic stops (which are in turn from Rotenberg):

¹⁸ See Dresher (1994) for discussion of other such phenomena. See also Fassberg (2013) for sandhi phenomena in Modern Hebrew.

(387) Judg 1:1

וַיִּשְׁאַלּוּ בְנֵי יִשְׂרָאֵל ←

w=yš^llw bny ysr^l

and=asked sons Israel

‘And the sons of Israel asked’

(388) Judg 1:8

וַיִּלְחַמוּ בְנֵי-יְהוּדָה ←

w=yllhmw bny yhw^{dh}

and=fought sons Judah

‘And the sons of Judah’

Since any prosodic word is necessarily included within a larger prosodic phrase, it follows that spirantisation will apply at morpheme boundaries within a prosodic word. A corollary, therefore, of the isomorphism of minimal prosodic words and orthographic words is that orthographic words are domains in which sandhi phenomena, such as spirantisation, are obliged to apply. Thus the preposition **בְּ-** *b-* induces spirantisation in the **כ** *k* of **כַּף** *kap* in (386). **לְ-** *l-* has the same effect, as may be seen from the following example, constituting a single prosodic word, where **לְ-** *l-* induces spirantisation in the **כ** *k* of **כַּף** *kap* ‘palm (of the hand), sole (of the foot)’:

(389) Gen 8:9

לְכַף-רַגְלָהּ ←

l=kp=rgl-h

for=sole=foot-her

‘for the sole of her foot’ (KJV)

However, spirantisation is blocked by the boundaries of a prosodic phrase, the edges of which are indicated by disjunctive accents (Dresher 1994, 10). Consider the following example, where the disjunctive accent on **נַפְשִׁי** *npš-y* marks the end of its prosodic phrase. Accordingly, the **ב** *b* in **בְּכַפִּי** *b=kp-y* is not spirantised. Together with (386), the example constitutes a minimal pair:

(390) 1Sam 28:21

וְאָשִׁים נָפְשִׁי בְּכַפִּי ←

w=š^ym npš-y b=kp-y

and=I_put life-my in=hand-my

‘and I have put my life in my hand’ (after NAS)

We have seen that graphematic words in Tiberian Hebrew may terminate both a prosodic word and a prosodic phrase. Any graphematic word, therefore, may or may not induce sandhi phenomena, such as spirantisation. By contrast, morphemes that cannot conclude a prosodic word or phrase, such as -ב *b*-, -ל *l*- and -כ *k*-, that are therefore obligatorily written together with any following word, must induce relevant sandhi phenomena. A corollary, therefore, of orthographic words mapping to minimal prosodic words is that, as well as constituting *minimal domains for stress assignment*, orthographic words constitute minimal domains for external sandhi.

11.7. Conclusion

I have argued in this chapter that graphematic word division in Tiberian Hebrew targets minimal prosodic words. This is to say that the orthography separates units that are capable of standing as independent prosodic words. This is in fact the same conclusion arrived at by Drescher (2009). In the present case, however, this has been via an entirely prosodic/phonological route, without the grammar of the writing system requiring direct reference to morphosyntax, as required by Drescher. Minimal prosodic words are distinguished from ACTUAL PROSODIC WORDS in context, which in Part I and Part II I have argued to be the target of graphematic word division in early Phoenician inscriptions, as well as in the Ugaritic ‘Majority’ orthography. As Drescher (2009) describes, minimal prosodic words may optionally be incorporated into neighbouring prosodic words, but are not required to do so.

Chapter 12

Minimal prosodic words in epigraphic Hebrew and Moabite

12.1. Introduction

In the previous chapter it was established that graphematic word division in the consonantal text of the Hebrew Bible corresponds to the separation of MINIMAL PROSODIC WORDS. This principle of word division is distinguished from strategies targeting ACTUAL PROSODIC WORDS in context, seen in most early Northwest Semitic inscriptions, by being a more abstract prosodic representation. It is distinguished too from morphosyntactic word division, as seen in the Ugaritic ‘Minority’ orthography, by the fact that the separation of graphematic sequences does not correlate with their morphosyntactic status, but rather with (minimal) prosodic status. It has been observed in the literature that this word division strategy is of considerable antiquity, occurring already in the pre-Masoretic consonantal text of the Hebrew Bible (Aronoff 1985, 47). The goal of the present chapter is to establish how far back this more abstract prosodic principle of word division might be said to go.

Although the Masoretic Text is provided to us in codices from the early medieval period, the antiquity of the word division strategy that we see in the Masoretic Text can be seen in the Biblical documents from the Dead Sea. To illustrate the almost identical orthographies of the Masoretic Text and the consonantal text as represented in the Dead Sea Scrolls, compare the opening of the book of Isaiah in 1QIsa^a and the Masoretic Text:

(391) Isa 1:1 1QIsa^a (text Parry & Qimron 1999)

חזון ישׁיהוּ בן אמוץ אשר חזה על יהודה וירושלם בימי עזיה ←
יותם אחז חזקיה מלכי יהודה

ḥzwn yš'yhw bn 'mwš 'šr ḥzh
vision Isaiah son Amos which saw

ʿl yhwdh w=yrwšlm b=y'my 'zyh
about Judah and=Jerusalem in=days Uzziah

ywtm	ʕhz	ʕhzqyh	mlky	yhw dh
Jotham	Ahaz	Hezekiah	kings	Judah

(392) Isa 1:1 Masoretic Text (text Westminster Leningrad Codex)

← תִּזוֹן יִשְׁעֵיהוּ בֶן־אָמוֹץ אֲשֶׁר הָיָה עַל־יְהוּדָה וִירוּשָׁלַם בְּיָמֵי עֲזִיָּהוּ
 יוֹתָם אָתוֹ יְחִיקָהוּ מִלְכֵי יְהוּדָה:

hzw n	yš' yhw	bn	ʕmws	ʕsr	hzh
vision	Isaiah	son	Amos	which	saw
ʔ	yhw dh	w=yrwšlm	b=y my	ʕz yhw	
about	Judah	and=Jerusalem	in=days	Uzziah	

ywtm	ʕhz	yhzqyh w	mlky	yhw dh
Jotham	Ahaz	Hezekiah	kings	Judah

'The vision concerning Judah and Jerusalem that Isaiah son of Amos saw during the reigns of Uzziah, Jotham, Ahaz and Hezekiah, kings of Judah' (NIV)

Note in particular for our purposes the identical treatment of the prefixes w- and b- in w=yrwšlm 'and=Jerusalem' and b=y my 'in=days' respectively. Furthermore, morphemes joined by *maqef* in the Masoretic Text are graphematically independent in 1QIsa^a. This includes items in construct, such as בֶן־אָמוֹץ $\text{bn } ʕmws$ vs. בֶן אָמוֹץ $\text{bn } ʕmws$ 'son of Amos', as well as prepositional phrases, such as עַל־יְהוּדָה $ʔ=yhw dh$ vs. עַל יְהוּדָה $ʔ yhw dh$ 'about Judah'. Finally, nouns in construct joined by conjunctive accents are also graphematically separated, e.g. מִלְכֵי יְהוּדָה vs. מִלְכֵי יְהוּדָה mlky yhw dh .

1QIsa^a has been dated paleographically to 125–100 BCE (Ulrich 2015). At least from a graphematic point of view, the same word division strategy must go back at least that far. How much further back can it be traced? This is the question addressed in the present chapter.

Word division in epigraphic Hebrew presents a different set of challenges from those raised by the Tiberian tradition. In particular, since word division is by means of a point, which is easily lost, in the case of many of the ostraca it is in principle difficult to know whether a word division point has been lost, or was never there in the first place (see Millard 2012b, 25). There are, however, cases, particularly inscriptions, where word division has been well preserved.

In §12.2 I consider the word division orthography of the Siloam Tunnel inscription, dated to some time in the 8th century BCE (Sasson 1982). There too a word division orthography is found that is identical to that seen in the consonantal text of the Masoretic Text. Then in §12.3 I turn to the Moabite inscriptions commissioned by Meshāʕ, king of Moab, from the second half of the 9th century BCE, where once again we find a near identical orthography of word division, although with one significant difference.

In §12.4 I seek to account for the word division strategy that we find in these inscriptions. I argue on typological and syntactic grounds that the most likely explanation is that word division targets MINIMAL PROSODIC WORDS, not only in Tiberian terms, but also in terms of the Hebrew and Moabite languages of the time.

12.2. Siloam Tunnel inscription

As mentioned, the Siloam Tunnel inscription is dated to the 8th century BCE (Sasson 1982; for context and further details see Rendsburg & Schniedewind 2010). In the literature, the inscription is regarded as having consistent word division (Naveh 1973b, 207). In fact, the word division orthography turns out to be identical to that of the consonantal Masoretic Text, despite the fact that it antedates our first texts of the Bible by many centuries.

Monoconsonantal prepositions, viz. $\text{-לֹ} l\text{-}$, $\text{-בֹ} b\text{-}$, the definite article $\text{-הַ} ha\text{-}$, and the conjunction $\text{-וַ} w\text{-}$ are all consistently univerrated with the following morpheme, just as we see in the consonantal text of the Masoretic Text, e.g.:¹

(393) $\text{הַנֶּחֱבֵה} h\text{-}nqbh$ [the-tunnel] (KAI⁵ 189.1)

(394) $\text{בְּיַד} b\text{-}wd$ [in=continuance] ‘while’ (KAI⁵ 189.1)

(395) $\text{בְּסֵר} b\text{-}sr$ [in=rock] (KAI⁵ 189.3)

(396) $\text{וַזֶּה} w\text{-}zh$ [and=this] (KAI⁵ 189.1)

(397) $\text{וַאֲלֵפ} w\text{-}lp$ [and=thousand] (KAI⁵ 189.5)

The similarity with the consonantal orthography of the Masoretic Text extends even further, to the univervation of $\text{מִן} min$, including the sandhi assimilation of final $/\text{-n}/$ (cf. §3.5 and §11.1.4 above):

(398) $\text{מִן־הַיָּם} m\text{-}ymn$ [from=south] (KAI⁵ 189.5)

Note too the univervation of clitic combinations:

(399) $\text{לְהַנְּחִיב} l\text{-}h\text{-}nq[b]$ [to=the-tunnel] (KAI⁵ 189.2)

(400) $\text{וַבַּיּוֹם} w\text{-}b\text{-}ym$ [and=on=day] (KAI⁵ 189.3)

By contrast, multi-consonantal prepositions are consistently separated from the following morpheme, e.g.:

(401) $\text{אֲשֶׁר־לְרֹאֲשֵׁי} l_{(w)} r\text{-}w_{(w)}$ [to_{(w)}} associate-his_{(w)}}] (x2, KAI⁵ 189.2, 3)

¹ The text followed is that of Sasson (1982).

The word division strategy is not unique to KAI 181. It also appears to be applied to the other (much shorter) inscription from Mesha^c's reign:

(416) KAI 306.1

[]אאֶגֶפֶת־מֶלֶךְ־מְבָרַח

*k]msyt*_{<ω>} *mlk*_{<ω>} *m'b*_{<ω>} *h-d[]*_{<λ>}

PN king TN the-TN-ite

'*Kmšyt*, king of Moab, the TN-ite'

In KAI 306 we see:

- The univerbation of the definite article *h-* with the following morpheme in *h=d[]*;
- Separation of the construct chain elements *mlk · m'b* 'king of Moab'.

12.3.2. Treatment of heavy suffix pronouns

There is, in fact, only one significant⁴ respect in which the word division strategy differs from the attested Tiberian Hebrew tradition, namely in the writing of the suffix pronoun אֶת־ *-hm* as an independent graphematic word (Segert 1961, 236):

(417) KAI 181.18

וַיִּדְרֹגְנִי־לְפָנֵי־יְהוָה־בְּמֹשֶׁה ←

*w=ʔšhb*_{<ω>} *hm*_{<ω>} *lpny*_{<ω>} *kms*_{<ω>}

and=I_dragged them before DN

'And I dragged them (*i.e.* the articles of Yahweh) before Kemosh'

Compare this instance with that at lines 12–13, where the same verb occurs with the light third-person masculine singular suffix, but this time joined with the pronoun:

⁴ The word division orthography also differs from that of Tiberian Hebrew in two minor respects: a) (l. 1) *bn · kms[yt]mlk · m'b* 'son of *Kmšyt* king of Moab', where there is no word divider between *kmš[yt]* and *mlk*; and b) (l. 4) *ky=hs'=ny* 'because he rescued me'. However, there is reason to believe that neither of these are in fact genuine deviations from the orthography we find in the consonantal text of the Masoretic Text: the first is found at a break in the inscription, while in the case of the second a word divider is actually visible in Lidzbarski's drawing of the inscription.

(418) KAI⁵ 181.12–13

[𐤀]𐤏𐤃 ←
 .𐤏𐤏𐤏𐤏𐤏𐤏𐤏𐤏𐤏

w=²[s]_{<λ>} ḥb-h_{<ω>} lpny_{<ω>} kmš_{<ω>}
 and=I_dragged-him before DN

‘And I dragged him (i.e. 𐤏𐤏) before Kemosh’ (for the interpretation see Lipiński 1971, 332)

In the literature, the separation of the pronoun in l. 18 has been regarded as a curiosity of limited significance.⁵ Thus Lehmann (2016, 44 n. 2) writes:

as yet the separation of this suffix seems to be the exception that still longs for a different explanation as [i.e. than] being a surplus ‘highlighting’ divider.

It turns out, however, that this orthography is not only explicable, but expected. In the case of consonantal Hebrew, I argued above that two factors are necessary for a word to be written together with a host, rather than independently, namely subsyllabicity and the inability to carry stress under any circumstances. It is noteworthy, then, that in Tiberian Hebrew, the ‘heavy’ suffix pronouns, that is, second- and third-person masculine and feminine plural, 𐤏𐤏- *-kem*, 𐤏𐤏- *-ken*, 𐤏𐤏- *-hem* and 𐤏𐤏- *-hen*, respectively, always take the accent, rather than the lexical word to which they are suffixed (van der Merwe, Naudé & Kroeze 2017, 93). These pronouns therefore serve as the prosodic hosts for the content words to which they are suffixed (cf. Kuryłowicz 1959, who argues that in origin the Hebrew verb is proclitic).

Indeed, it is very occasionally written as an independent graphematic word even in the Masoretic Text. Normally, the preposition 𐤀𐤏 ‘d suffixes pronouns to the base

⁵ Thus Donner & Röllig (1968, ad loc.), citing Segert (1961, 217–218, 236), state ‘Beachte, da 𐤏𐤏 durch einen Worttrenner vom Verbum getrennt ist; doch ist dies wohl nur eine graphische Besonderheit’. (Translation: ‘Note where *-hm* is separated by means of a word divider; though this is of course only a graphical anomaly.’) Segert (1961, 217 n. 93) notes that ‘Lidzbarski: K, S. 8 halt diese Abtrennung mit Recht für graphisch und äußerlich’. (Translation: ‘Lidzbarski, K, p. 8 rightly takes this separation as graphical and superficial.’) Cf. Andersen (1966, 97): ‘The writing of *hm* as a separate word does not prove that it was a free form syntactically, although it suggests it. Only the occurrence of one or more words between the verb and such an object would settle the point. In the absence of such evidence we cannot say that here we have anything more than an orthographic convention similar to that found in Ugaritic and Aramaic texts. It is the verb as a free form, not the suffix, that is indicated by the dot. A comparable usage is the traditional convention of the scripts derived from the original Phoenician of writing prepositions (all of which are probably proclitic) as part of the following word if they are written with one consonant, but as a distinct word if they are written with more than one consonant. Beyer (2012, 117) states Such a statement begs the question of the difference between a suffix and a pronoun, which in turn presupposes an understanding of the difference between a suffix and a word. Beyer accents with two stresses as follows: *ʾašōb-hʾema*.’

עֲדִי ^cdy (van der Merwe, Naudé & Kroeze 2017, 369). However, in the following example it is written as an independent graphematic word before םה- *hm*:

(419) 2Kgs 9:18

בָּאֵלֶּיָם לְאֵדֵי עֲדִי ←

$b^{\prime}=h\text{-}ml^{\prime}k_{\omega}$

$d^{\prime}=hm_{\omega}$

came=the-messenger up_to=them

‘The messenger came up to them’

Moabite in fact adheres more closely to the orthographic principle of separating (Tiberian) minimal prosodic words than Tiberian Hebrew itself: since םה- *hem* can take the accent, it is a minimal prosodic word; furthermore, since the *wayyiqtol* form וַאֲצַחֲב $w^{\prime}shb$ serves as a prosodic host for the enclitic pronoun ה *h* in ll. 12–13, וַאֲצַחֲב $w^{\prime}shb$ is also a minimal prosodic word. Therefore, since both are minimal prosodic words, they should be written separately. It is therefore the Tiberian strategy of writing the ‘heavy’ pronouns together with their verbs that is unexpected and requiring of explanation.

12.4. Accounting for word division in the Mesha^c and Siloam inscriptions

From a purely graphematic perspective, the word division orthography of the consonantal text of the Hebrew Bible is very ancient. So far we have tacitly assumed that word division in the Siloam and Mesha^c inscriptions targets minimal prosodic words in the same way that the consonantal text of the Hebrew Bible appears to do in Tiberian terms. The goal of the present section is to assess how likely this is to be the case.

A priori such a claim might seem far-fetched. After all, the Tiberian cantillation tradition is only directly attested from the early medieval period, long after the inscriptions we have been considering were written. However, in favour of such a view are the following considerations. First, the cantillation tradition in Tiberian Hebrew, although directly attested only late, is itself believed to have very ancient roots (§1.7.4.1; Revell 1971, 222; Khan 2020, 51; cf. also Revell 1976). Secondly, we have observed in the foregoing chapters the isomorphy of prosodic words in Tiberian Hebrew and graphematic words in the Ugaritic ‘Majority’ orthography, as well as graphematic words in the Byblian and Phoenician royal inscriptions. We have used this, along with morphosyntactic and typological arguments, to suggest that graphematic words in Ugaritic and Phoenician correspond to prosodic words in these languages. The corollary of this isomorphy that prosodic words in Ugaritic and Phoenician had a similar distribution to prosodic words in Tiberian Hebrew. Given such a parallel distribution of ACTUAL PROSODIC WORDS in Ugaritic, Phoenician and

Tiberian Hebrew, it seems highly plausible that ACTUAL PROSODIC WORDS in Moabite and Archaic Hebrew would have also had a similar distribution. In such a scenario, an orthography that targeted MINIMAL PROSODIC WORDS would be expected to look just like the orthographies that we see in the Moabite and Hebrew inscriptions we have been considering.

To these language-specific considerations can be added the following typological and syntactic arguments. First, it is clear that graphematic words in the Mesha' stela and the Siloam tunnel inscriptions cannot target morphosyntactic entities, for exactly the same reasons that we rejected this hypothesis in the case of Tiberian Hebrew, chief among which is that the treatment of prepositions depends on graphematic/syllabic weight, and not on morphosyntactic status. Second, the consistency of the separation of graphematic words rules out a prosodic explanation, since, cross-linguistically, it would be expected for longer function words to incorporate with neighbouring morphemes on a sporadic basis. What we see in these inscriptions, however, is that longer function words *never* incorporate with neighbouring morphemes. For precisely the reasons advanced in the case of Tiberian Hebrew, therefore, even without the added information of the system of accents to which we have access in the Tiberian tradition, the profile of word division in the orthography matches what one might expect if graphematic words targeted MINIMAL PROSODIC WORDS.

A final consideration in favour of a MINIMAL PROSODIC WORD analysis of word division in Moabite is precisely the separation of the heavy suffix pronoun 𐤁^{h} -*hem* discussed in §12.3.2. In this respect the Moabite orthography was found to conform even more closely to that expected on a Tiberian basis than the consonantal text of the Hebrew Bible.

12.5. Conclusion

In this chapter I have sought to demonstrate that the principle of word division that we see in the consonantal text of the Hebrew Bible is very ancient, dating to at least the 9th century BCE. I have argued that the most economical way to account for this orthography is to interpret graphematic words as corresponding to MINIMAL PROSODIC WORDS, just as they do in Tiberian Hebrew.

12.6. Conclusion to Part III

In Part III of the study I set out to establish the linguistic domain of word division in the consonantal text of Hebrew Bible. After discussion of the morphosyntactic status of graphematic affixes in Tiberian Hebrew, as well as proposal, along with ultimate rejection, of morphosyntactic and graphematic explanations for graphematic wordhood (Chapter 10), I set out to explore a prosodic explanation in Chapter 11.

I started by considering graphematic affixes in terms of minimal prosodic wordhood. After this I considered Drescher's proposal for a combination of prosodic and morphosyntactic constraints (Drescher 2009). This was rejected as it was found not to be able to account for the observed phenomena. Instead I argued that the principle governing word division in the consonantal Masoretic tradition is the identification of minimal prosodic words, that is, morphemes that are fully determined in the shape CVC at the lexical level. Since heavy suffix pronouns are of this shape, but are not written as independent graphematic words, I proposed that a morpheme must be marked explicitly as a prosodic word in the lexicon.

In view of the antiquity of the consonantal and prosodic traditions of Tiberian Hebrew (cf. §1.7.4.1), I argued that it is reasonable to suppose that the correspondences we have observed between orthographic wordhood and minimal prosodic wordhood are not simply artefacts of the medieval tradition, but have a genuine basis in the prosody of Hebrew in antiquity. From this starting point I considered two very early instances of the same or very similar word division strategies being employed, namely, the Mesha^c stele, and the Siloam tunnel inscription. I argued on comparative, typological and morphosyntactic grounds that the most economical explanation for the word division orthographies of these inscriptions is that they too separate MINIMAL PROSODIC WORDS.

PART IV
Epigraphic Greek

Chapter 13

Introduction

13.1. Overview

The claim in Parts I to III has been that word division in a significant proportion of early Northwest Semitic inscriptions targets prosodic units. In most cases this is at the level of the prosodic word, but it is also possible to find demarcations at the level of the prosodic phrase (KAI 10). Unlike Northwest Semitic, the suggestion that word division is prosodic in an important subset of Archaic and Classical Greek inscriptions is not new, and might be said in fact to represent the consensus (§1.6.2). For our purposes, the fact that a contemporary and related writing system demarcates word-level units on the basis of prosody adds support for the idea that prosody might govern word division in Northwest Semitic.

There are, however, *prima facie* difficulties with identifying graphematic words and prosodic words in the Greek epigraphic material, to do with the disparity between what is expected on grounds of pitch accentuation and lexical status. These have not to date been addressed head-on (§13.5). Yet doing so has the potential to pay dividends in terms of further refining our understanding of the purpose of word division, and how this practice is interpreted in different linguistic circumstances. The argument I wish to put forward is that graphematic words in Greek inscriptions are best seen not as accentual units, as has often been suggested (Kaiser 1887; Morpurgo Davies 1987; Wachter 1999), but rather rhythmic units.

Although I consider here principally the demarcation of *word*-level units, this is not to deny that a wide variety of word division practices existed in the Ancient Greek world, ranging from not punctuating at all to punctuating prosodic phrases and even larger units (Larfeld 1914, 303; Devine & Stephens 1994, 388–390; Wachter 1999). We have also seen such variation in Phoenician inscriptions of broadly comparable date (Chapter 4). However, for reasons of scope, it is not possible to include analyses of these types here.

Before embarking, I first itemise the inscriptions that I will be using to describe word division in Ancient Greek (§13.2). I then outline the basis for identifying prosodic words in Ancient Greek on grounds independent of epigraphic word division (§13.3).

13.2. Corpus

The study will be conducted on the basis of the analysis of a small set of inscriptions,¹ detailed in Table 13.1.²

Table 13.1: Inscriptions considered in regard to word-level punctuation in Greek

<i>Inscription</i>	<i>City</i>	<i>Approx. date BCE</i>	<i>Source text used</i>
SEG 14.604	Ischia, Italy	late 8th	Watkins (1976)
SEG 23.530	Dreos, Crete	650–600	Woodhead (1968)
SEG 11.314	Argos	575–550	Probert & Dickey (2015)
IG I ³ 5	Eleusis, Attica	500	Lewis (1981)
IG I ³ 699	Athens, Attica	500–480	Lewis & Jeffery (1994)
IG I ³ 775	Athens, Attica	500–480	Lewis & Jeffery (1994)
IG I ³ 1084	Athens, Attica	end 5th	Lewis & Jeffery (1994)
IG IV 554	Argos	6th or early 5th	Jameson (1974)
IGA 497 (= SEG 4.616)	Teos, Ionia	5th	Roehl (1882)
Peek (1957) 207	Attica	4th	Peek (1957)
IGA 42	Olympia, Elis		Roehl (1882)
IGA 499	Ephesus, Ionia		Roehl (1882)

As the list immediately makes clear, word-level punctuation is found right across the Greek-speaking world in the first half of the 1st millennium BCE. This includes the very earliest known Greek inscription with punctuation, namely, the Nestor's Cup inscription (SEG 14:604) (Wachter 2010, 53). This suggests that, although the practice of punctuating at the word level is most commonly found in texts dated to the 6th and 5th centuries BCE and is locationally restricted (Morpurgo Davies 1987, 270), it is a practice that likely goes back to the inception of the Greek alphabet itself.

¹ It is important to be careful about the texts used. IGA 497 has been lost (Morpurgo Davies 1987, 270). More recent editions, however, are unreliable in their reading of punctuation, with both Schwyzler (1923) and Meiggs & Lewis (1969) not infrequently missing punctuation before lexicals (see further Morpurgo Davies 1987, 277 n. 21). It is therefore often important to go back to the 19th-century editions of Boeckh (1828) and Roehl (1882) (Morpurgo Davies 1987, 277 n. 21).

² Sources of dating information: for SEG 14:604, Faraone (1996, 77); for SEG 23.530, Gagarin & Perlman (2016, 209); for SEG 11.314, Probert & Dickey (2015); for IG IV 554, Buck (1998[1955], 284); for IGA 497, Morpurgo Davies (1987, 270). For Peek (1957) 207 and IG I³, dates are provided as listed under each inscription at <https://inscriptions.packhum.org/>, accessed July and August 2021.

13.3. Prosodic wordhood in Ancient Greek

Various features have been associated with prosodic words in Greek in the literature, notably several kinds of junctural phenomena (Golston 1995, 346–347; Agbayani & Golston 2010, 157; Vis 2013). We have seen junctural phenomena such as nasal assimilation in Northwest Semitic languages, notably Phoenician (§3.5, cf. §1.4.2.2). However, Ancient Greek attests a wider range of these phenomena than ancient Northwest Semitic languages. This is in large part because vowels are obligatorily written in the former in contrast to the latter. Accordingly, the following junctural phenomena are associated with prosodic wordhood in Ancient Greek: vowel coalescence known in traditional grammar as ‘crasis’ (de Haas 1988), vowel deletion, known in traditional grammar by the terms ‘elision’ and ‘apocope’ and nasal assimilation.

Greek also provides the following further sources of information on the prosodic word:

- The distribution of words around canonical breaks in metrical compositions, known as ‘caesurae’;
- Constraints on where words are permitted to end, also in metrical compositions, known as ‘bridges’;
- The formation discontinuous syntactic structures in Ancient Greek, traditionally known as ‘hyperbaton’; these have been argued to be driven by phonology (Agbayani & Golston 2010). Insofar as prosodic words are minimal prosodic constituents, at least for the purposes of hosting clitics, such rearrangements of the syntax can reveal useful information on prosodic word boundaries (Goldstein 2016).

13.3.1. Vowel coalescence (*crasis*)

As in many languages, prosodic words in Greek are characterised by vowel coalescence at prosodic word-internal morpheme boundaries. However, unlike the Northwest Semitic languages so far considered here, Greek provides very good evidence of the phenomenon (Sihler 1995, 232, §239.4) since the writing system stays very close to the surface phonology, especially in terms of representing vowels.³ Since writing systems for Northwest Semitic languages for the most part eschew the writing of vowel phonemes, any vowel coalescence that may have occurred cannot be observed.

The following provides an example of the phenomenon (a comprehensive list of examples in Attic inscriptions can be found in Threatte 1978, 174–175):

(420) TAΘENAI AI *tathēnaíai* < *têi=athēnaíai* [the.DAT=DN.DAT] (IG I³ 699.A2, 500–480 BCE)

³ Vedic and Classical Sanskrit also provide excellent evidence of vowel coalescence. For sandhi rules in Classical Sanskrit see Macdonell 1927, 10–32.

(421) IGA 42 (Olympia)

ΚΑΡΓΕΙΑΔΑΣ : ΗΑΓΕΛΑΙΔΑ : ΤΑΡΓΕΙΟΥ

*kargeiádas*_{<ω>} *hagelaída*_{<ω>} *targeíou*_{<ω>}*kai=Argeiádas* *ho=Agelaída* *toû=Argeíou*

and=PN the=PN.GEN the.GEN=TN.ADJ.GEN

Devine & Stephens (1994, 268) take the domain of vowel coalescence in Ancient Greek to be what is in our terms the (recursive) prosodic word (see §1.4.2.3). It does, however, seem to have been restricted to particularly cohesive or frequent units within the recursive prosodic word (Devine & Stephens 1994, 268; citing Ahrens 1891[1845]).

13.3.2. Vowel deletion (elision)

Vowel deletion, also known as elision, is also a characteristic of prosodic wordhood in Greek:

(422) : ΗΕΠΙΔΙΩΤΗΙ : *ēpidiōtēi*_{<ω>} (IGA 497, A.3)

This is the equivalent of the following:

(423) *ē=ep'=idiōtēi*_{<ω>} [or=on=private] (IGA 497, A.3)

The same inscription provides evidence that elision does not occur across prosodic word boundaries:

(424) IGA 497, A.10 (Teos; 5th century BCE)

: ΗΕΣΑΧΘΕΝΤΑ : ΑΝΩΘΕΟΙΗ :

‘or push back what has been brought in’

Note the interface of two lexicals, separated by a word divider. The fact that the final /a/ of ΗΕΣΑΧΘΕΝΤΑ *ē=esakhthénta* is not elided before the initial /a/ of ΑΝΩΘΕΟΙΗ *anōtheoíē* is consistent with the domain of elision being the prosodic word.

13.3.3. Consonant assimilation

Consonant assimilation, that is, where features of a consonant, such as [+labial], are assimilated to those of an immediately following consonant, can be seen in the following example from Teos (for discussion of the phenomenon more broadly in Greek, see Devine & Stephens 1994, 305; Sihler 1995, 232, §239.6):

(425) IGA 499, 1–2 (Ephesus)

: HMMEN : ΑΠΟΚΡΥΨΕ

[ΙΔΕ]ΞΙΟΣ : ΗΝΔΕ : ΕΠΑΡΕΙ :

$\grave{\epsilon}m=m\grave{\epsilon}n_{\langle\omega\rangle}$	<i>apokrípse</i> _{⟨λ⟩} [<i>i</i>] _{⟨ω⟩}	<i>dé]xios</i> _{⟨ω⟩}	$\grave{\epsilon}n=d\grave{\epsilon}_{\langle\omega\rangle}$
if=PTCL	hide.FUT.IND.ACT.3SG	the_right.NOM	if=PTCL

*epárei*_{⟨ω⟩}

lift.FUT.IND.ACT.3SG

‘if the right hides, but if it lifts ...’

In the example HMMEN and ΗΝΔΕ provide a minimal pair in terms of the final consonant of the conditional particle HN *én* ‘if’. HMMEN involves the nasal assimilation of /n/ to /m/ in HN *én* before MEN *mén*. Before the dental /d/, however, HN *én* retains /n/.

13.4. Metre and natural language

Considerable evidence for the properties of prosodic words in Ancient Greek comes from metrical compositions. This might be viewed as problematic (cf. discussion at Devine & Stephens 1994, 100–101). After all, poetic metres such as the iambic trimeter or the dactylic hexameter are highly constrained verse forms, where each line must conform to a very precise structure, in a way that is quite different from natural speech. The key issue, however, is not whether people generally speak in metre – they don’t – but whether the language metrical compositions is sensitive to prosodic units in the same way as spoken language. The latter statement does indeed seem to be true: metrical compositions across the world’s languages are sensitive to the phonological status of morphemes beyond the segmental level (Fortson 2008, 9). Accordingly, important facts about the prosodies of both Greek and Latin have been inferred on this basis in recent years (see *e.g.* Devine & Stephens 1994; Fortson 2008; Goldstein 2010). Indeed, it could be argued that the highly constrained nature of metrical compositions makes exceptions to the rules that exist all the more significant, and have the potential to tell us much more about prosody than prose compositions, where there are far fewer rules.

13.5. Problems with identifying graphematic words with prosodic words

Whilst most of the literature tacitly assumes that graphematic words correspond to prosodic words in word-separating inscriptions, this is doubted in some quarters.⁴

⁴ Thus Goldstein (2010, 56) is somewhat circumspect in his identification of graphematic words with prosodic words in IG I³ 702, stating that ‘the function of these triple-punct markers remains somewhat

The following problems with the correspondence between graphematic words and prosodic words in Ancient Greek inscriptions can be identified:

- A discrepancy between the distribution of (non-)accented forms and graphematically dependent morphs, including in terms of clitic polarity (§13.5.1);
- A discrepancy between the distribution of lexical words and graphematic words (§13.5.2);
- Graphematically dependent morphs for the most part have to be trimoraic rather than bimoraic, *i.e.* larger than the supposed minimal prosodic word (§13.5.3);
- Monomoraic function words receive the pitch accent (§14.1).

Before these issues are outlined, it is first important to survey some difficulties in identifying graphematic words in the first place.

13.5.1. Graphematic word and pitch accentual domains are heteromorphic

13.5.1.1. Introduction

Absent from the list of features used in the identification of prosodic words (§13.3) is the position of the accent. This might be surprising given that cross-linguistically prosodic words are characterised by having a single primary accent (§1.4.2.2). In the present study the position of the accent in Hebrew is important for identifying prosodic words there, and, by extension, in Phoenician and Ugaritic. Although Vis and Golston do not give the pitch accent as a diagnostic of prosodic wordhood, it is brought into consideration in other scholars' discussions of the clitic group/(recursive) prosodic word in Greek, *e.g.* Klavans (2019[1995], 134–135), Goldstein (2016, 49–60) and Devine & Stephens (1994).

It should be noted at the outset that the accent in Greek is not necessarily a direct correlate of the accent in the Semitic languages we consider elsewhere in this study. This is because Ancient Greek, for the most part, had a pitch accent, whereas Northwest Semitic languages are assumed to have had a stress accent (cf. for Tiberian Hebrew, Suchard 2019, 85–86; Khan 2020, 268, *passim*; for Ugaritic, Tropper 2012, 88–89; for Proto-Semitic in general and references, see Suchard 2019, 86–89).⁵

The tonal nature of the Ancient Greek accent deserves to be emphasised, since several studies in recent years refer to the accent as one of stress, *e.g.* Klavans (2019[1995], 134–135), Goldstein & Haug (2016, 301–302). Yet there is evidence cross-linguistically that the nature of the accent, whether stress, pitch or pitch differentiated stress, carries important consequences for its distribution (§15.3.2).

It is therefore important to understand the relationship between prosodic wordhood and pitch accentuation: if graphematic words correspond to prosodic words, an account needs to be provided for cases where graphematic word divisions contradict the demarcation(s) expected on pitch accentual grounds. Before exploring

opaque' (cf. Goldstein 2016, 67 in regard to IG I³ 775).

⁵ For evidence that Thessalian had a stress accent, see Probert (2006, 73–74) and Chadwick (1992).

this, however, first a brief overview of the history of the system of accentuation as we have received it is in order.

13.5.1.2. History of the orthographic accent

The system of orthographic accentuation that we find in modern printed texts is likely to have originated in the post-classical period, specifically in 2nd-century BCE Alexandria (Probert 2006, 15, 21). Ancient tradition has it that it was invented by the librarian there, Aristophanes of Byzantium (Probert 2006, 21). While it is difficult to know for sure if it was indeed he who was responsible, the fact we have our first extant papyri with accent marks from the 2nd century is consistent with invention around this time (Probert 2006, 22).

The fact that our system of orthographic accentuation in Greek has its roots in the post-classical period rather than that of the Classical language raises the question of what such a system might have to tell us about the system of accentuation in the Classical and Archaic periods (cf. Probert 2006, 25–26). Indeed, it might be supposed that such an accentuation system might in fact have very little to tell us about the prosody of the Classical language, given the view in some quarters that the Alexandrian system of accentuation was pure invention (for the existence of this view, see Probert 2006, 26; see also the discussion in Goldstein 2010, 50–54).

That the system of accentuation does reflect prosodic reality is indicated by the general agreement in lexicals on the position of the accent with modern Greek cognates (Probert 2006, 26–27). Although for the most part Alexandrian accentuation may be said to reflect the Koine variety spoken by the Alexandrian grammarians (Probert 2006, 70), the agreement in accentuation with cognates in other Indo-European languages, notably Sanskrit, with which the Alexandrian grammarians were not familiar, shows that in principle the prosody underlying the orthographic accentuation reaches back to before the post-classical period (Probert 2006, 26). In addition, there is evidence that the Alexandrian grammarians derived at least some of their information on accentuation from oral recitation traditions, in addition to the language of their own time (West 1981; Probert 2006, 33–45). Finally, there is statistical evidence that, in addition to rhythm, the position of the accent played an important role in Homeric poetry (Abritta 2015; 2018). That Homeric poetry has been shown to be sensitive to accentuation according to the Alexandrian tradition provides support for the validity of that tradition for Homeric Greek, which must mean that to a considerable degree, at least, the accentual system described by the Alexandrians was also that of Homeric Greek.

The system of accentuation described by the Alexandrian grammarians can therefore be said to have its roots deep in antiquity, and, crucially, may be said to reach back prior to the variety of Greek spoken by those grammarians. It is important, therefore, to deal seriously with instances in Classical Greek inscriptions where graphematic word division does not correspond to what one would expect to find if

graphematic words demarcate prosodic words. Such cases fall into the following three categories:

- Tonally enclitic forms written as orthographically independent words, demarcated to the left as well as, potentially, to the right (§13.5.1.3);
- Orthotonic words or morphs written as graphematically dependent forms (§13.5.1.4);
- Lack of agreement between actual and traditional clitic polarity (§13.5.1.5).

13.5.1.3. Graphematic independence of enclitic forms

Of the first kind an important subset are where indicative present forms of the verb ‘to be’, which are usually tonally enclitic (Devine & Stephens 1994, 355–356; Barrett 2001[1964], 425–426), are graphematically demarcated both to the left and right (Devine & Stephens 1994, 328; cf. for Cypriot Egetmeyer 2010, 528), *e.g.*:⁶

(426) SEG 14:604 1 ‘Nestor’s Cup inscription’ (8th century BCE, text per Watkins 1976, 40; cf. Faraone 1996, 77; original written right-to-left, accentuation by this author)

→ ΝΕΣΤΟΡΟΣ : Ε[ΣΤ]Ι : ΕΥΠΙΟΤ[ΟΝ] : ΠΟΤΕΡΙΟΝ :

*néstorós*_{<ω>} *e[st]í*_{<ω>} *eúpot[on]*_{<ω>} *potérion*_{<ω>}

PN.GEN be.3SG good_drinking.NOM cup.NOM

‘Nestor’s cup is good to drink from’ (trans. Watkins)

The conditions under which present indicative forms of ‘to be’ are orthotonic are (Devine & Stephens 1994, 355–356):

- At the beginning of a clause;
- At the beginning of a verse;
- Following one of: ΟΥ(Κ) *ou(k)*, ΜΗ *mḗ*, ΑΛΛΑ *allá*, ΚΑΙ *kaí* and ΤΟΥΤΟ *toúto*;
- When the sense is existential.⁷

The context of (426) fulfils none of these conditions.

In the inscriptions that they consider, Devine & Stephens (1994, 328) state that ‘Forms of the verb *to be* are generally punctuated whether they follow a focused word or not’. There are instances where forms of the verb *to be* are not graphematically separated from a preceding graphematic sequence. However, those I have found are paradoxically outside of the indicative, *e.g.*:⁸

⁶ The Nestor’s Cup inscription happens to be the first Greek inscription attested with punctuation (Wachter 2010, 53).

⁷ Barrett (2001[1964], 425–426) rejects this possibility, but Devine & Stephens (1994, 356) provide evidence from musical settings for its validity.

⁸ Similar is ΕΝΟΧΙΟΝΤΟ *enokhoi ’entō* ‘let them be liable’ (IG IV 554, 7; text per Fraenkel 1902).

(427) SEG 23:530 2 (Dreros, 650–600 BCE; original text is written in ‘boustrophedon’, with the line below written left-to-right)

→ | ENIKAΔI | OPONHMEN

*en=ikádi*_{<ω>} *óron=* *êmen*_{<ω>}

in=twenty limit.ACC.SG be.PRS.INF

‘the limit is to be on the twentieth day’ (trans. with ref. to Van Effenterre 1961, 547)

Other enclitics may be demarcated to the left in inscriptions, e.g. TE *te* in the following:

(428) SEG 11:314 5–6 (Argos, 575–550 BCE)

→ ΣΥΛΕΥΣ : TE

KAI : EPATYIIOS

*suleús*_{<ω>} *te*_{<λ>} *kai*_{<ω>} *erátuiios*_{<λ>}

PN.NOM PTCL and PN.NOM

‘Syleus and Eratyios’ (trans. Probert & Dickey 2015, 115)

The example comes in a sequence of personal names, in which all forms of the prepositive KAI *kai* are also demarcated to the right (cf. also the discussion in Morpurgo Davies 1987, 277 n. 22).

Elsewhere in the same inscription TE *te* is graphematically dependent on the preceding lexical:

(429) SEG 11:314 3 (Argos, 575–550 BCE)

→ : KAITAXPEMATATE :

kai= *tà*= *khrématá*= *te*_{<ω>}

and the treasures PTCL

‘and both the treasures ...’

If possession of a (primary) accent is diagnostic of prosodic wordhood, the face-value interpretation of the fact that forms of ‘to be’ and TE *te* are written as independent graphematic words is that graphematic words do not correspond to prosodic words.

13.5.1.4. Graphematic dependence of orthotonic forms

The reverse problem can also be found, namely, that function words, which are accented in our texts, are written in inscriptions as graphematically dependent forms. In the following example the postpositive ΔE *dé* is written together with the preceding lexical exactly as one would expect if it were an enclitic like TE *te*, as at (429):

(430) SEG 11:314 10–11 (Argos, 575–550 BCE)

: ΔΑΜΟΣ

ΙΟΝΔΕ : ΧΡΟΝΣΘΟ :

*damós*_{<λ>} *ion=dè*_{<ω>} *khronsthō*_{<ω>}

state.NOM=but use.SUBJ.NACT.3SG

‘But the state may use them’ (trans. Probert & Dickey 2015, 115)

Multiple lexical items may also be written as a single graphematic word. A case in point is IG I³ 5. This inscription is in general punctuated consistently, with proclitics univerted with the following lexical, and with each graphematic word comprising at most one lexical (cf. §13.5.4). An important exception, however, is where two lexicals are punctuated as a single graphematic word:

(431) IG I³ 5, 3 (Eleusis, ca. 500 BCE; vowel lengths provided by this author)

→ : ΗΕΡΜΕΙΕΝΑΓΟΝΙΟΙ :

hermé̂= *enagōnīōi*_{<ω>}

DN.DAT of_the_games.DAT

‘to Hermes, president of the games’

IG I³ 699 furnishes a further example. In this inscription word-level units are generally punctuated (on the graphematic prefixation of the pronoun ME *me*, see §13.5.1.5). However, in the second half of the second line, given here, a phrase is punctuated:

(432) IG I³ 699 (Athens, ca. 500–480 BCE; vowel lengths provided by this author)

→ : ΗΟΣΜΙΚΥΘΟΗΥΙΟΣ

ho= *smikúthō* *huiós*

the Smikuthos.GEN son.NOM

‘the son of Smikuthos’

13.5.1.5. Labile polarity in appositives

ENCLITICS PUNCTUATED AS PREPOSITIVES

Differentiating enclitics from postpositives more generally is the fact that they may bring about changes in the accentuation of preceding morpheme sequences. In the following example, the enclitic pronoun ME *me* changes the realisation of the pitch accent on the connective ΓΑΡ *gár* from grave to acute:

(433) Eur. *Hipp.* 10, 13 (text per Barrett 2001[1964])

→ ὁ γάρ με Θησέως παῖς, Ἀμαζόνος τόκος

...

λέγει κακίστην δαιμόνων πεφύκεναι

<i>ho</i>	<i>gár</i>	<i>me</i>	<i>thēséōs</i>	<i>paîs</i>
the	for	me	PN.GEN	child.NOM

<i>amazónos</i>	<i>tókos</i>	...	<i>légei</i>	<i>kakístēn</i>
PN.GEN	offspring.NOM.SG		says	worst.ACC.F.SG

<i>daimónōn</i>	<i>pephukénai</i>
gods.GEN	become.PRF

‘For the son of Theseus, the offspring of an Amazon ... says that I am the worst of the gods.’

Punctuation can, however, group such ‘enclitics’ with *following* lexicals (example cited in Goldstein 2016, 67):⁹

(434) IG I³ 775 (Athens, ca. 500–480 BCE; vowel lengths provided by this author)

ΗΙΕΡΟΚΛΕΙΔΕΣ : ΜΑΝΕΘΕΚΕΝ :

<i>hierokleídēs</i> _{<ω>}	<i>m=anéthēken</i> _{<ω>}
PN.NOM	me=dedicated

‘Hierocleides dedicated me’

This example and others like it have been used as evidence of variable clitic polarity in Greek. This is to say that particles traditionally classed as enclitics may behave as proclitics (Goldstein 2016, 67–68; cf. Devine & Stephens 1994, 365–368). The discrepancy between accentuation and graphematic word division in inscriptions is, however, a problem for the alignment of prosodic words with graphematic words.

The issue is that tonal spreading, the process by which enclitics affect the accent of the previous morpheme, ‘typically occurs within a particular prosodic domain’ (Goldstein 2016, 59). In Greek the domain in question is either the prosodic word or the recursive prosodic word (Goldstein 2016, 59). Under the prosodic hierarchy, prosodic words cannot overlap. Accordingly, if tonal spreading occurs within the prosodic word, it must belong to the same prosodic word as the morpheme immediately prior. This, however, is incompatible with ‘enclitics’ being able to appear first in their prosodic word, if the punctuation of (434) is taken to mark prosodic words.

⁹ IG I³ 1008 provides a parallel without elision of *me*.

PREPOSITIVES PUNCTUATED AS POSTPOSITIVES

It is also possible to find ‘prepositive’ morphs that are graphematically dependent on a preceding host, that is, that are written as though they are postpositive. Consider the treatment of the article in the following example:

(435) IG IV 554, 4–5 (Argos; 6th or early 5th century BCE; vowel lengths this author)

ΤΟΝΓΡΑΣΣΜΑΤΟΝ : ΗΝΕΚΑΤΑΣ : ΚΑΤΑ

ΘΕΣΙΟΣ : ΕΤΑΣ : ΑΛΙΑΣΣΙΟΣ

*tôn=grassmátōn*_{<ω>} *héneka=tâs*_{<ω>} *katathésios*_{<ω>} *ē=tâs*
 the.GEN.PL because_of=the.GEN.SG desposition.GEN.SG or=the.GEN.SG

*aliássios*_{<ω>}

assembly.GEN.SG

‘on account of the deposition of written proposals or the act of the assembly’
 (trans. Buck 1998[1955], 284)

In the first determiner phrase the article is graphematically dependent on the following word, ΤΟΝΓΡΑΣΣΜΑΤΟΝ *tôn=grassmátōn*_{<ω>}. By contrast, in the next two determiner phrases, the article is written together with a preceding graphematic host, respectively ΗΝΕΚΑ *héneka* and ΗÈ. Postpositive placement of the article is paralleled not only in word-punctuating inscriptions, e.g. IGA 321 from Oeantheia (Roehl 1882, 69–73), but also in inscriptions that punctuate larger prosodic units (Wachter 1999, 375).

13.5.1.6. Summary

Punctuation in word-separating inscriptions sometimes contradicts what is expected for graphematic words correspond to prosodic words:

- Enclitic forms may be written as independent graphematic words;
- Orthotonic forms may be written together with other words;
- The expected polarity of appositives is not always observed.

13.5.2. Graphematic words need not include lexical words

In addition to the problems associated with the distribution of the pitch accent, there are issues of word class to consider. Under one definition of prosodic wordhood in Greek, a one-to-one mapping exists between lexicals and prosodic words. Thus Golston (1995, 346) states (also adopted in Vis 2013):

a Greek utterance has as many phonological words as it has lexical items and the right edge of each phonological word is coterminous with the right edge of a lexical item.

However, epigraphic punctuation, if it corresponds to prosodic word demarcation, is inconsistent with Golston’s notion of prosodic wordhood. We have already seen

(§13.5.1.4) that sequences comprising multiple lexicals may be univerted. In addition, although graphematic words *usually* include at least one lexical, this is not always so, and non-lexical sequences may on occasion stand as independent graphematic words. Minimally, such sequences can comprise a single non-lexical. In SEG 11:314 we find several instances of the dative plural definite article τοῖσι written as an independent graphematic word (cf. Probert & Dickey 2015, 120):

(436) SEG 11:314 5–6 (Argos, 575–550 BCE)

ΤΟΙΣΙ : ΧΡΕΜΑΣΙ : ΤΟΥΣΙ : ΧΡΕΣΤΕΡ
 ΗΟΙΣΙ :

*toîsi*_{<ω>} *khrēmasi*_{<ω>} *toûsi*_{<ω>} *khrestēr*_{<λ>} *íioisi*_{<ω>}
 the.DAT.PL treasures.DAT the.DAT.PL utensils.DAT

‘the treasures that are utensils’ (trans. Probert & Dickey 2015, 115)

Parallels may be given from IGA 497, which is well known for the high degree of consistency in its graphematic word demarcation (§13.5.4), where the following non-lexicals are written as independent graphematic words:¹⁰

(437) ΟΣΤΙΣ: *óstis* (A.1, B.8)

(438) ΟΙΤΙΝΕΣ: *oítines* (B.29)

(439) :KENON: *kênon* (A.3–4)

(440) :TOKENO: *toû=kênou*_{<ω>} _{<λ>} (A.5, cf. A.12, B.7–8)

(441) :KAIA _{<λ>} ΥΤΟΝ: *kai=a*_{<λ>} *utón*_{<ω>} (A.4–5, cf. A.11–12, B.6–7)

(442) :ΕΝΗΙΣΙΝ : *enêisi*_{<ω>} (B.36)

Sequences of multiple non-lexicals may also be written as a single graphematic word. In IGA 499 the non-lexical sequences HM MEN *ê̄m=mèn*_{<ω>} and HN ΔΕ *ê̄n=dè*_{<ω>} are written as independent graphematic words:

(443) IGA 499, 1–3 (Ephesus)

: ΗΜΜΕΝ : ΑΠΟΚΡΥΨΕ

[ΙΔΕ]ΞΙΟΣ : ΗΝΔΕ : ΕΠΑΡΕΙ : ΤΗ

[ΝΕ]ΥΩΝΥΜΟΝ : ΠΤΕΡΥΓΑ :

*ê̄m=mèn*_{<ω>} *apokrúpse*_{<λ>} [*i*_{<ω>}] *dé]xios*_{<ω>} *ê̄n=dè*_{<ω>}
 if=PTCL hide.FUT.IND.ACT.3SG the_right.NOM if=PTCL

¹⁰ Text per Roehl (1882, 135). On considerations regarding the text to use, see n. 1 above.

<i>epárei</i> _{<ω>}	<i>tè[n=e]juónumon</i> _{<ω>}	<i>ptéruga</i> _{<ω>}
lift.FUT.IND.ACT.3SG	the=left.ACC	wing.ACC
‘if the right hides, but if it lifts the left wing’		

That punctuated units correspond to prosodic words is indicated by the presence of prosodic word-internal sandhi phenomena, specifically, nasal assimilation of /n/ to /m/ in before MEN *mén* in HMMEN *è̃m=mèn* (§13.3.3). Furthermore, note that nasal assimilation is absent across a divider, as in the case of E]YΩNYMON : ΠΤΕΡΥΓΑ *e]juónumon*_{<ω>} *ptéruga*_{<ω>} : if nasal assimilation were to have taken place here, we would expect *E]YΩNYMOM : ΠΤΕΡΥΓΑ *e]juónumom*_{<ω>} *ptéruga*_{<ω>} (cf. the sequence αμ ποταμον *am’ pótamon* (Devine & Stephens 1994, 305)).

If Golston is correct that a given prosodic word can only contain one lexical, graphematic words in these and other Greek inscriptions cannot correspond to prosodic words. It is, therefore, important to establish whether or not lexical words and prosodic words must correspond in the way Golston describes. See further on this issue in Chapter 16.

13.5.3. *Bimoraic non-lexical graphematic words are rare*

A further difficulty concerns what comprises a minimal prosodic word. We saw in the Introduction that there is a cross-linguistic tendency for prosodic words to be binary, either syllabically or moraicly (§1.4.2.4). This is the Prosodic Minimality Hypothesis (PMH). For the purposes of this study the binarity of the foot has been especially important for the identification of the target of graphematic words in Tiberian Hebrew (Part III).

For lexical words, at least, Ancient Greek can be said to adhere to the PMH, provided that the final consonant is excluded from the mora calculation (Steriade 1988; Golston 1991; 2013), e.g. γή *gḗ* ‘land, earth’, ὄϊς *oḯs* ‘sheep’ etc. Where a lexical word consists of only a single mora under this calculation, and a final consonant is present, the final consonant is incorporated into the moraic calculation (Blumenfeld 2011). This condition is able to account for words that would otherwise be monomoraic, namely, δός *dós* [give.IMP], θές *thés* [put.IMP], ἔς *hés* [hurl.IMP] (Blumenfeld 2011; Golston 2013).

Non-lexical words, both in Greek and in other languages, do not necessarily adhere to the PMH. In Greek, therefore, a number of non-lexical words are monomoraic (Devine & Stephens 1994, 304): ME *me*, ΣΕ *se*, ΗΕ *he*, ΣΦΕ *sphe*, ΓΕ *ge*, ΤΕ *te*, ΤΙ *ti*, ΔΕ *dé*, (Η)Ο *ho*, ΤΟ *tó*, ΤΑ *tá*, ΣΥ *sú*, (Η)Α *há* and ΠΡΟ *pró*. Excluding final consonants yields three more monomoraic morphs (Devine & Stephens 1994, 304): EN *en*, AN *án* and ΠΡΟΣ *prós*.

Non-lexicals in Greek (or indeed other languages) are by no means necessarily monomoraic, however. Plenty of both bimoraic and trimoraic non-lexicals can also be found (cf. Devine & Stephens 1994, 357):

- Prepositions: ANA *aná*, ΑΠΟ *apó*, ΔΙΑ *diá*, ΕΞ *ex*, ΕΠΙ *epí*, ΚΑΤΑ *katá*, ΜΕΤΑ *metá*, ΠΑΡΑ *pará*, ΠΕΡΙ *perí*, (Η)ΥΠΕΡ *hupér*, (Η)ΥΠΟ *hupó*, (Η)ΩΣ *hōs*;
- Conjunctions: ΑΛΛΑ *allá*, ΑΤΑΡ *atár*, ΕΙ *ei*, ΕΠΕΙ *epeí*, Η, ΗΔΕ *ēdé*, ΗΜΕΝ *ēmén*, ΚΑΙ *kaí*, ΟΥΔΕ *oudé*, ΜΗΔΕ *mēdé*, ΦΗ *phē*, (Η)ΩΣ *hōs*;
- Negatives: ΟΥ(Κ) *ou(k)*, ΜΗ *mē*;
- Adverb: ΙΔΟΥ *idouí*.

The problem with the epigraphic evidence in these terms is that it is difficult to find examples of bimoraic non-lexicals written as independent graphematic words in Greek inscriptions (cf. Devine & Stephens 1994, 327). Thus appositives consisting of two light syllables are not generally written as independent graphematic words (Devine & Stephens 1994, 327), e.g. ΕΠΙ *epí*:

(444) SEG 11:314 1 (Argos, 575–550 BCE)

→ ΕΠΙΤΟΝΔΕΟΝΕΝ : ΔΑΜΙΙΩΡΓΟΝΤΩΝ :

*epì tōndeōnēn*_(w) *damiiorgōntōn*_(w)

on the_following serve_as_δαμιοργοί

‘When the following were *damiorgoi*’ (trans. after Probert & Dickey 2015, 115)

However, that bimoraic morphemes can, albeit rarely, be written as independent graphematic words is shown by instances of the article written as an independent graphematic word. This can happen when the article specifies a branching constituent, in which case it is more likely to be separated from neighbouring morphs (Devine & Stephens 1994, 327), e.g.:

(445) IG IV 554, 1 (Argos; 6th or early 5th cent. BCE; vowel lengths this author)

→ [Θ]ΕΣΑΘΡΟΝ[: ΤΟ]Ν : ΤΑ : Σ : : ΑΘΑΝΑΙΑΣ

*thēsaurōn*_(w) *tōn*_(w) *tâ*_(w) *s*_(w) *Athanaías*_(w)

treasuries.GEN the.GEN.PL the.GEN.SG DN.GEN

‘of the treasuries of Athena’

(446) IG IV 554, 2 (Argos; 6th or early 5th century BCE; vowel lengths this author)

→ [ΕΤ]ΑΝΒΟΛΑΝ : Τ[Α]Ν : ΑΝΦΑΡΙΣΣΤΟΝΑ

*ē=tân=bōlân*_(w) *tân*_(w) *anph’=arísstona*

or=the.ACC.SG=council.ACC the.ACC.SG around=PN.ACC

‘or the council around Arisston’

Perhaps not too much should be read into the word divisions in the first line of (445). There is clearly something a bit odd going on here, in view of the irregular word divider in ΤΑ:Σ *ta*_(w) *s* in the first line, and the two word dividers after it. This

is likely related to issues with a nail hole (Jameson 1974; Devine & Stephens 1994, 389). Elsewhere in the inscription, however, word division is more regular. It may, therefore, be significant that in (446) a word divider appears after the article T[A]N *tān*. Such cases are, in any case, rare. It is far more common for a bimoraic non-lexical to be unverbated with its prosodic host.

13.5.4. Inscriptions vary in the level of consistency of word division practice

Many of the issues described in the foregoing sections amount, in both prosodic and morphosyntactic terms, to (at least apparent) inconsistency. Thus we have seen that graphematic words may comprise either no lexical words at all (§13.5.2) or multiple lexical words amounting to whole phrases (§13.5.1.4). This is an issue we have, however, encountered in Northwest Semitic inscriptions. In particular, in both written Ugaritic and written Phoenician nouns with dependent nouns and noun phrases are found written as a single unit (§3.4.2, §6.3), as are the combinations Verb + NP and Verb + PP (§3.4.3, §6.4).

In the Northwest Semitic material, however, there is an important distinction to be made between the orthography of the consonantal Masoretic Text and the Meshah^c stelae, on the one hand, and Phoenician and Ugaritic ('Majority' orthography) material, on the other. I have argued that the important distinction here is between the representation of actual prosodic words in a particular prosodic context in Phoenician (Chapter 3) and Ugaritic (Chapter 8), versus minimal prosodic words, *i.e.* units that must stand as prosodic words in any context (Chapter 12). The difference between the two orthography types comes down to morphosyntactic consistency. A natural consequence, for example, of the minimal prosodic word orthography is that a graphematic word will comprise a maximum of one lexical, since lexical words in Moabite and Hebrew are valid minimal prosodic words.

A similar variability in levels of consistency can be seen in the Greek material. For example, the high level of consistency in the word division orthography of IGA 497 (Roehl 1882, 135–136) has been noted since at least Kaiser (1887, 17) (cf. also Larfeld 1914, 303; Meiggs & Lewis 1969, 62; Morpurgo Davies 1987, 271, 277 n. 21). IGA 497, and other inscriptions like it,¹¹ are characterised by punctuating such that a single graphematic word comprises a maximum of one lexical. Some of these, as we have seen, show that a single graphematic word need not comprise any lexicals at all (§13.5.2). From a typological perspective, therefore, the word division orthography of IGA 497 is the same as that which we find in the consonantal text of the Masoretic Text and in the Meshah^c stele. The implications of this typological similarity are taken up at §17.4 below.

¹¹ Cf. SEG 11:314, IGA 499, IG I³ 5. Larfeld (1914, 303) notes the following as similarly consistent in word division practice: IGA 5, 42, 43a, 359, 498b, 502, 544. IGA 359 and 498b have no non-lexicals.

13.6. Conclusion

Prosodic words in Greek have previously been identified based on their status as a lexical or non-lexical, and their pitch accentual status. The difficulties with identifying graphematic words with prosodic words in Greek word-punctuating inscriptions therefore fall into two basic categories:

- Pitch accentual status (enclitic, orthotonic) is not a good predictor of graphematic clisis: enclitics can be written as independent graphematic words, or as prepositives;
- The presence of a lexical is not a good predictor of the status of graphematic wordhood: a graphematic word may contain multiple lexicals, or no lexicals at all.

These two issues are addressed in the following chapters. First, Chapter 14 outlines the pitch accentual status of pre- and postpositives. Pitch accent alone is found not to be sufficient to explain the observed behaviour. For this reason, Chapter 15 brings rhythm into consideration, and argues that graphematic words correspond to a prosodic unit determined by rhythmic rather than pitch accentual properties. Finally, Chapter 16 sets out to account for the punctuation of lexicals. Once again, it is argued that bringing rhythm into consideration is able to provide at least part of the solution.

Chapter 14

The pitch accent and prosodic words

14.1. Introduction

The difficulties outlined at §13.5.1 above can be seen as part of a wider set of issues with the mapping of Greek pitch accentual domains with those of prosodic wordhood (cf. Golston 1990; Revithiadou 2013). These centre on the following fundamental concerns:

- **Syntactic distributions of appositives and clitics** Although appositives and clitics are treated differently in terms of pitch accentuation, their syntactic distributions are hard to distinguish (Golston 1990, 69; 2016, 50).
- **Asymmetry between enclitics and proclitics** Enclitics, *i.e.* morphemes dependent for their pitch accentuation on a preceding morpheme, comprise a subset of the larger class of postpositives. However, the group of prepositives are likely not to have had a subset of proclitics to parallel the enclitics (Devine & Stephens 1994, 358–361).
- **Pitch accentuation of monomoraic morphs** Per §13.5.3, a minimal prosodic word is bimoraic. Insofar as the possession of a primary accent is a cross-linguistic diagnostic of prosodic wordhood, monomoraic morphs should not in principle carry an accent at all. However, a number of monomoraic morphs in Greek do precisely this, notably, ΔΕ *dé*, ΤΟ *tó*, ΤΑ *tá*, ΣΥ *sú*, (Η)Α *há*, ΠΡΟ *pró*, ΑΝ *án*, ΠΡΟΣ *prós*, ΓΑΡ *gár* (cf. Devine & Stephens 1994, 304).

These three facts suggest that Greek pitch accentuation is somewhat independent of prosodic wordhood as construed more broadly. How, then, should clitics and appositives be analysed in terms of pitch accentuation? This is the question I set out to address in this chapter, before bringing rhythm into consideration in the next chapter.

14.2. Prosody of postpositives and enclitics

Goldstein (2016) and Goldstein & Haug (2016) provide some of the most thorough treatments of the prosody of postpositives and enclitics in Greek, and therefore offer important foundations for the work here. These are first reviewed in turn.

14.2.1. Goldstein (2016)

Goldstein (2016, 49–51) proposes that, although there is no distributional distinction between clitics and postpositives (also known as syntactic clitics), they incorporate differently with their prosodic hosts. In the case of enclitics, the host and clitic together project a recursive prosodic word (Goldstein 2016, 51). By contrast, a postpositive incorporates with its host at the level of the prosodic phrase (Goldstein 2016, 51). In support of this Goldstein points out that enclitics are distinguished from postpositives in the following respects:

- '[P]ostpositives never trigger a secondary accent on a host' (Goldstein 2016, 56), whereas enclitics do (Goldstein 2016, 53–56);
- The accent on a postpositive undergoes 'lulling' when followed by a prosodic word, whereas this happens with enclitics only in very specific circumstances (Goldstein 2016, 57–58).
- Clitic chains demonstrate 'tonal spreading', whereas chains of postpositives do not (Goldstein 2016, 59–60).

However, Goldstein acknowledges that postpositives and enclitics are aligned, over against lexicals, on the following points:

- Both enclitics and postpositives are capable of standing at Porson's Bridge (Goldstein 2016, 52–53).
- For non-clitic non-postpositive words, it is normally the case that 'When the vowel that hosts a high tone is deleted [as in elision], the tone docks on the next vowel to the left' (Goldstein 2016, 56). However, in the case of both enclitics and postpositives, this does not happen. Instead, the high tone disappears altogether (Goldstein 2016, 56–57). High tones also evanesce in the case of the prepositive prepositions (Goldstein 2016, 57).

Goldstein (2016, 60) concludes that, 'postpositives do exhibit behavior characteristic of prosodic words, especially when it comes to lulling and tonal spreading'. Postpositives nevertheless demonstrate prosodic dependency, in common with clitics, in their capacity to stand at Porson's Bridge. Furthermore, 'Whatever [postpositives'] orthographic accent means prosodically, its behaviour differs from that of true lexical accents' (Goldstein 2016, 60), that is in respect of the fact that their high tones evanesce in cases of elision.

For Goldstein (2016, 51), the fact that many postpositives do not meet the threshold of prosodic word minimality is not a problem because 'the minimal word requirement in Greek is category specific ... and in particular restricted to nouns' (Goldstein 2016, 51). Goldstein supports this with the observation that, 'Certain monosyllabic verb forms, such as the imperatives *δός* *dós* "give!" and *θέξ* *thés* "put!" also fail to meet the minimality threshold'.

However, Goldstein does not address the question of why it should be that minimal prosodic word requirements would be category specific. By contrast, for

Devine & Stephens (1994), non-lexicals can fall below the threshold of prosodic word minimality precisely because they are prosodically dependent upon a host word. If this is correct, it cannot also be the case that monomoraic non-lexicals project a prosodic word of their own.¹

14.2.2. Goldstein & Haug (2016)

Goldstein & Haug (2016, 301–303) offer an alternative hypothesis, that both enclitics and postpositives incorporate with their hosts by projecting a recursive prosodic word. Where they differ is in the calculus for the position of the secondary accent triggered by clitic incorporation:

- **Enclitics** trigger a secondary accent ‘on the host, the clitic, or none at all’ (Goldstein 2016, 302);
- **Postpositives** trigger a secondary accent which always falls on the postpositive.

An account along these lines has some advantages over that of Goldstein (2016):

- It provides a context for understanding how the accent on postpositives might differ from that on lexicals (Devine & Stephens 1994; Goldstein 2016). Specifically, if the accent on a postpositive is a secondary accent, in common with that on enclitics, it is easier to see why it might evanesce in the same way as the accent on enclitics.
- It avoids the need for supposing that sub-prosodic words can serve as prosodic hosts, as well as the oddity that sub-prosodic words carry an accent at all;
- It explains why accented postpositives might nevertheless be written as graphematically dependent forms (cf. §13.5.1.4).

Nevertheless, the account does not explain why an appositive, especially a postpositive, might associate rightward rather than leftward (§13.5.1.5). However, this is an important step to understand in order to understand labile polarity in general. In order to start to answer this question, it is therefore important first to outline the basis of the traditional distinction between prepositives (§14.3.1). From this I move on to address the accentual basis for labile polarity in prepositives (§14.3.2).

14.3. Prosody of prepositives and ‘proclitics’

14.3.1. Traditional identification of prepositives and proclitics

The discussion in the preceding section has considered only those non-lexicals traditionally classified as either enclitic or postpositive: proclitics (or prepositives) are not considered. It was observed above (§14.1), however, that there is an asymmetry in the Greek system of clitics and appositives. Specifically, there is not likely to have

¹ It is also the case that an account has been offered as to why a handful of verbal forms might appear to fall below the threshold of minimality without compromising the PMH (Blumenfeld 2011; Golston 2013).

been a class of accentless proclitics parallel to the class of enclitics: proclitics were likely to have been part of the rising pitch trajectory, and are therefore not to be distinguished from prepositives (Devine & Stephens 1994).

Prepositives can be divided into two groups. The items of the first group are not marked with an orthographic accent in the manuscripts (cf. Devine & Stephens 1994, 357):

- Nominative masculine/feminine forms of the definite article, *e.g.* ὁ *ho*;
- Monosyllabic vowel-initial prepositions, *e.g.* ἐκ *ek* or ἐξ *ex*;
- Negative particle, *e.g.* οὐ(κ) *ou(k)*;
- Monosyllabic vowel-initial conjunctions, *e.g.* εἰ *ei*, ὡς *hōs*.

These items have orthographically orthotonic counterparts, namely:

- Nominative masculine/feminine forms of the relative pronoun, *e.g.* ὃ *hó*;
- Number 'six', *e.g.* ἕξ *héx*;
- 2SG form of the verb 'to be', *e.g.* εἶ *eî*;
- Number 'one', *e.g.* εἷς *heîs*, ἕν *hén*;
- Relative 'where': οὗ *hoû*;
- Adverb 'so, thus': ὥς *hōs*.

Given that each of the non-accented morphs is paired with an accented morph, it is generally assumed that the practice of variably accentuating these minimal pairs was for the purpose of orthographic disambiguation and does not reflect any underlying prosodic difference (Devine & Stephens 1994, 357).

In addition to these, there exist a number of other items that are taken to be 'proclitics', but which are always accented in our texts. The set of proclitics includes all of the above-mentioned forms, but also includes a number of other polysyllabic morphemes, namely (cf. Devine & Stephens 1994, 357):

- Prepositions, *e.g.* ἀνά *aná*, ἀπό *apó*;
- Conjunctions, *e.g.* ἀλλά *allá*, ἐπεὶ *epeî*;
- Negatives, *e.g.* μή *mē*;
- Adverb, ἰδοῦ *idouí*.

On the basis of Greek musical settings, Devine & Stephens (1991, 284–286) argue that prepositives/proclitics are accented in such a way as to be part of the rising trajectory towards the primary (usually lexical) accent of the prosodic word (see also Devine & Stephens 1994). This is true both of orthotone prepositives and (orthographically at least) atonic forms: in both cases the prepositive forms a pitch accentual unit with the following lexical.²

² Probert (2006, 69 n. 35) takes the view that such orthographic accentuation is a matter of convention (see also Probert 2003, 133–142).

Given that all prepositives are likely to have been accented, it is worth asking where prepositives fit into the picture we have so far drawn. Extending Goldstein & Haug (2016) I propose that prepositives, together with any following lexicals, project a recursive prosodic word, whereby the prosodic word's secondary accent falls on the prepositive.

14.3.2. Accentual basis for labile polarity in prepositives

Just as postpositives may on occasion associate rightward, prepositives may also on occasion associate leftwards, especially in the case of prepositions. Such cases of hyperbaton have been argued to entail the phonological leftward movement of the dependent noun phrase (Agbayani & Golston 2010, 138). The traditional term for such cases is 'anastrophe' (Devine & Stephens 1994, 364–365; Goodwin 1894, §116). In such instances the position of the orthographic pitch accent changes (Goodwin 1894, §116; Devine & Stephens 1994, 364–365), *e.g.*:

(447) Eur. *Hipp.* 8 (text per Barrett 2001[1964])

τιμώμενοι χαίρουσιν ἀνθρώπων ὑπο

timómēnoi

khaírousin

anthrópōn

húpo

honour.PRS.PTCP.NACT.NOM.PL

rejoice.PRS.IND.ACT.3PL

people.GEN

by

'They rejoice when they are honoured by people'

Here the preposition ὑπο *húpo* 'by' governs the noun immediately to its left, viz. ἀνθρώπων *anthrópōn* 'people', rather than governing a noun to its right, as would be expected. In such circumstances, the preposition is accented on the first syllable rather than the second.

Following and extending Goldstein & Haug (2016, 301–302), I take it that ὑπό in this case is tonally dependent on its (preceding) host, namely ἀνθρώπων *anthrópōn*, and that the two project a recursive prosodic word, with the primary accent on ἀνθρώπων *anthrópōn*, and the secondary accent on ὑπο *húpo*.

14.4. Conclusion

The goal of the chapter was to account for the prosodic behaviour of clitics and appositives. In reviewing the literature we saw that appositives generally pattern with lexicals over against clitics in the position of the orthographic pitch accent. This was accounted for by following Goldstein & Haug (2016) in proposing that appositives, together with their host, project a recursive prosodic word, whose secondary pitch accent falls on the appositive. The exception is where an oxytonic vowel is elided. In this case there is no secondary accent. Enclitics are a subset of postpositives that follow different rules for the position of the pitch accent, so that the secondary accent may appear on the host, on the enclitic or nowhere.

Goldstein (2016) and Goldstein & Haug (2016) deal only with postpositives and enclitics. Since my ultimate goal is to account for the labile proclisis of *e.g.* ME *me* in inscriptions, it is also necessary to have an understanding of proclitics. The pitch accentual behaviour of proclitics was therefore described in §14.3. Unlike postpositives, among prepositives there are no proclitics parallel to enclitics, whose pitch accent is projected forward on to a following morpheme, at least in the inherited accent tradition. I therefore proposed that all prepositives project a secondary accent together with their host, accounting for their orthotonic status. Finally, in §14.3.2 we saw that there is some accentual basis for labile polarity in prepositives, since ‘prepositions’ may occur in postpositive position, and when they do, the accentuation changes.

These findings, however, still do not allow us to account for the graphematic proclisis of enclitics in inscriptions, or indeed, the graphematic enclisis of postpositives such as the article (§13.5.1.5). In order to do this, it is necessary to take rhythm into account. It is to this goal that I turn in the next chapter.

Chapter 15

Domains of pitch accent and rhythm

15.1. Introduction

At §13.5.1.5 we gave the following example of the graphematic proclisis of an enclitic:

(448) IG I³ 775 (Athens, ca. 500–480 BCE; vowel lengths this author)

HIEPOKΛEΙΔΕΣ : ΜΑΝΕΘΕΚΕΝ :
*hierokleidēs*_{<ω>} *m=anēthēken*_{<ω>}
PN.NOM me=dedicated
‘Hierocleides dedicated me’

In starting to provide an account for this behaviour, the previous chapter gave an overview of the pitch accentual status of pre- and postpositives, as well as the limited evidence in pitch accentuation for labile polarity. However, the behaviour of the pitch accent is not enough to explain examples like (448).

The present chapter seeks to provide such an account by bringing rhythm into consideration. Indeed, evidence for the kind of proclisis that we see in (448) can be found in metrical compositions, such as the following:

(449) Aesch. *Suppliants* 785 (text Page 1972; ex. quoted at Devine & Stephens 1994, 368; Goldstein 2016, 64)

→ κελαινόχρως | δὲ πάλλεταιί | μου καρδία·
kelainóchrōs | *dè* *pálletaí* |
black_coloured PTCL quiver.PRS.3SG

mou *kardía*
me.GEN heart.NOM.F.SG
‘My black heart is aquiver’

Porson’s Law states that a (prosodic) word break is not permitted after a heavy syllable in the first syllable of the third metron (the ninth element of the line;

cf. Devine & Stephens 1994, 105; Goldstein 2016, 52). In (449) the metra boundaries are indicated by |. The first syllable of the third metron is μου *mou*, and is therefore heavy. Porson's Law states that a prosodic word break cannot occur in this position. For this to be the case, μου *mou* must associate rightward with καρδία *kardía*, rather than leftward with πάλλεται *pálletai*. So far, so good. The problem is that in terms of pitch accentuation, the association of μου *mou* is leftward, in view of the secondary accent on πάλλεταιί *pálletaí*. This is exactly the phenomenon that we see in (448), and that this chapter sets out to explain.

We can also note that the same kind of proclisis can occur at the caesura (for the phonological status of the caesura, cf. Goldstein 2010, 108–109). This is to say that an enclitic pronoun can occur immediately to the right of the major break in the line (caesura indicated by |):

(450) Aesch. *Choephoroi* 181 (text Page 1972; ex. quoted at Devine & Stephens 1994, 365)

→ οὐχ ἧσσον εὐδάκρυτά | μοι λέγεις τάδε

<i>oukh</i>	<i>hêsson</i>	<i>eudákrutá</i>		<i>moi</i>
NEG	less	deserving_of_tears.N.PL.ACC		me.DAT

<i>légeis</i>	<i>táde</i>
say.PRS.2SG	this.N.PL.ACC

'These things you say to me are not less deserving of tears' (trans. with ref. to Sommerstein 2009)

In both cases, therefore, polarity determined on the basis of rhythm is different from that determined on the basis of the pitch accent.

This is not, of course, to say that accentual and rhythmic polarity always disagree. Indeed, they may align with one another against the syntax (Devine & Stephens 1994, 288) as in the following example:

(451) Eur. *Helen* 471 (text Diggle 1994; partially quoted at Devine & Stephens 1994, 288)

→ πῶς φήις; τίν' εἶπας μῦθον; αὐθίς μοι φράσον.

<i>pôs</i>	<i>phêis</i>	<i>tín'</i>	<i>eîpas</i>	<i>mûthon</i>
how	say.PRS.2SG	what.ACC.SG	say.PST.2SG	word.ACC

<i>authís</i>	<i>moi</i>	<i>phráson</i>
again	me.DAT	tell.IMP.SG

'What you say? What speech? Tell me again!'

Although the enclitic μοι *moi* coheres syntactically with φράσον *phráson*, it coheres prosodically with preceding αὐθίς *authís*. This is the case both in terms of the pitch

accent, as indicated by the acute accent on the final syllable of αὐθίς *authís*, and rhythmically, since under Porson's Law a prosodic word cannot end after a heavy syllable in the line's ninth element. Such, however, would be required if a prosodic word boundary occurred immediately to the right of αὐθίς *authís* (Devine & Stephens 1994, 288).

The relevance of this issue for epigraphic word division is that enclitics are liable to be written either independently or together with a host situated to the right, and that prepositives may on occasion be written together with a host situated to the left (§13.5.1.5). The labile polarity observed in inscriptions is therefore parallel to the labile polarity seen in verse.

From cases where postpositives cohere to the right Devine & Stephens (1994, 367) draw the following conclusion:

The strong association of postcaesural enclitic pronouns with rightward syntactic cohesion indicates that they were both rhythmically prepositive and accentually proclitic ... these pronouns must be either orthotonic or proclitic.

It must be the case that these examples involve rhythmic *proclisis* of normally *enclitic* personal pronouns. However, Devine and Stephens go beyond this and assert that the pronouns must also be accentually proclitic. Devine & Stephens (1994, 367) justify their assertion with the claim that, 'rhythmical demarcation to the left is not compatible with accentual enclisis'. My purpose in this chapter is to challenge this assertion, and propose that if rhythmic demarcation and accentual enclisis were not seen as incompatible, many of the problems associated with labile polarity both in metrical compositions and in inscriptions disappear.

15.2. Challenging the inherited tradition of accentuation

Devine and Stephens do not make explicit *ad loc.* the basis on which they claim that rhythmic demarcation to the left is incompatible with accentual enclisis. However, at p. 152 they state that 'The Greek word is a prosodic domain not only for rhythmic organization but also for tonal organization'. This can be taken to mean that the domain of both pitch accentuation and rhythmic prominence is the (recursive) prosodic word-level unit, also known as the clitic group or the appositive group. If the domain of both rhythmic prominence and pitch accentuation is the same, it follows that, where they conflict, either the domain implied by the orthographic accent or the domain implied by rhythmic organisation corresponds to that unit, but not both. Since the rhythm is 'baked in' to the text, so to speak, it is inherently more likely that it is the position of the orthographic accent that is incorrect.

As stated, the suggestion that the tradition of accentuation in Ancient Greek is faulty is only implicit in Devine & Stephens (1994). There is, however, some evidence that the tradition of accentuation as we have received it may not reflect the prosody of at least the Classical variety of the language. We saw at §13.5.1.2 that the inherited system of orthographic accentuation originated in Hellenistic Alexandria, and,

therefore, most immediately reflects the accentuation of learned Koine Greek there. Whilst there are good reasons to trust the reliability of the system in general terms (§13.5.1.2), aspects of the system, particularly in respect of clitics, have been questioned (Wachter 1999, 366). The objections have to do, in particular, with the phenomenon of ‘tonal spreading’, whereby a sequence of enclitics are accented on all but the last member (Goldstein 2016, 59–60). Barrett (2001[1964], 426–427) questions whether such a rule is likely to reflect any phonological reality (cf. also Probert 2003, §297, with references). An important objection is that, where such an accentuation scheme entails that consecutive syllables are marked with an acute, this is at variance with the rule that prohibits adjacent syllables from being accented Barrett (2001[1964], 427).¹ In its place, Barrett (2001[1964], 427) proposes that ‘each enclitic affects the accentuation of the preceding syllables in the same way in which it would affect it if they were comprised in a single orthotone’, remarking that the medieval MSS of Euripides *Hippolytus* follow this system.

If it is possible for the grammarians to err in the accentuation of enclitic chains, it is not out of the question that they might be incorrect in other respects. Note that such a view would not necessarily entail that the accentuation tradition was incorrect at the point in time when it was codified. Rather, it would simply need to entail that by the point in time when the tradition of Greek accentuation was codified, the memory of the tonal proclisis of the class of ‘enclitics’ had been lost. If the tonal proclisis of enclitics was lost from the language earlier than tonal enclisis of proclitics, this is at least plausible.

15.3. Pitch accentuation and rhythmic prominence have different domains

15.3.1. Introduction

An alternative to querying the validity of the grammatical tradition is to challenge the premise that the domains of pitch accentuation and rhythmic prominence must be the same. That the two might be different is suggested by typological evidence, largely from Japanese, that in languages with a pure pitch accent, the assignment of that accent is in principle independent of rhythmic prominence (see §15.3.2 below). If the domains of pitch accentuation and rhythmic prominence can be different, it will be possible for a given morpheme to associate rightwards in terms of rhythm, but leftwards in terms of the pitch accent.

15.3.2. Accent from a cross-linguistic perspective

Prosody in the world’s languages involves one or more of three prosodic features:

¹ Although Barrett puts this in terms of syllables, in fact the issue is that adjacent morae may not be accented (Steriade 1988, 290). Nevertheless, by either rule we would still not expect to find instances of e.g. Eur. *Hipp.* 876: εἶ τί μοι (Diggle 1984), cf. εἶ τι μοι (Barrett 2001[1964]).

- Pitch, *i.e.* frequency of a given sound (cf. Devine & Stephens 1994, 198–199);
- Duration, *i.e.* temporal extent of a given sound;
- Intensity, *i.e.* the amplitude of a given sound (cf. Devine & Stephens 1994, 195–198).

Languages avail themselves of these features to different extents and to fulfil different functions. For our purposes it is helpful to distinguish the following types (for a full survey, see Devine & Stephens 1994, 198–215):

- Stress languages;
- Pitch differentiated stress languages;
- Pitch accent languages.

In STRESS LANGUAGES, like English, a stressed vowel is characterised by longer duration and greater intensity than its unstressed counterpart; such vowels may also have higher pitch, although they need not do so, and the pitch may in fact be lower (Devine & Stephens 1994, 204). By contrast, in a PITCH DIFFERENTIATED STRESS LANGUAGE, like Swedish, accented vowels are characterised by all three of higher pitch, longer duration and greater intensity (Devine & Stephens 1994, 207).

PITCH ACCENT LANGUAGES differ from both their stress and pitch differentiated stress counterparts in that a high tone is independent of any greater intensity or duration (Devine & Stephens 1994, 211). Such languages are rare, but one such is Japanese. Although it has been found that accented vowels are a little longer than unaccented ones, this difference cannot be perceived (Devine & Stephens 1994, 212). In general, ‘The role of duration and intensity as exponents of the Japanese accent is marginal compared with their role as exponents of the English stress accent’ (Devine & Stephens 1994, 212).

The Japanese case is significant for Ancient Greek prosody because speech rhythm there does not take account of the pitch accent: speech has an iambic rhythm, and the alternation between longer and shorter syllables occurs regardless of the position of the pitch accent (Devine & Stephens 1994, 213). Furthermore, Japanese verse is arranged according to rhythmic principles only, with the position of the pitch accent ignored (Devine & Stephens 1994, 214).

It is generally accepted that the orthographic accents in modern and medieval Ancient Greek texts – *viz.* acute, grave and circumflex – indicate the position and nature of a tonal, *i.e.* pitch, accent (Goldstein 2013). However, it has also been argued, notably in Allen (1973) and Allen (1966), that Ancient Greek also had a stress accent. This proposal has received a mixed reception, and further analyses have yielded mixed results (for a survey, see Golston 2013).

A central issue is, of course, what actually comprises stress. As outlined immediately above, stress is a composite feature comprising greater duration and intensity. Ancient Greek verse, at least, is governed by metrical principles where the key distinction is between heavy and light – *i.e.* moraically long and moraically short – syllables. It is well known, furthermore, that the mapping of linguistic sequences

on to Greek metrical forms is not determined by the position of the pitch accent, but rather by the moraic lengths of the syllables of those sequences. Duration is one of the components of the ‘accent’ in stress and pitch differentiated stress languages: stressed vowels in such languages are longer than unstressed ones. In terms of duration, therefore, moraicly longer syllables in Greek verse can be said to be more prominent than moraicly shorter ones. If Greek verse, especially the iambic trimeter, can be said to have a strong correlation with Greek speech, it follows that Greek speech too can be said to have had an alternation between rhythmically prominent syllables and less prominent ones.

In making the assertion that some syllables in Ancient Greek were more rhythmically prominent than others, I do not mean to assert that Ancient Greek had ‘stress’, or that the rhythmic prominence corresponds exactly to stress in stress and pitch differentiated stress languages. On the other hand, to the extent that ‘non-accentual durational prominence’ shares certain phonological properties with ‘stress’, there are typological grounds for discussing it in the same context as ‘stress’ (for these points and further discussion see Devine & Stephens 1994, 206, 214–215). To propose, therefore, that the domain of pitch accentuation could be independent of the domain of rhythmic prominence is at least possible from a cross-linguistic perspective.

15.3.3. *The E-domain*

That the accentuation of enclitics is governed by a specific domain in Ancient Greek has been suggested before. Steriade (1988) proposes two sub-phrase levels to account for the position of the accent in host-enclitic sequences. As Golston (1990, 71) notes, the domains of enclitic accentuation and of [function word + content word] or [content word + function word] cannot be the same because phenomena such as crasis apply beyond the E-domain, at the level of the appositive group. Golston (1990) argues that, since it is possible to derive tonal accent position on the basis of metrical principles, it is not necessary to posit a separate domain to cover pitch accentuation vis-à-vis incorporation of the clitic group. However, Golston does not address the issues of labile appositive polarity described above (§13.5.1.5). If the domains of the pitch accent and of rhythmic prominence are to be distinguished, something along the lines of the E-domain is therefore needed to account for the distribution of the pitch accent.

The typological evidence adduced in the previous section suggests that rhythmic and pitch accentual word-level domains can be different. However, this does not prove that they *are* different in Classical Greek. For this positive evidence from within Greek is needed. It is this issue to which I now turn.

15.3.4. *Accounting for non-initial position of postpositives*

At §13.5.1.5 and §14.3.2 above I gave evidence that appositives may on occasion reverse their ‘normal’ polarity. This comes largely from metrical compositions, where

both enclitic pronouns, *e.g.* ME *me*, MOY *mou*, and orthotone postpositives, *e.g.* MEN *mén*, ΔE *dé*, ΓAP *gár*, may associate rightwards. Despite the possibility of rightward association, however, it remains the case that postpositives may not come first in their clause (Goldstein 2010, 113; Luraghi 2013). This is not what we would expect if postpositives can indeed associate rightward, as the metrical evidence suggests: instead, we would expect to find at least a small number of clauses where they come first, but this is not found.

The constraint against initial placement of postpositives could in principle be either syntactic or phonological. A phonological explanation is most natural in the case of enclitics, given their projection of an accent on to a preceding syllable. However, in the case of orthotone postpositives, a syntactic explanation could be offered, namely, that, despite bearing an accent, such postpositives are obliged by syntactic rules not to occur first in their clause. For this reason orthotone postpositives have sometimes been referred to as ‘syntactic clitics’ (see Goldstein 2016, 50, with references).

That the constraint is in fact phonological is suggested not least by the fact that orthotone postpositives are prohibited from occurring not only first in their clause, but also first in their verse line (cf. the statement that ‘[clausal] clitics can be hosted outside of second position by the first prosodic word of the metrical line or the first metrical word after the caesura’ (Goldstein 2010, 99)). Since verse lines need not – and indeed regularly do not – coincide with the start of a clause or sentence, this latter constraint cannot be a syntactic one. On the other hand, if verse lines are in principle taken to be intonational phrases (Goldstein 2010, 84–85, 97–99, 102–103; Goldstein & Haug 2016; cf. Devine & Stephens 1994, 398, 414–425; for a caveat see Devine & Stephens 1994, 400), the constraint can be analysed as phonological. Similarly Agbayani & Golston (2010, 6) define a postpositive to be ‘a word that cannot occur at the beginning of a phonological phrase’.

However, if the constraint against initial position in postpositives is phonological, it cannot be rhythmic: as we have seen, it is precisely on rhythmic grounds that labile polarity in appositives is posited. The only remaining possibility is that the constraint is accentual. This is to say that the polarity of an appositive could be labile in terms of rhythm, but fixed in terms of the assignment of the pitch accent. It is this possibility that I now explore.

15.3.4.1. Enclitics

Sauzet (1989) (cf. Golston 1990) proposes that (recessive) orthotone words are accented by first identifying the metrically prominent syllable and assigning Low tone to that syllable. The mora immediately prior to the metrically prominent one is then assigned High tone, *i.e.* the syllable receiving the accent. The metrically prominent syllable is identified by first splitting the word into feet consisting of syllabic trochees, right-to-left, and then ascribing prominence to the first syllable

in the right-most foot. For example, the noun ἄνθρωπος *ánthrōpos* can be split into syllabic trochees in the following way (Sauzet 1989: 98):

(452) ἄνθρωπος
 {(anth.)(rō.pos.)}

The first syllable in the right-most foot is /rō/, which is then the metrically prominent syllable. High tone is then ascribed to the (vocalic) mora immediately prior to that syllable, which in this case is /a/, *i.e.*:

(453) ἄνθρωπος
 H L*
 {(ánth.) (rō. pos.)}

Sauzet (1989, 98) (cf. Golston 1990, 72) holds that enclitics are also associated with the melodic sequence High-Low; enclitics form their own feet (Golston 1990, 72). Where they differ from orthotonic words is in that they may ascribe (part of) this melody to prosodic words situated to their left, *e.g.*:

(454) ἄνθρωπός τις
 H L* H L*
 {(ánth.) (rō. pós.)} (tis.)

If enclitics are lexically associated with the tone sequence High-Low, where in principle the High tone is mapped to the preceding prosodic word, it follows that enclitics cannot be placed first in their clause.²

In Sauzet's analysis (Sauzet 1989, 103–105; followed by Golston 1990, 76), it is also possible for High tone to be assigned lexically, as in the case of oxytones such as καθαρός *katharós*:³

² There are a number of details here that need elaboration, but which are beyond the scope of the present analysis. Sauzet's analysis gives the wrong predictions in cases such as φοῖνιξ τινός *phoînix tinós*: Sauzet would give *φοῖνίξ τινος *phoînix tinos* (per Golston 1990, 74, 81). Golston (1990, 75–77) therefore reformulates Sauzet in a number of important respects, substituting syllabic for moraic trochees, and proposing that enclitics are associated with a High tone only, rather than a High-Low sequence.

³ Footing is not indicated since this is not relevant for accent assignment in these cases.

(455) καθαρός

		H	
{ka.	tha.	rós.}	

In these cases the lexically assigned position of the High tone trumps all other factors that might govern its position. Thus, the High tone associated with enclitics can simply dock to this existing High tone:⁴

(456) καθαρός τις

		H			L*
{ka.	tha.	rós.}	+	{X	(tis.)}
>					
{ka.	tha.	rós.		(tis.)}	

Following Goldstein & Haug (2016, 302), it would be possible to say that the High tone of enclitics is assigned by means of clitic incorporation into a recursive prosodic word.

15.3.4.2. Orthotone postpositives

Goldstein & Haug (2016, 302–303) propose that orthotone postpositives (*e.g.* MEN *mén*, ΔΕ *dé*, ΓΑΡ *gár*) are also assigned the accent by means of secondary prosodic word incorporation. Where postpositive orthotones differ from enclitics is in the fact that the position of the secondary accent is fixed, *i.e.* it always occurs on the postpositive in a fixed position. In Sauzet's terms, we can say that this (secondary) accent is a lexically assigned High tone.

At §15.3.4.1 above, the obligatory non-initial position of enclitics was accounted for by the fact that they map a High tone on to a preceding syllable. We could therefore seek to account for the obligatory non-initial position of orthotone postpositives by proposing that they map the tone sequence NH-H (NH = Non-High, H = High) on to their first syllable, and the last mora of the preceding prosodic word:

(457) καθαρός δέ

		H		NH	H
{ka.	tha.	rós.}	+	{X	dé.}

⁴ For an alternative analysis, and a description of some related issues, see Golston (1990, 76, 81).

	NH	H	
>	{ka.}	tha.	{ròs.} dé.}

In support of the High tone associated with ΔE *dé* is evidence from the musical notation of the Delphic Hymns. In general, in orthotone non-lexicals ‘the High-Low movement of the non-lexical accent is reduced but not eliminated’ (Devine & Stephens 1994, 363). Writing with reference specifically to MEN *mén*, ΔE *dé* and ΓΑΡ *gár*, Devine & Stephens (1994, 354) state that ‘these postpositives have grave accents with the regular lowering on nonlexicals’.⁵ This is to say that the High tone of non-lexicals is lower than the high tone of lexicals, but still higher than a Low tone. That orthotone postpositives are associated with a lowered High tone is consistent with their accent being secondary in nature, per Goldstein & Haug (2016, 302–303).

This in turn explains how such morphs can carry an accent at all, since in many cases they fall below the threshold of prosodic word minimality (§13.5.3): their accent is not primary, and the form is dependent for its primary accenthood on a neighbouring lexical, or, possibly on the projection of a primary accent together with a neighbouring non-lexical.

15.4. Rhythmic words are canonically trimoraic or greater

If rhythmic words are to be distinguished from pitch accentual words, it is worth asking how the minimality constraints on rhythmic words compare to those on prosodic words in general. As we have seen, there is general agreement in the literature that a prosodic foot in Greek is bimoraic, with the caveat that word-final consonants are ‘extrametrical’ (Golston 1991; Devine & Stephens 1994, 93; Goldstein 2016, 51). Insofar as a minimal prosodic word must consist of at least one foot, it follows that a minimal prosodic word must be bimoraic. This appears to be true, at least in terms of the calculation of the position of the pitch accent in lexicals (Blumenfeld 2011). There are, however, reasons for thinking that rhythmic words are minimally ternary rather than binary (Devine & Stephens 1994, 121, 128–129).

The significance for epigraphic word division is this. At §13.5.3 we saw that examples of bimoraic (and monomoraic) non-lexicals written as independent graphematic words are hard to find. If rhythmic words are canonically at least trimoraic, and if graphematic words correspond to rhythmic words rather than pitch accentual words, the lack of bimoraic graphematic words follows as a natural consequence: mono- and bimoraic rhythmic words would not be written as independent graphematic words because they would need to be incorporated with a neighbouring unit in order to reach rhythmic canonicity.

⁵ Cf. Goldstein (2010, 52–53), who seems to interpret Devine and Stephens differently.

The canonical ternarity of the rhythmic word follows from the fact that any alternating rhythmic sequence based on sequence length in terms of minimal units (*i.e.* *morae*) can only minimally be three units (*i.e.* *morae*) long: a bimoraic sequence would not be able to have any alternation in sequence length (cf. Allen 1973, 283; Devine & Stephens 1994, 128–129). Support for the canonical ternarity of the rhythmic foot in Greek comes from the following statement by the metrician Aristoxenus that a verse foot smaller than three *morae* is illicit (Devine & Stephens 1994, 128):

Τῶν δὲ ποδῶν ἐλάχιστοι μὲν εἰσιν οἱ ἐν τῷ τρισήμῳ μεγέθει· τὸ γὰρ δίσημον μέγεθος παντελῶς ἄν ἔχοι πυκνὴν τὴν ποδικὴν σημασίαν. Γίνονται δὲ ἰαμβικοὶ τῷ γένει οὗτοι οἱ ἐν τρισήμῳ μεγέθει· ἐν γὰρ τοῖς τρισὶν ὁ τοῦ διπλασίου μόνος ἔσται λόγος. (text Pighi 1959)

The smallest of the feet are those in the three-unit magnitude: for the two-unit magnitude would have a foot whose articulation [*sēmasia*] was completely crowded together [*pyknos*]. These feet in a three-unit magnitude are iambic in genus, for in the three there will be only the ratio of the duple [2:1 or 1:2]. (trans. Barker 1989, 189)

Insofar as the rhythms of Greek poetry, especially iambic, can be taken as a subset of the rhythms of speech (§13.4, Devine & Stephens 1994, 100–101), *mutatis mutandis* it would follow that the smallest rhythmic foot in Greek speech is also trimoraic. Given the PMH, it would follow in turn that the smallest rhythmic word would also be trimoraic.

There is an obvious question, here, however: how does minimal rhythmic ternarity square with the existence of bimoraic lexicals, both bisyllabic, *e.g.* οἷς οἷς ‘sheep’, and monosyllabic, *e.g.* παῖς *paîs* ‘child’, γῆ *gê* ‘land, earth’. Here it is important to establish what is meant by bimoraicity. For in terms of the calculation of the pitch accent, it is generally taken to be the case that a single final consonant is extrametrical (Devine & Stephens 1994, 93; Golston 2013). From the perspective of rhythm in context, however, final consonants are always metrical. Thus in the following example, the final consonants of both MEN *mén* and δὸς *dòs* render their syllables heavy before the following consonant:

(458) Eur. *I.T.* 501 (text per Diggle 1981)

οὐ τοῦτ’ ἐρωτῶ· τοῦτο μὲν δὸς τῆι τύχηι.

<i>ou</i>	<i>toút’</i>	<i>erōtô</i>	<i>toûto</i>	<i>mèn</i>
not	this.N	I_ask	this.N	PTCL

<i>dòs</i>	<i>têi</i>	<i>túkhēi</i>
give.IMP.SG	the.DAT.SG	chance.DAT
‘I do not ask for this: give this to chance’		

It is therefore relevant that whilst rhythmically heavy (*i.e.* bimoraic) monosyllabic prepositions such as EK *ek*, EN *en* and ΠΡΟΣ *prós* may regularly be found before Porson’s

bridge, there is only one instance of a trimoraic preposition standing in an equivalent position, namely ANEY *aneú* ‘without’ at O. C. 664 (Devine & Stephens 1994, 347): trimoraic non-lexicals have considerably greater rhythmic autonomy than their bimoraic counterparts.

It is similarly noteworthy that the imperative δός *dós* ‘give!’, a lexical form, may stand at before Porson’s bridge, e.g. at Eur. *I.T.* 501, quoted above, where it stands before the article. In terms of the calculation of the pitch accent, this form can be considered monomoraic, since for these purposes a single final consonant are considered extrametrical (Devine & Stephens 1994, 93; Golston 2013; although cf. Blumenfeld 2011). However, from a rhythmical perspective it is bimoraic. Its rhythmic incorporation, therefore, shows that even rhythmically bimoraic monosyllabic lexicals may, at least on occasion, be subordinated to neighbouring lexicals.⁶

15.5. Graphematic words correspond to rhythmic words

The canonical ternarity of the rhythmic word is striking given the graphematic word constraints we have observed (§13.5.2, §13.5.3): mono- and bimoraic non-lexicals are in most cases graphematically dependent, but bisyllabic trimoraic non-lexicals may be written as independent words. Compare the bimoraic preposition ΕΠΙ *epí* with the trimoraic article ΤΟΙΣΙ *toîsi*:

(459) SEG 11:314 1 (Argos, ca. 575–550 BCE)

→ ΕΠΙΤΟΝΔΕΟΝΕΝ : ΔΑΜΙΙΩΡΓΟΝΤΩΝ :

epi *tōndeōnēn*_{<ω>} *damiōrgōntōn*_{<ω>}

onthe_following serve_as_damiorgoí

‘When the following were *damiorgoi*’ (trans. Probert & Dickey 2015, 115)

(460) SEG 11:314 5–6 (Argos, ca. 575–550 BCE)

→ ΤΟΙΣΙ : ΧΡΕΜΑΣΙ : ΤΩΙΣΙ : ΧΡΕΣΤΕΡ

ΙΙΟΙΣΙ

*toîsi*_{<ω>} *khrémasi*_{<ω>} *toîsi*_{<ω>} *khrestēr*_{<λ>} *íioisi*_{<ω>}

the.DAT.PL treasures.DAT the.DAT.PL utensils.DAT

‘the treasures that are utensils’ (trans. Probert & Dickey 2015, 115)

The correlation between the constraints on the graphematic word and the smallest size of the canonical rhythmic word suggests that graphematic words in fact correspond to rhythmic words rather than pitch accentual words.

⁶ Devine & Stephens (1994, 121) remark in this connection that ‘Although monosyllabic and bimoraic feet are permissible word shapes, they are not primary feet but marked mappings’. This is to say, at the least, that such word shapes are liminal in their status as fully-fledged rhythmic words.

This proposal has the potential to explain other otherwise puzzling facts about graphematic word division in Greek inscriptions. First, I have noted that enclitic ΕΣΤΙ *estí* is written as an independent graphematic word in the Nestor's Cup inscription, and this turns out to be true in general of present indicative forms of the verb (§13.5.1.3). This is obviously a problem if graphematic words correspond to accentual units. However, if graphematic words demarcate rhythmic words, rather than pitch accentual words, the explanation emerges transparently: ΕΣΤΙ *estí* is trimoraic, and is therefore valid as an independent graphematic word, even if, from a pitch-accentual perspective, it is dependent on a preceding unit. Furthermore, all present indicative forms of the verb 'to be' are bisyllabic and trimoraic, except the second person singular form ΕΙ *éí*.

The second puzzling fact explained by the suggestion that graphematic words correspond to rhythmic words is that pitch accentual enclitics may be univerted with a following rather than a preceding lexical (§13.5.1.5). We have seen that the evidence for labile clitic polarity in Ancient Greek comes from metrical, that is, rhythmical, compositions, where the presence of enclitics at metrical bridges shows that these particles must associate rightward on occasion. Despite the fact that such an arrangement does not correspond to what one would expect on pitch accentual grounds, graphematic words in inscriptions are also known to write enclitics as postpositives, as we saw in the case of the pronoun ΜΕ *me*. However, if graphematic words correspond to rhythmic words, rather than pitch accentual words, the problem disappears.

I observe in closing that, although the canonical rhythmic word is ternary, this does not mean that bimoraic rhythmic words cannot be found. Where bimoraic units are demarcated as graphematic words in their own right, these tend to be nouns, *e.g.*:

(461) Peek (1957) 207, 7–8 (400–300 BCE)

→ ΜΥΡΤΑΛΗ

ΓΡΑΥΣΚΑΙ : ΠΑΙΣ : ΘΕΟΓΕΝΗΣ

*murtálē*_{<λ>} *graûs=kai*_{<ω>} *paîs*_{<ω>} *theogénēs*_{<λ>}

PN old_woman=and child PN

'Myrtale, old lady, and child, Theogenes'

(462) IG I³ 1084 (end of 5th century BCE; vowel lengths this author)

→ ΗΙΕΡΟΝ :

ΔΙΟΣ : ΜΙ

ΛΙΧΙΟ : <Γ>

ΗΣ : ΑΘΗΝ

ΑΙΑΣ

*hieròn*_{<ω>} *diòs*_{<ω>} *mi*_{<λ>} *likhiō*_{<ω>} <g>_{<λ>} *ēs*_{<ω>}

temple.NOM.N.SG DN.GEN gracious.GEN.SG land.GEN

*athēn*_(λ) *aías*_(ω)

DN.GEN

‘Temple of Zeus the gracious, of Ge, and of Athena’ (trans. with ref. to Hallof; see Lewis & Jeffery 1994 online)

Devine & Stephens (1994, 327) (cf. §13.5.3) also point out that bimoraic non-lexicals may be written as independent graphematic words when they are followed by a branching constituent (§13.5.3). A possible reason for this is that, because of the phenomenon of edge alignment, the non-lexical in such contexts is followed immediately by a prosodic phrase boundary, which would necessitate the non-lexical standing as an independent rhythmic word, conceivably followed by a pause.

15.6. Conclusion

The goal of this chapter was to account for the apparent disagreement between the prosodic domain implied by tonal spread from enclitics, on the one hand, and that implied by occasional rhythmic and graphematic proclisis of postpositives, on the other. This behaviour is problematic because it has been assumed up to now that the pitch accentual and rhythmic domains must be coextensive. However, I have argued that if one allows pitch accentual and rhythmic words to be at least partly independent of one another, it is possible to provide an account of behaviour that is otherwise impossible to reconcile other than by invalidating the inherited accent tradition.

If this is correct, the implication for the semantics of punctuation is that graphematic words in Greek word-punctuating inscriptions correspond to rhythmic words, rather than pitch accentual words. If it is not granted for pitch accentual and rhythmic domains to have differing extents, the implication of the optional graphematic proclisis of postpositives is that graphematic word division corresponds to prosodic domains as indicated by the rhythm of metrical compositions, and not those of the inherited accentual tradition. Either way, graphematic word division follows the structure of rhythmical compositions over against that implied by the pitch accent.

This conclusion fits well with the broader Northwest Semitic context of alphabetic writing. As previously remarked (§13.5.1.1), the accent in Northwest Semitic was likely one of stress. We have seen that rhythmic prominence has some affinities with stress in stress languages (§15.3.2). It is therefore possible to imagine that word division in Greek inscriptions was understood by its adapters to mark out rhythmic units rather than tonal ones.

Chapter 16

Graphematic words with multiple lexicals

16.1. Introduction

The preceding chapters of Part IV have addressed the problems with associating prosodic words with graphematic words in word-punctuating Greek inscriptions. Most of these are resolved if pitch accentual words and rhythmic words are distinguished, and graphematic words are identified with rhythmic words rather than pitch accentual words. One difficulty remains, however: while graphematic words need not contain any lexicals (§13.5.2), they may comprise multiple lexicals. In the inscriptions we have considered for this chapter, two principal types may be identified:

- Noun + Verb
- Noun + dependent genitive

16.1.1. Noun + dependent genitive

We find two examples of this kind of lexical-lexical univerbation. The first is the following:

(463) IG I³ 699 (Athens, 500–480 BCE; vowel lengths this author)

→ : ΗΟΣΜΙΚΥΘΟΗΥΙΟΣ

ho=	smikuthō	huiós
the	Smikuthos.GEN	son.NOM
‘the son of Smikuthos’		

In this case a single noun phrase is univerbation, comprising a determiner, a dependent genitive and the head noun.

The second case is similar:¹

¹ Van Effenterre’s translation (Van Effenterre 1961, 547) reads: ‘s’il leur arrivait de chasser, il a été décidé combien (ils chasseraient): pour la battue, le mois d’Hyperboios, au vingtième jour, sera la limite’.

(464) SEG 23.530 (Dreros, 650–600 BCE; original text is written in ‘boustrophedon’, with the first line written right-to-left, and the second written left-to-right; the first (reading) line is also placed below the second)

→ ... ΕΦΑΔΕ | ΟΖΑ | ΕΛΑΣΙΤΟΥΠΕ
 ΡΒΟΙΟ | ΜΗΝΟΣ | ΕΝΙΚΑΔΙ | ΟΡΟΝΗΜΕΝ

*éwade*_{<w>} *óza*_{<w>} *elási=* *tô=*
 decree.AOR.IND.ACT.3SG PTCL.REL **driving.DAT.SG** **the.GEN.SG**

*upe*_{<λ>} *rboío*_{<w>} *mēnòs*_{<w>} *en=ikádi*_{<w>} *óron=*
hyperboios.GEN.SG month.GEN.SG in=twenty limit.ACC.SG

*êmen*_{<w>}
be.PRS.INF

‘... it has been decreed how much (they should hunt): **for the driving (out of the animals)**, the limit is to be on the twentieth day **of the month of Hyperboios**’ (trans. with reference to Van Effenterre 1961, 547)

This inscription univerbates the dative noun ΕΛΑΣΙ *elási* with the genitive noun phrase ΤΟΥΠΕΡΒΟΙΟ *tô=uperboío*. ΤΟΥΠΕΡΒΟΙΟ *tô=uperboío* is graphematically phrased together with the dative ΕΛΑΣΙ *elási*. If the passage is interpreted per Van Effenterre 1961, 547, word division is out of step with the syntactic phrasing: Van Effenterre’s translation suggests that ΤΟΥΠΕΡΒΟΙΟ *tô=uperboío* is to be taken with ΜΗΝΟΣ *mēnòs*. ΤΟΥΠΕΡΒΟΙΟ | ΜΗΝΟΣ *tô=uperboío*_{<w>} *mēnòs* ‘of the month of Hyperboios’ would be a genitive phrase giving the period during which the driving out is to take place (cf. George 2014, 308–310).

An alternative, which accounts fully for the punctuation, is to take ΕΛΑΣΙΤΟΥΠΕΡΒΟΙΟ *elási tô uperboío* as a single (prosodic and syntactic) unit, *i.e.* ‘for the driving out (in the period) of (the month of) Hyperboios’. The whole would then be translated:

it has been decreed how much (they should hunt): for the driving (out of the animals) in (the month of) Hyperboios, the limit is to be on the twentieth day of the month.

16.1.2. Noun + Verb

16.1.2.1. SEG 14.604 ‘Nestor’s Cup inscription’

There are two instances of Noun–Verb univerbations in the inscriptions considered for this study. The first occurs in the Nestor’s Cup inscription. The first line of this inscription is given above (426). The remaining lines are given here:

(465) SEG 14.604:2–3 (Ischia, 8th century BCE; original right-to-left; accents this author)

→ ΗΟΣΔΑΤΟΔΕΠΙΕΣΙ : ΠΟΤΕΡ[ΙΟ] : ΑΥΤΙΚΑΚΕΝΟΝ
 ΗΙΜΕΡΟΣΗΑΙΡΕΣΕΙ : ΚΑΛΛΙΣΤΕΦΑΝΟ : ΑΦΡΟΔΙΤΕΣ

hòs= *d'*= *à<n>*= *tôde*= *piēsi*_{<ω>}
 who PTCL PTCL this.GEN drinks

*potērīō*_{<ω>} *autika*= *kēnon*_{<λ>} *hímeros*=
 cup.GEN immediately that.ACC desire.NOM

*hairēsei*_{<ω>} *kalliste[phá]nō*_{<ω>} *aphrodítēs*_{<ω>}
 seize.FUT.3SG beautifully_crowned DN.GEN

‘but he who drinks from *this* cup, forthwith him will seize desire of fair-garlanded Aphrodite’ (trans. Watkins)

In general the inscription is word-punctuating. The following are all written as independent graphematic words:

- ΝΕΣΤΟΡΟΣ *néstoros* (1)
- Ε[ΣΤ]Ι *e[st]i* (1)
- ΕΥΠΟΤ[ΟΝ] *eúpot[on]* (1)
- ΠΟΤΕΡΙΟΝ *potérion* (1)
- ΠΟΤΕΡ[ΙΟ] *potēr[iō]* (2)
- ΚΑΛΛΙΣΤΕΦΑΝΟ *kallistephánō* (3)
- ΑΦΡΟΔΙΤΕΣ *aphrodítēs* (3)

There are, however, three instances in which longer units appear to be punctuated as one:

- ΗΟΣΔΑΤΟΔΕΠΙΕΣΙ : *hòs=d'=à<n>tôde=piēsi*_{<ω>} (2)
- : ΑΥΤΙΚΑΚΕΝΟΝ *autika=kēnon*_{<λ>} (2)
- ΗΙΜΕΡΟΣΗΑΙΡΕΣΕΙ : *hímeros=hairēsei*_{<ω>} (3)

It might be supposed that in these cases it is prosodic phrases rather than prosodic words that are punctuated. This is in view of the fact that ΗΑΙΡΕΣΕΙ *hairēsei* could be seen as (syntactic) phrase-final. Before reaching for this conclusion, it is worth observing that only one instance involves two lexicals, namely ΗΙΜΕΡΟΣΗΑΙΡΕΣΕΙ : *hímeros=hairēsei*_{<ω>}, and in this case it is a Noun + Verb sequence.

An important consideration is that the inscription is metrical. Devine & Stephens (1994, 400) observe that ‘some metrical inscriptions punctuate words or appositive groups’, whilst in others ‘the punctuated element may be not only a word or appositive group but also an extended appositive group or an easily formed minor phrase’, giving the Nestor’s Cup inscription as an example of the latter kind. They account for this vacillation by supposing (p. 401) that ‘the punctuation of these metrical inscriptions apparently accesses a slower rate of speech in which

single words are less readily joined into minor phrases'. In other words, for Devine and Stephens metrical inscriptions essentially punctuate (minor) prosodic phrases, but at a rate of speech slow enough that many individual words are pronounced as though they were minor prosodic phrases. Since one of the distinguishing features of prosodic phrases, as opposed to prosodic words, is the presence of pause, the implication is that graphematic word separation corresponds to pauses in recitation.

Devine and Stephens' inference could in principle, however, be turned on its head, so that these inscriptions punctuate prosodic words, that is, autonomous rhythmic units, but in parts the rate of speech accessed is such that certain elements are subordinated to others in the recitation.

Either way, it may be significant that two of the three cases of longer unit punctuation involve verbs, and the other involves two non-lexicals. By contrast, those morphosyntactic units demarcated as single graphematic words in this inscription are either nouns or adjectives. (The only exception to this is Ε[ΣΤ]Ῑ ε[st]ī (1), for discussion of which see §13.5.1.3.) This point is developed at §17.3.3 below.

16.1.2.2. SEG 23.530 (Dreros)

The graphematic subordination of verbal forms is not, however, limited to metrical inscriptions. SEG 23.530, introduced earlier at (427), is a legal inscription from Dreros in Crete. This provides another example of Noun-Verb univerbation:

(466) SEG 23.530 (Dreros; 650–600 BCE; original text is written in 'boustrophedon', with the first line written right-to-left, and the second written left-to-right; the first (reading) line is also placed below the second)

→ ... ΕΦΑΔΕ | ΟΖΑ | ΕΛΑΣΙΤΟΥΠΕ
 ΡΒΟΙΟ | ΜΗΝΟΣ | ΕΝΙΚΑΔΙ | ΟΡΟΝΗΜΕΝ

<i>éwade</i> _{<ω>}	<i>óza</i> _{<ω>}	<i>elási=</i>	<i>tô=</i>
decree.AOR.IND.ACT.3SG	PTCL.REL	driving.DAT.SG	the.GEN.SG

<i>upe</i> _{<λ>} <i>rboíō</i> _{<ω>}	<i>mēnòs</i> _{<ω>}	<i>en=ikádi</i> _{<ω>}	<i>óron=</i>
hyperboios.GEN.SG	month.GEN.SG	in=twenty	limit.ACC.SG

*êmen*_{<ω>}
be.PRS.INF

'... it has been decreed how much (they should hunt): for the driving (out of the animals), **the limit is to be** on the twentieth day of the month of Hyperboios' (trans. with ref. to Van Effenterre 1961, 547)

In this case the subject and verb are in a dependent infinitive clause controlled by ΕΦΑΔΕ *éwade* 'it has been/was decreed'. As with the Nestor's Cup inscription, the inscription is, for the most part, word-punctuating.

16.1.2.3. Summary

In principle the following interpretations are available:

- The inscriptions in question punctuate word-level units for the most part, but occasionally punctuate prosodic phrase level units. This is to say that the inscriptions are inconsistent;
- The inscriptions are consistent in demarcating prosodic words, in particular, rhythmic words, and the univerbation of lexicals indicates the prosodic subordination of one lexical to another.

These possibilities are explored in the following subsections.

16.2. Inconsistency of levels of graphematic representation

That the inscriptions in question are simply inconsistent in their level of phonological representation – sometimes punctuating prosodic words, and sometimes punctuating prosodic phrases – is an *a priori* plausible explanation of the facts, since inscriptions that are inconsistent in this respect clearly exist (Devine & Stephens 1994, 389–390). The question, however, is how to interpret this inconsistency, *i.e.* to ask where the inconsistency lies. In principle there are once again two possibilities:

- The author/writer of the inscription varies the prosodic level of the punctuation, sometimes punctuating prosodic words and sometimes punctuating prosodic phrases;
- The author/writer is consistent in representing the same prosodic level, *e.g.* prosodic words; the variation arises out of variable phrasing in natural speech (cf. Devine & Stephens 1994, 389).

That at least in some cases the first explanation is correct is suggested by the fact that there are inscriptions where the phonological level of punctuation appears to vary widely. Thus Devine & Stephens (1994, 389) report that LSAG 15.4 ‘seems to start out with word punctuation, proceed to minor phrase punctuation and end up with major phrase punctuation’.² However, it is worth considering the second possibility, that the author/writer is in fact consistent, before resorting to the inconsistency of the author/writer.

Cross-linguistic support for this possibility comes from the Northwest Semitic material considered in the earlier chapters of this study, where I have argued that the punctuation of multiple lexicals as a single graphematic word in both Ugaritic and Phoenician can be profitably compared to the distribution of actual prosodic words in the cantillation tradition of the Masoretic Text. This is to say that the

² In their study Devine & Stephens (1994) distinguish three prosodic levels above the word: the appositive group, the minor phrase, and the major phrase. In our terms these correspond to the (recursive) prosodic word, the prosodic phrase, and the intonational phrase.

morphosyntactically inconsistent distribution of word dividers can be accounted for with reference to variable prosodic phrasing, rather than to inconsistency on the part of the inscriptions' authors or writers. It would be desirable, however, to ground such an argument not only in cross-linguistic evidence from Northwest Semitic, but also in evidence from within Greek itself. It is to this task that I now turn.

16.3. Prosodic subordination of one lexical to another

16.3.1. General lexical-lexical subordination in Ancient Greek

If graphematic words consistently correspond to prosodic words in these inscriptions, the univertation of two lexicals should imply that the resulting unity comprises a single prosodic word. This would entail the following lexical-lexical subordinations:

- Noun + verb
- Noun + dependent genitive

We have already seen evidence for the prosodic incorporation of bimoraic lexicals (§15.4). There is also slight evidence for the prosodic subordination of larger lexical units (Goldstein 2010, 52). In the following line from Sophocles *Ajax*, the verbal form ἡγεῖτ' *hēgeit'* stands immediately before Porson's bridge:³

(467) Soph. *Ajax* 1101 (text per Finglass 2011; discussed at Goldstein 2010, 52)

→ ἔξεστ' ἀνάσσειν ὧν ὄδ' ἡγεῖτ' οἴκοθεν;

<i>éxest'</i>	<i>anássein</i>	<i>hôn</i>	<i>hód'</i>
be_permitted.PRS.3SG	command.PRS.INF	whom.GEN.PL	this.M.SG

<i>hēgeit'</i>	<i>oíko-then</i>
lead.IPRF.IND.NACT.3SG	home-from

'[How] can [you] command [the men] that he led from his home?' (trans. after Finglass 2011, 447)

Since a heavy word-final syllable is not normally permitted before Porson's bridge, it is assumed that the two constitute a single prosodic unit, albeit not necessarily a single prosodic word (Goldstein 2010, 52). Such an interpretation is helped by the elision, which could be taken to suggest that the two constitute a single prosodic phrase (cf. Goldstein 2010, 22; for the domains of elision, see Devine & Stephens 1994, 262–265).⁴

³ Square brackets in the translation indicate elements from the previous line.

⁴ The 'violation' of Porson's Law at this point has of course induced several suggested emendations, for which see Finglass (2011) *ad loc.* There is also a variant reading for ἡγεῖτ' *hēgeit'*, namely ἡγαγ' *égag'*.

It should be emphasised, however, that the evidence for such lexical–lexical prosodic subordination is not plentiful: examples of ‘violations’ of Porson’s Law involving a sequence of two lexicals are very limited.⁵ It is therefore worth trying to adduce other evidence.

16.3.2. *Lexicalised expressions*

Fixed phrases – that is, lexicalised expressions – are known cross-linguistically to have the properties of prosodic words rather than prosodic phrases (Devine & Stephens 1994, 348–349): whereas a phrase might be expected to have one prosodic prominence or ‘accent’, fixed phrases often have only one. In an inscription punctuating prosodic words, such phrases might be expected to be written as a single graphematic word. Evidence for the univerbation of fixed phrases in the history of Greek comes from the Mycenaean corpus, where the sequence *pa-si-te-o-i* ‘to all the gods’ is written as a single graphematic word (Morpurgo Davies 1987; Devine & Stephens 1994, 344–345).

Phrase-lexicalisation could offer an explanation of the instances of the univerbation of a noun with attributive or dependent genitive phrase. IG I³ 5 at (431) *ermeî enagōníōi* is very plausibly seen as an example of this kind. Furthermore, if the interpretation suggested at §17.1.1 is adopted, the phrase ΕΛΑΣΙΤΟΥΠΕΡΒΟΙΟ *elási tô uperboío* ‘for the driving out of (the month of) Hyperboios’ might amount to a quasi-lexicalised expression, if this was a recognised event in the society of Dreros. Finally, a similar argument could be made for the patronymic phrase ΗΟΣΜΙΚΥΘΟΗΥΙΟΣ *ho smikúthō huiós* ‘the son of Smikuthos’ at (432): if Onesimus was widely known as *ho smikúthō huiós*, it could well be imagined that this phrase would function as a single prosodic word.⁶

16.3.3. *Lower degree of phonological prominence in verbs*

The other type of univerbation we see in the Greek inscriptions is that of a noun with a following verb. Verbs are cross-linguistically less phonologically prominent than nouns (Devine & Stephens 1994, 303–304; Fortson 2010, 110). This is seen within the history of Indo-European, where finite verbs in main clauses are generally unaccented (Devine & Stephens 1994, 303; Barrett 2001[1964], 426; Fortson 2010, 109–110). In the history of Greek, clitic-like behaviour of main clause verbs beyond EIMI *eimí* and ΦΗΜΙ *phēmi* can be seen in the Mycenaean texts. Here main clause verbs are regularly drawn to second position after an introductory particle (Thompson 2010, 197), *e.g.*:

⁵ The only other case of which I am aware is that of the first line of Euripides *Ion* (Irvine 1997; Finglass 2011, 448).

⁶ It may be relevant that in the Iambographers, while four-syllable words with a heavy first syllable are generally avoided in the third metron, there is one example of a proper name, Γορτυνίης, in this position at Archilochus 24.2 (Devine & Stephens 1994, 143; text West 1971). This perhaps shows a greater tendency for subordination in personal names than in other nouns.

(468) PY Ta 711.1 (Pylos; text from DAMOS, Aurora 2015; word division indicated by commas in syllabic transcription)

o-wi-de, pu₂-ke-qi-ri, o-te, wa-na-ka, te-ke, au-ke-wa, da-mo-ko-ro

*hō=wide*_{<ω>} Phugegwri_{<ω>} *hote*_{<ω>} *wanax*_{<ω>}
 thus=saw.AOR.IND.ACT.3SG PN when king.NOM.SG

*thēke*_{<ω>} Augewān_{<ω>} *dāmokoron*_{<λ>}
 appoint.AOR.IND.ACT.3SG PN *dāmokoros*

‘Thus Phugegwri_{<ω>} saw when the king appointed Augewān_{<ω>} as dāmokoros’ (trans. Thompson 2010, 197)

In this example a contrast can be observed between the second position of the main clause verb *wide* ‘saw’ hosted by the introductory particle *hō*, with the non-initial position of *thēke* ‘appointed’ in the subordinate clause.

In sentences without the introductory particle *hō*, the verb may occur in second position after the subject, e.g.:

(469) PY Ep 704.5 (Pylos; text from DAMOS, Aurora 2015; word division indicated by commas in syllabic transcription)

e-ri-ta, i-je-re-ja, e-ke, e-u-ke-to-qe, e-to-ni-jo, e-ke-e,

*Erithā*_{<ω>} *hierēia*_{<ω>} *hekhei*_{<ω>} *eukhetoi=k^we*_{<ω>}
 PN priestess.NOM.SG has claims.NACT=CONJ

*etonion*_{<ω>} *hekhehen*_{<ω>}
 superior_lease have.PRS.INF.ACT

‘Eritha the priestess has and claims to have a superior lease’ (trans. Rupert Thompson, pers. comm.)

Note in this case that, if *hekhei* is hosted by the first prosodic word in the sentence, the sequence PN–Noun must constitute a single prosodic word.

Evidence for the rhythmically weaker status of verbs as opposed to nouns in Classical Greek emerges from their treatment in the Iambographers, where of four-syllable words with heavy first syllable, only verbs may stand in the third metron (Devine & Stephens 1994, 143), e.g.:

(470) Semonides 7.109 (text per West 1972)

→ ἦτις δέ τοι μάλιστα σωφρονεῖν δοκεῖ,
 αὕτη μέγιστα τυγχάνει **λωβωμένη**·

hētis *dé* *toi* *málista* *sōphroneîn*
 INDEF.NOM.F.SG PTCL PTCL very_much be_wise.PRS.INF

*dokēi*_(λ) *hautē* *mégista* *tugkhánei*
 seem.PRS.IND.ACT.3SG this.F.NOM.SG most_greatly happen.PRS.IND.ACT.3SG

lōbōménē

maltreat.PTCP.NACT.NOM.F.SG

‘She who seems to be the most discreet, that one is often the one to be most grievously **maltreated**’

If it is not by chance that verbs are the only word class to occur in this position, the restriction provides evidence of weaker prosodic status for verbs because four-syllable words with heavy first syllable are generally avoided in the third metron (Devine & Stephens 1994, 106). That verbs appear in this position, and nouns do not, would suggest that the first syllable of the former were capable of subordination (*i.e.* shortening) (Devine & Stephens 1994, 143).

It may be relevant in this connection that in the Nestor’s Cup inscription *HAIPEΣEI hairései* ‘will seize’ occurs second in the line, after the subject noun *HIMEPOΣ hímeros* ‘desire’. If the verse line serves as an intonational phrase (§15.3.4, Goldstein 2010, 99; Goldstein & Haug 2016), this main clause verb could be seen to be hosted by the first prosodic word in the line, splitting the noun phrase *HIMEPOΣ ... ΚΑΛΛΙΣΤΕ[ΦΑ]ΝΟ ΑΦΡΟΔΙΤΕΣ hímeros ... kallistephánō Aphrodítēs*. If hyperbaton in Greek is phonological movement (Agbayani & Golston 2010), and head-initial ordering is pragmatically neutral (Agbayani & Golston 2010, 138), the second line of the inscription could be seen as focus-preposing of the subject *HIMEPOΣ hímeros* before the verb, with the verb cliticising to that subject.

16.3.4. Different kinds of rhythmic subordination

The explanations given in the foregoing sections aim merely to show that possible motivations for lexical–lexical subordination can be found. This is not to argue, however, that these explanations offer a complete solution. It should be highlighted, for example, that there is a disparity in the treatment of the verb *EIMI eimí* ‘to be’ compared with the other verb forms: while the latter are graphematically dependent, the former is graphematically independent (§13.5.1.3). Earlier I accounted for the graphematic independence of *EIMI eimí* ‘to be’ by distinguishing between rhythmic and accentual dependence, proposing that *EIMI eimí* ‘to be’ was rhythmically independent, and that it is rhythmic rather than accentual units that are punctuated (§15.5).

It is therefore worth following up these points with the observation that the three lines of the Nestor’s Cup inscription likely represent two kinds of metre: whilst the first is likely a (slightly irregular) iambic trimeter, whilst the second two lines are, forming as they do two hexameter lines (Watkins 1976, 34, with references; Hackstein 2010, 418–419). Indeed, the first word divider in each verse line corresponds to the

position of a caesura (Watkins 1976, 34).⁷ This supports the suggestion that the punctuation demarcates rhythmic units, but, if correct, also implies that what might be understood to count as an rhythmic word in hexameter might be different from that which might be identified in prose or iambic texts.

16.4. Punctuating canonical rhythmic words

In §16.3 I sought to provide evidence from within the Greek language for a prosodic account of the univertation of multiple lexicals in a subset of Greek inscriptions. I noted, however, that not all word-punctuating inscriptions univertate multiple lexicals (§13.5.4). In particular, IGA 497 (Teos), IGA 499 (Ephesus), SEG 11:314 (Argos) and IG I³ 5 (Eleusis) are consistent in not univertating lexicals in this way. The rationale for punctuation in these inscriptions has been observed to be related to the accentual status of a given morpheme: clitics are univertated with lexicals, while orthotonic words may stand as independent graphematic words (Kaiser 1887, 19; Morpurgo Davies 1987).

We have already seen that this account of graphematic word division, relying, as it does, on a classical understanding of the tonicity of pre- and postpositives in Greek, is in need of some adjustments (Chapter 15). In particular, I have argued that it is a morpheme's *rhythmic* status, rather than its accentual status, that is key to determining its treatment in graphematic word division. Only by making this adjustment is it possible to account both for the obligatory second-positioning of postpositives, while allowing for alternating clitic polarities in both verse and in inscriptional punctuation.

How, though, are we to account for the fact that certain inscriptions display a much greater level of consistency in terms of the prosodic level of punctuation compared to others? In Chapter 11 and Chapter 12 I argued, on the basis of the discrepancy between graphematic words in the Masoretic consonantal text, on the one hand, and the prosodic words of the cantillated text, that graphematic words in the consonantal Masoretic Text and in the Meshah^c stela correspond to minimal prosodic words, where each graphematic word comprises at most one lexical (Chapter 12, §13.5.4). Here I propose a similar explanation for the difference in the punctuation of Greek inscriptional texts, namely, that these inscriptions punctuate CANONICAL RHYTHMIC WORDS. A canonical rhythmic word must be at least bisyllabic and trimoraic if non-lexical; lexicals may fall below this threshold, but must still be minimally bimoraic (§15.4, §15.5). Accordingly, non-lexicals are never in these inscriptions written as independent graphematic words unless they are minimally bisyllabic and trimoraic, whilst a given graphematic word may have a maximum of a single lexical. If a non-lexical or a sequence of non-lexicals meet the threshold of canonical rhythmic wordhood, they may, at least in IGA 499 and SEG 11:314 be written as independent graphematic words.

⁷ I am much indebted to Torsten Meissner, pers. comm., for first alerting me to this point.

Such a proposal should, however, be regarded as preliminary: the number of inscriptions of length that adopt such an orthography is relatively few. It is therefore desirable to amass a greater dataset to test it. Nevertheless, the explanation appears to fit the facts of Ancient Greek prosody.

Of course, any explanation of word division in written texts must ultimately find its origin in the social and physical context(s) of a given inscription's creation. This topic is taken up in the Conclusion (Chapter 17).

16.5. Conclusion to Part IV

In Part IV I have explored the demarcation of prosodic words in Greek word-punctuating inscriptions. In these inscriptions appositive non-lexicals are regularly written either together with a neighbouring lexical, or as a single graphematic word with neighbouring non-lexicals. The units so grouped make sense as prosodic units – that is, as one that, from a cross-linguistic perspective, might be expected to share a primary accent – but little sense as syntactic units. These facts strongly support a prosodic interpretation of the facts.

While the prosodic word hypothesis appeals at a general level, when it comes to the details of its relationship to Ancient Greek prosody, and especially to the pitch accent, the following problems are encountered:

- Enclitic forms may be written as separate graphematic words (§13.5.1.3);
- For a given appositive, conventional polarity is not necessarily maintained, and frequently does not correspond what is traditionally accepted to be the case (13.5.1.5);
- Graphematic words may comprise multiple lexicals (§13.5.1.4), while:
- Graphematic words need not comprise any lexicals at all (§13.5.2).

The first two issues were addressed by investigating the relationship between accent and rhythm in Ancient Greek, with two possibilities considered. The first was that appositives may indeed be labile in their polarity, with enclitics, notably enclitic pronouns, retaining the possibility of switching polarity in certain contexts, and becoming proclitics (§15.2). Such a view would entail the complete loss of knowledge of such a state of affairs in the grammatical tradition. The second possibility was to distinguish between the domains of the pitch accent and that of rhythm: while pitch accentual polarity is maintained in Ancient Greek, rhythmic polarity need not be (§15.3). This necessitated the positing of two prosodic word-level units in Greek, the rhythmic word and the pitch accentual word. Evidence for the independence of rhythmic and pitch accentual domains was provided both on both cross-linguistic and Greek-internal grounds. Cross-linguistic evidence was adduced from Japanese, where the domains of pitch accent and rhythm are in principle independent (§15.3.2). Evidence from within Greek was provided from the obligatory non-initial position of orthotonic postpositives (§15.3.4). In §15.4 I argued that if graphematic words are

taken to represent rhythmic words rather than pitch accentual words, a number of the difficulties associated with the identification of graphematic words with prosodic words are resolved. The major remaining issue was the fact that graphematic words can, on occasion, comprise multiple lexicals, an *a priori* unexpected result. For the instances found in the inscriptions considered in this chapter a number of explanations were proposed. Ultimately, however, in order to account for the data it seems one must accept a certain level of fluidity between levels of phonological representation (§16.2).

Chapter 17

Epilogue: The context of word division

17.1. Overview

This study set out to explain why it is that in Northwest Semitic and Greek alphabetic inscriptions from the late 2nd through to the first half of the 1st millennium BCE words are divided in ways that, to a person rooted in Western European traditions, seem strange. In both contexts smaller function morphemes are normally written together with neighbouring words, a practice that is quite alien to modern Western European writing, as well as to modern editors of Ancient Greek texts. Furthermore, word division practices have seemed inconsistent to modern scholars, a fact that has often led to implicit or explicit criticism of the original writers of the documents, *i.e.* that they were not sufficiently skilled or ‘developed’ to know how to separate words correctly.

The present study has started from the premise that the writers of these texts were skilled at their craft and aimed for consistency in word division. From this starting point the goal has been to find the linguistic level targeted by word division. In principle word division might be governed by principles operating at one of several linguistic levels, including morphosyntax, phonology/prosody, graphematics, and semantics. In the Greek and Semitic alphabetic writing investigated here, however, we have seen that for the most part it is phonology/prosody that is most significant: graphematic words correspond either to prosodic words or prosodic phrases. The appearance of inconsistency arises because the regular application of prosodic principles in word division, whereby each graphematic word corresponds to an actual graphematic word in context, can result, at least in the Northwest Semitic context, in inconsistent word division from the perspective of morphosyntax. The only exception was found to be the ‘Minority’ orthography of the Ugaritic texts, where graphematic words correspond to morphosyntactic words, *i.e.* to words as we might expect to find them in an English text.

The aim in the Conclusion is to explore the implications of these findings both for our understanding of the world in which these writing systems were situated and for our understanding of the development of writing systems.

17.2. Orality and literacy

The fact that word division has been shown to conform to prosodic principles in both Northwest Semitic and Ancient Greek inscriptions does not mean or require that the writers of these inscriptions had a well-developed linguistic understanding of the language(s) that they were writing. It is therefore worth asking how the cultural environment in which they were operating resulted in a system that produced prosodic word division.

In modern western societies we are used to drawing a sharp distinction between the written and the spoken word. This is shown, not least, by the fact that we are taught to read and write silently. The context of writing in the ancient world was very different from this. Writing practices there were rooted in orality, and presupposed an oral interaction with the written word. As Thomas (1992, 74) puts it:

the written word in the ancient world often has such a close relationship to the background of oral communication that it cannot properly be understood in isolation from that background.

This may be said to have been the case both in Late Bronze Age Ugarit and in the Iron Age societies of Greece and the Levant. While the development and spread of (alphabetic) writing may have had some effect on literacy, that effect need not have been to replace orality with literacy. Instead the two seem to have coexisted (Thomas 1992; Whitley 2021, 280).

The present study has considered three contexts for word division: Late Bronze Age Ugarit, the Iron Age Levant and the Ancient (Archaic and Classical) Greek world. The oral context of writing in each of these is now briefly surveyed in turn.

17.1.1. Late Bronze Age Ugarit

Although literacy was likely the preserve of a small minority in Ugaritian society (Boyes 2021, 194, 278–279), the oral and written word were heavily interdependent (Boyes 2021, 189–195).¹ In regard to poetic literary texts there has been considerable debate concerning the date of their composition, especially those of the Baʿl Cycle. On the one hand it has been argued that, in considerable measure, these poems are the original composition of a certain ʾIlimilku living in the 13th century BCE (*e.g.* Wyatt 2005, 251–252; Pardee 2012; Tugendhaft 2018, 29, 41 n. 11). Others, by contrast, prefer to see ʾIlimilku as a (primarily) a copyist acting at the end of a long line of oral composition (Greenstein 2014, 216–217; for an overview, see Curtis 2016). Whatever one thinks of the origin of the epics, however, it seems highly probable that they were intended to be performed orally, and perhaps even composed orally too, in

¹ The debate concerning the relationship between text and oral tradition in Ugarit has taken place within wider consideration of the question in the Mesopotamian context. While some scholars have sought to emphasise the textual elements of the poetic literary tradition in Mesopotamia, others have highlighted the oral element, particularly in performance (for details and references, see Boyes 2021, 189–195).

conjunction with being written down (Boyes 2021, 193; citing Redford 2000). Nor was oral performance limited to literary texts. Letters are assumed to have been dictated and then read aloud to the person to whom they were written (Boyes 2021, 193). Ritual and magical texts have a close relationship to oral performance, and this may well have been the case for legal documents too (Boyes 2021, 193–194).

17.2.2. Iron Age Levant

The earliest phases of the Biblical tradition were most likely oral in nature (Schniedewind 2004, 52–53). However, these had likely entered a written tradition by the 10th century BCE (Schniedewind 2004, 55–56). Yet the existence of a textual tradition does not negate the largely oral nature of interaction with the written word on the part of the majority of the population until a late date. We see this in action in the Biblical book of Nehemiah (passage cited at Schniedewind 2004, 48–49):

(1) Neh 8:3

וַיִּקְרָא-בוֹ לְפָנָי הַרְחֹב אֲשֶׁר ׀ לְפָנָי שַׁעַר-הַמַּיִם מִן-הָאֹרֶחַ עַד-מַחְצִית הַיּוֹם נֹגֵד
הָאֲנָשִׁים וְהַנְּשִׂים וְהַמְּבִינִים וְאִזְגֵי כָל-הָעָם אֶל-סֵפֶר הַתּוֹרָה:

‘And he [*i.e.* Ezra] read therein before the street that [was] before the water gate from the morning until midday, before the men and the women, and those that could understand; and the ears of all the people [were] attentive unto the book of the law’ (KJV)

Compare also Nehemiah 8:8:

(2) Neh 8:8

וַיִּקְרְאוּ בְּסֵפֶר בְּתוֹרַת הָאֱלֹהִים מִפֶּה שׁוֹם וְשׁוֹם שָׁכַל וַיְבִינּוּ בְּמִקְרָא: ׀

‘So they read in the book in the law of God distinctly, and gave the sense, and caused [them] to understand the reading’ (KJV)

As Schniedewind (2004, 48–49) points out, the word for ‘reading’ here is the verb *קרא* *qr* with its original meaning of ‘call out, proclaim’: reading was the act of proclaiming something written on the page.

This long-lived interplay of the oral and the written is consistent with the fact that the oral components of the Hebrew Bible, especially the traditions of vocalisation and cantillation, have histories related to but independent of that of the consonantal text, where we see evidence of the interplay of oral and written traditions within the text of the Hebrew Bible itself (§1.7.4.1).

17.2.3. Ancient Greece

Literacy and orality also co-existed in the Greek setting. While several studies in the 20th century sought to present the arrival of alphabetic writing as something of a

watershed moment for the Greek-speaking world by entailing the supplanting of orality with literacy (for surveys see Thomas 1992), over the last few decades there has been a growing recognition that orality and literacy are not mutually exclusive. Instead the two seem to have co-existed in Greek-speaking societies over a long period of time (Thomas 1992). Recent studies have also shown that the period between the literacy of the Mycenaean palaces which used the Linear B syllabary and the adoption of alphabetic writing – previously thought to have been of the order of 300 years in length – may have been much less than that, if, in fact, there was a gap at all (see Waal 2020). Thus, in Classical Greece, as in the Iron Age Levant, reading aloud was the norm, particularly works with any literary pretensions (Knox & Easterling 1985, 14; Thomas 1992, 13; Knox 1968, 435).² It is reasonable to expect that public documents, such as laws or imprecations, would have been read aloud: even if literacy rates were higher in Classical Greece than they had been in earlier times, there are likely to have been a considerable number of people who relied on others for access to the documents (see for Athens, Thomas 2009, 24).

17.3. Prosodic word level punctuation is a function of the oral performance of texts

The long-lasting significance of orality in all three of the contexts we have considered provides important background for the central thesis of the present monograph, namely, that word division in Ugaritic, Phoenician, Hebrew/Moabite and Ancient Greek demarcates prosodic rather than morphosyntactic (or any other) units. It remains to explore exactly how this might have played out in individual acts of reading and writing.

17.3.1. From spoken to written: the role of dictation

Reconstructing the context of the creation of individual inscriptions is of course fraught with difficulty. However, at least two scholars (Devine & Stephens 1994, 390; Wachter 1999, 379–380) have suggested that word-level punctuation in Ancient Greek inscriptions corresponds to the intonational units of dictation (Devine & Stephens 1994, 390; Wachter 1999, 380).

Such an explanation is supported by evidence from elision. At §13.3.2, I provided evidence that in the *Teiae Dirae* elision does not occur across a prosodic word boundary. This was taken as evidence that the domain of elision was the prosodic word. However, it has been argued that elision also occurs at the level of the prosodic phrase, and that it is only limited to the domain of the prosodic word in careful or deliberate speech (Devine & Stephens 1994, 264). In fluent speech it may even have occurred at the level of the intonational phrase (Devine & Stephens 1994, 264). That elision does not occur at the level of the prosodic phrase in (424) suggests that the domains marked

² It should be said, however, that silent reading was by no means unknown (Knox 1968, 435).

out by word division correspond to those of careful or deliberate speech, such as one might expect in a dictation scenario.

However, this does not explain why such pauses should be indicated in the text: from the perspective of the writer or inscriber it would surely have sufficed simply for the person dictating to wait until the writer/inscriber had caught up. For this, it is necessary to invoke readers.

17.3.2. Aids to the reader

If one of the primary purposes of writing down a text is to then read the text out loud, it will likely be of considerable benefit to the reader to break that text up into pronounceable units.³

The importance of punctuation for readers has been highlighted in recent studies of Ancient Greek writing in relation to word division on Crete. Gagarin & Perlman (2016, 54) link the fact that inscriptions on Crete do not show word division after about 500 BCE to the increasing literacy on the island, suggesting that ‘as readers became more proficient, this aid was no longer needed’ (cf. Gagarin & Perlman 2016, 51). In a similar vein, Steele (2020, 139) has pointed out that the standardisation of writing direction, paragraphing and word division on Crete can all be seen as developments that contribute to making a text more accessible to the readership, particularly in legal documents, where there may have been a need for a wider audience to have help navigating longer texts (Steele 2020, 148). Consistent with this suggestion is the fact that word division is much more common on Crete than elsewhere in the Greek-speaking world, where there are no inscriptions comparable to the Cretan laws (Steele 2020, 148). The fact that word division corresponds to prosodic words rather than morphosyntactic words in these texts, as we have seen, helps us understand the context of word division further: if one of the primary purposes of a written text is for it to have been read aloud, the chunking of that text into pronounceable units seems to be a self-evidently helpful thing to do.

The punctuating of public inscriptions with a view to facilitating the public reading of these documents has parallels in the Levant at a similar period. Lehmann (2005, 91) points out that the Yehawmilk stele (KAI 10, see Chapter 4) ‘is not a written text alone, but that it is a written *public display* text’ (original emphasis), placed as it was at the entrance of the temple of the Lady of Byblos. Since most of those who viewed the inscription likely could not read it, their access to the text was through oral recitation of the inscription, that is, through an oral tradition (Lehmann 2005, 91). The use of spaces would have been important for such oral recitation: it would have shown the reciter how to break the text up into prosodic phrases.⁴

³ Cf. Wachter (1999, 367), who is rather circumspect in his assessment of the function of punctuation. Wachter does note, however, that texts marked up in this way are much easier to read than those which are not.

⁴ Cf. Lehmann’s statement (Lehmann 2005, 92): ‘Our detection of spaces and their distribution is indeed

17.3.3. (In)consistency

It remains to explain, why, in Hebrew (and Moabite) orthography, it is the *minimal* prosodic word that is demarcated, and not *actual* prosodic words in context, the unit of demarcation in Ugaritic and Phoenician texts. We have seen throughout this study that the investigation of the semantics of word division in both Northwest Semitic and Greek material has been beset by the problem of (morphosyntactic) inconsistency. This inconsistency is particularly a feature of the Ugaritic and Phoenician documents we have seen, but may also be found in Greek texts. We have argued that word division is in fact not inconsistent, for the most part, but accurately reflects prosodic chunking, at the levels of either the prosodic word or prosodic phrase. Nevertheless, the fact remains that, as we see both from the cantillation tradition of Tiberian Hebrew and from cross-linguistic study of prosodic phrasing (Devine & Stephens 1994, 225–226), the process of parsing a syntactic unit into prosodic words can yield different results in different contexts. The punctuation of a given text is therefore highly specific to that text.

It is conceivable that, as the practices of recitation changed subtly from generation to generation, graphematic demarcation into actual prosodic words in context in the inscriptions of a previous eras was felt to be rather inconsistent by the writers of the inscriptions themselves. Writers may therefore have sought a more systematic approach to their art, one that could be generalised, in principle at least, to all inscriptions. Yet in a culture where the oral performance of written text is still important, it could be thought desirable to retain the connection between the graphematic and the prosodic: the demarcation of minimal prosodic words fulfils both goals of facilitating recitation and ensuring graphematic consistency.

Demarcation into minimal prosodic words appears to have become standard practice among the writers of Hebrew, and, to the extent that we have evidence, Moabite documents. Word division in Phoenician texts outside of Byblos is, however, very rare. The writers of these texts appear to have achieved the goal of graphematic consistency by dispensing with word division altogether. The same may be said for the Greek setting, where word division is mostly found in texts of the 6th and 5th century BCE, but thereafter are less easy to find (Morpurgo Davies 1987, 270). In the Greek context, at least in official documents, punctuated texts were displaced by texts written in the ‘Stoichedon’ format, whereby letters are arranged in a grid (Wachter 2010, 54). The motivation in this case appears to have been to reduce the chances of forgery (Wachter 2010, 54).

17.3.4. Demarcating morphosyntactic words

At Ugarit we find a major exception to the generalisation that graphematic words correspond to prosodic units: texts written in the Ugaritic ‘Minority’ orthography separate words at the level of morphosyntax rather than prosody (§9.4). What might

an account of such “oral poetry” and can be read as a kind of vocal score for the public performance’.

this fact imply for the intended use of these documents? In particular, the fact that these texts tend to be administrative in nature, and that literary texts appear not to have been written with this word division orthography, opens up the possibility that these texts were not, primarily at least, intended to be read aloud.

We have already seen that silent reading was not common in the ancient world, and it is likely that most people who came into contact with the written word will have done so through oral performance. However, at least in the Greek context, claims that silent reading was unknown in the ancient world have been shown to be overstated (Knox 1968). In the Ugaritic context, it seems *a priori* plausible, at least, that administrative documents intended for 'internal' use might have been written down with the primary intention of being read silently by administrators, rather than publicly proclaimed.

Of course, such a proposal would also need to account for the fact that a number of letters are also written in this orthography. The use of this orthography might be taken to imply that these documents too were not primarily intended to be read aloud. Such a situation might be understood to pertain where the letters in question were translations or transliterations out of another medium, kept for administrative purposes. I leave the pursuing of this question to further research.

17.3.5. Conclusion: the oral context of word division

The prosodic separation of words in both ancient Northwest Semitic and Greek is consistent with what we know of the importance of orality in these cultures. While some of the academic discourse of the second half of the 20th century has sought to emphasise the revolutionary effect of alphabetic writing on literacy, especially in leading to the demise of orality, the evidence from word division practices is consistent with a continuation of the importance of oral performance alongside literacy until well into the 1st millennium BCE.

As discussed in the previous section, the one exception to this, namely, the 'Minority' orthography of Ugaritic, implies a writing practice that was less concerned with the connection to oral performance. It is surely not accidental that the texts that tend to be written in this orthography are primarily administrative in nature. By contrast, texts directly linked to oral performance, notably, the literary epics, regularly separate prosodic words, whilst private letters, which are likely to have been written down in order to be read aloud, also do so in many cases.

Bibliography

- Abritta, A. 2015. 'On the role of accent in ancient Greek poetry. Pitch patterns in the Homeric hexameter.' *Quaderni urbinati di cultura classica* 111(3): 11–27.
- Abritta, A. 2018. 'On the role of accent in ancient Greek poetry.' *Mnemosyne* 71(4): 539–554.
- Agbayani, B. and Golston, C. 2010a. 'Phonological movement in classical Greek.' *Language* 86(1): 133–167.
- Agbayani, B. and Golston, C. 2010b. 'Second-position is first-position: Wackernagel's law and the role of clausal conjunction.' *Indogermanische Forschungen* 115: 1–21.
- Aguilar-Guevara, A., Pozas Loyo, J. and Vázquez-Rojas Maldonado, V. 2019. 'Definiteness across languages: an overview.' In Aguilar-Guevara, A., Pozas Loyo, J. and Vázquez-Rojas Maldonado, V. (eds), *Definiteness across languages*, iii–xx. Berlin, Language Science Press.
- Ahl, S.W. 1973. 'Epistolary texts from Ugarit: Structural and lexical correspondences in epistles in Akkadian and Ugaritic.' Brandeis University doctoral dissertation.
- Ahrens, H.L. 1891[1845]. 'De crasi et aphaeresi.' In Haeberlin, C. (ed.), *Kleine Schriften*, 52. Hannover, Hahn.
- Albright, W.F. 1966. *The Proto-Sinaitic inscriptions and their decipherment* (Harvard theological studies 22). Cambridge, MA, Harvard University Press.
- Allen, W.S. 1966. 'Prosody and prosodies in Greek.' *Transactions of the Philological Society* 65(1): 107–149.
- Allen, W.S. 1973. *Accent and rhythm: Prosodic features of Latin and Greek. A study in theory and reconstruction* (Cambridge studies in linguistics 12). Cambridge, Cambridge University Press.
- Amiran, R. and Eitan, A. 1965. 'A Canaanite-Hyksos city at Tell Nagila.' *Archaeology* 18(2): 113–123.
- Andersen, F.I. 1966. 'Moabite syntax.' *Orientalia* 35(2): 81–120.
- Anderson, S.R. 2005. *Aspects of the theory of clitics* (Oxford Studies in Theoretical Linguistics 11). New York, Oxford University Press.
- Aronoff, M. 1985. 'Orthography and linguistic theory.' *Language* 61(1): 28–72.
- Aurora, F. 2015. 'DAMOS (Database of Mycenaean at Oslo). Annotating a fragmentarily attested language.' Current Work in corpus linguistics: Working with traditionally-conceived corpora and beyond. Selected papers from the 7th International Conference on Corpus Linguistics (CILC2015), 21–31.
- Barker, A. (ed.) 1989. *Greek musical writings: Harmonic and acoustic theory*. Vol. 2. Cambridge, Cambridge University Press.
- Barrett, W.S. (ed.) 2001[1964]. *Euripides Hippolytos*. Oxford, Clarendon.
- Beck, D. 1999. 'Words and prosodic phrasing in Lushootseed narrative.' In Hall, T.A. and Kleinhenz, U. (eds), *Studies on the phonological word*, 23–46. Amsterdam, John Benjamins.
- Beeston, A.F.L. 1984. *Sabaic Grammar* (Journal of Semitic Studies Monograph Series 6). Manchester, Journal of Semitic Studies.
- Berlejung, A. 2010. 'Das Ostrakon aus Qubur al-Walaydah.' *Die Welt des Orients* 40(2): 273–284.
- Beyer, K. 2012. 'The languages of Transjordan.' In Gzella, H. (ed.), *Languages from the world of the Bible*, 111–127. Boston, De Gruyter.
- Bloomfield, L. 1933. *Language*. New York, Henry Holt.
- Bloomfield, L. 1935. *Language*. Delhi, Motilal Banarsidass.
- Blumenfeld, L. 2011. 'Coercion and minimality.' *The Linguistic Review* (2): 207–240.

- Boeckh, A. (ed.) 1828. *Corpus Inscriptionum Graecarum; auctoritate et impensis Academiae Litterarum Regiae Borussicae*. Berlin, Ex Officina Academica.
- Bordreuil, P. and Pardee, D. 2009. *A manual of Ugaritic*. Winona Lake, Ind., Eisenbrauns.
- Botwinik-Rotem, I. 2004. 'The category P: Features, projections, interpretation.' Tel-Aviv University doctoral dissertation.
- Botwinik-Rotem, I. and Terzi, A. 2008. 'Greek and Hebrew locative prepositional phrases: A unified case-driven account.' *Lingua* 118(3): 399–424.
- Boyes, P.J. 2021. *Script and Society: The social context of writing practices in late bronze age Ugarit*. Oxford, Oxbow Books.
- Breuer, M. 1982. *Ta'ame hamiqra bekhaf-'alef sefarim uvesifre 'emet* [The Biblical Accents in the Twenty-One Books and in the Three Books]. Jerusalem, Mikhala.
- Brown, B. 2008. 'The Kilamuwa relief: Ethnicity, class and power in Iron Age North Syria.' In Córdoba, J.M., Molist, M., Pérez, M.C., Rubio, I. and Martínez, S. (eds), *Proceedings of the 5th international congress on the archaeology of the ancient Near East, Madrid, April 3–8 2006*. Vol. 1, 339–355. Madrid, Universidad Autónoma de Madrid.
- Buck, C.D. 1998[1955]. *The Greek dialects*. London, Bristol Classical Press.
- Chadwick, J. 1992. 'The Thessalian accent.' *Glotta* 70(1/2): 2–14.
- Christophe, A., Gout, A., Peperkamp, S. and Morgan, J. 2003. 'Discovering words in the continuous speech stream.' *Journal of Phonetics* 31(3–4): 585–598.
- Clackson, J. 2007. *Indo-European linguistics: An introduction*. Cambridge, Cambridge University Press.
- Collins, T. 1971. 'The Kilamuwa inscription – a Phoenician poem.' *Die Welt des Orients* 6(2): 183–188.
- Cook, V. 2004. *The English writing system*. London, Taylor & Francis.
- Coulmas, F. 1996. *The Blackwell encyclopedia of writing systems*. Oxford, Blackwell.
- Cross, F.M. 1954. 'The evolution of the Proto-Canaanite alphabet.' *Bulletin of the American Schools of Oriental Research* 134(134): 15–24.
- Cross, F.M. 1967. 'The origin and early evolution of the alphabet.' *Eretz-Israel: Archaeological, Historical and Geographical Studies* 8: 8*–24*.
- Cross, F.M. 1980. 'Newly found inscriptions in Old Canaanite and Early Phoenician scripts.' *Bulletin of the American Schools of Oriental Research* 238: 1–20.
- Cross, F.M. 1984. 'An Old Canaanite inscription recently found at Lachish.' *Tel Aviv* 11(1): 71–76.
- Cunchillos, J.-L., Vita, J.-P. and Zamora, J.-Á. (eds) 2003. *Ugaritic data bank: The texts*. Madrid, Laboratorio de Hermeneumática.
- Curtis, A. 2016. 'Ilimilkū of Ugarit: Copyist or creator?' In Römer, T. and Davies, P.R. (eds), *Writing the Bible: Scribes, scribalism and script*. London, Routledge.
- Danon, G. 1996. 'The syntax of determiners in Hebrew.' Tel Aviv University doctoral dissertation.
- DeCaen, V. 2004. 'The pausal phrase in Tiberian Aramaic and the reflexes of *I.' *Journal of Semitic Studies* 49(2): 215–224.
- DeCaen, V. and Drescher, B.E. 2020. 'Pausal forms and prosodic structure in Tiberian Hebrew.' In Hornkohl, A.D. and Khan, G. (eds), *Studies in Semitic vocalisation and reading traditions*. Cambridge, University of Cambridge & Open Book.
- del Olmo Lete, G. 2008. 'The postpositions in Semitic.' *Aula Orientalis* 1(26): 25–59.
- del Olmo Lete, G. 2011. 'Grammaticalization, lexicalization, and semantic universals.' *Aula Orientalis* 29: 27–41.
- del Olmo Lete, G. and Sanmartín, J. 2015. *A dictionary of the Ugaritic language in the alphabetic tradition*. 3rd edn. Leiden & Boston, Brill.
- Devine, A.M. and Stephens, L.D. 1991. 'Dionysius of Halicarnassus, De compositione verborum XI.' *Transactions of the American Philological Association* 121: 229–286.
- Devine, A.M. and Stephens, L.D. 1994. *The prosody of Greek speech*. Oxford, Oxford University Press.

- Dietrich, M. and Loretz, O. 2009. 'Die keilalphabetischen Briefe aus Ugarit (I).' *Ugarit-Forschungen* (41): 109–164.
- Dietrich, M., Loretz, O. and Sanmartín, J. (eds) 1976. *Die Keilalphabetischen Texte aus Ugarit*. 1st edn. Neukirchen-Vluyn, Ugarit-Verlag.
- Dietrich, M., Loretz, O. and Sanmartín, J. (eds) 2013. *Die keilalphabetischen Texte aus Ugarit, Ras Ibn Hani und anderen Orten*. 3rd edn (Alter Orient und Altes Testament 360/1). Münster.
- Diggle, J. (ed.) 1981. *Euripidis fabulae*. Vol. 2. Oxford, Clarendon.
- Diggle, J. (ed.) 1984. *Euripidis fabulae*. Vol. 1. Oxford, Clarendon.
- Diggle, J. (ed.) 1994. *Euripidis fabulae*. Vol. 3. Oxford, Clarendon.
- Donner, H. and Röllig, W. 1968. *Kanaanäische und aramäische Inschriften*. 2nd edn. Vol. 2. Wiesbaden, Harrassowitz.
- Donner, H. and Röllig, W. 2002. *Kanaanäische und aramäische Inschriften*. 5th edn. Vol. 1. Wiesbaden, Harrassowitz Verlag.
- Dover, K.J. 1960. *Greek word order*. Cambridge, Cambridge University Press.
- Dresher, B.E. 1994. 'The prosodic basis of the Tiberian Hebrew system of accents.' *Language* 70(1): 1–52.
- Dresher, B.E. 2009. 'The word in Tiberian Hebrew.' In Hanson, K. and Inkelas, S. (eds), *The nature of the word: Studies in honor of Paul Kiparsky*, 95–111. Cambridge, MA, MIT.
- Dresher, B.E. 2013. 'Biblical accents: Prosody.' In Khan, G., Bolozky, S., Fassberg, S., Rendsburg, G.A., Rubin, A.D., Schwarzwald, O.R. and Zewi, T. (eds), *Encyclopedia of Hebrew language and linguistics*. Leiden, Brill.
- Egetmeyer, M. 2010. *Le dialecte grec ancien de Chypre*. Berlin, de Gruyter.
- Ellison, J.L. 2002. 'A paleographic study of the alphabetic cuneiform texts from Ras Shamra/Ugarit.' Harvard University doctoral dissertation.
- Elvira Astoreca, N. 2020. 'ΕΛΛΗΝΙΚΑ ΓΡΑΜΜΑΤΑ: A linguistic analysis of the early Greek alphabets.' University of Cambridge doctoral dissertation.
- ERV: *The Holy Bible Containing the Old and New Testaments Translated Out of the Original Tongues: Being the Version Set Forth A.D. 1611 Compared with the Most Ancient Authorities and Revised*. 1885. Oxford, Oxford University Press; BibleWorks, v.9.
- Evertz, M. 2016. 'Minimal graphematic words in English and German.' *Written Language & Literacy* 19(2): 189–211.
- Evertz, M. 2018. *Visual prosody: The graphematic foot in English and German* (Linguistische Arbeiten 570). Berlin & Boston, De Gruyter.
- Evertz, M. and Primus, B. 2013. 'The graphematic foot in English and German.' *Writing Systems Research* 5(1): 1–23.
- Faraone, C.A. 1996. 'Taking the 'Nestor's Cup Inscription' seriously.' *Classical Antiquity* (1): 77–112.
- Fassberg, S.E. 2013. 'Sandhi: Modern Hebrew.' In Khan, G., Bolozky, S., Fassberg, S., Rendsburg, G.A., Rubin, A.D., Schwarzwald, O.R. and Zewi, T. (eds), *Encyclopedia of Hebrew language and linguistics*. Leiden, Brill.
- Finglass, P.J. (ed.) 2011. *Sophocles Ajax* (Cambridge Classical Texts and Commentaries 48). Cambridge & New York, Cambridge University Press.
- Fortson, B.W. 2008. *Language and rhythm in Plautus: Synchronic and diachronic studies*. Berlin, De Gruyter.
- Fortson, B.W. 2010. *Indo-European language and culture: An Introduction*. 2nd edn. Wiley-Blackwell.
- Freedman, D.N. 1964. 'A second Mesha inscription.' *Bulletin of the American Schools of Oriental Research* 175: 50–51.
- Fraenkel, M. (ed.) 1902. *Inscriptiones Graecae IV: Aeginae Pityonesi Cecyphaliae Argolidis*. Fasc. 1. Berlin, Georg Reimer.

- Friedrich, J., Röllig, W. and Amadasi Guzzo, M.G. 1999. *Phönizisch-Punische Grammatik*. 3rd edn (Analecta Orientalia 55). Rome, Pontificio Instituto Biblico.
- Gagarin, M. and Perlman, P. (eds) 2016. *The laws of Ancient Crete: c.650-400 BCE*. Oxford, Oxford University Press.
- George, C.H. 2014. *Expressions of time in ancient Greek*. Cambridge, Cambridge University Press.
- Givón, T. 1991. 'The evolution of dependent clause morpho-syntax in Biblical Hebrew.' *Approaches to grammaticalization: Focus on types of grammatical markers*. Vol. 2, 257–310. Amsterdam, John Benjamins.
- Gnanadesikan, A.E. 2017. 'Towards a typology of phonemic scripts.' *Writing Systems Research* 9(1): 14–35.
- Goldstein, D.M. 2010. 'Wackernagel's law in fifth-century Greek.' UC Berkeley doctoral dissertation.
- Goldstein, D.M. 2013. 'Pitch.' In Giannakis, G. (ed.), *Encyclopedia of ancient Greek language and linguistics*. Leiden, Brill.
- Goldstein, D.M. 2016. *Classical Greek Syntax: Wackernagel's law in Herodotus*. Leiden & Boston, Brill.
- Goldstein, D.M. and Haug, D.T.T. 2016. 'Second-position clitics and the syntax-phonology interface: The case of ancient Greek.' In Arnold, D., Butt, M., Crysmann, B., Holloway King, T. and Müller, S. (eds), *Proceedings of the joint 2016 conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland*. Stanford, CA, CSLI.
- Golston, C. 1990. 'Floating H (and L*) tones in ancient Greek.' In Myers, J. and Pérez, P.E. (eds), *Arizona phonology conference*. Tucson, Department of Linguistics, University of Arizona.
- Golston, C. 1991. 'Minimal word, minimal affix.' *Proceedings of the North East Linguistic Society* (21): 95–109.
- Golston, C. 1995. 'Syntax outranks phonology: Evidence from ancient Greek.' *Phonology* (3): 343.
- Golston, C. 2013. 'Minima.' In Giannakis, G. (ed.), *Encyclopedia of ancient Greek language and linguistics*. Leiden, Brill.
- Goodwin, W.W. 1894. *A Greek grammar*. Basingstoke & London, Macmillan.
- Gordon, C.H. 1965. *Ugaritic textbook: Grammar, texts in transliteration, cuneiform selections, glossary, indices* (Analecta Orientalia 38). Rome, Pontifical Biblical Institute.
- Gordon, C.H. 1998. *Ugaritic textbook. Revised reprint* (Analecta Orientalia 38). Rome, Pontifical Biblical Institute.
- Greene, N.E. 2017. 'The Qubur al-Walaydah bowl.' *Bulletin of the American Schools of Oriental Research* 377(1): 39–47.
- Greenstein, E.L. 2014. 'Review of Dennis Pardee, "The Ugaritic texts and the origins of West Semitic literary composition".' *Bulletin of the American Schools of Oriental Research* 371: 215–218.
- Gzella, H. 2007a. 'Parallelismus und Asymmetrie in Ugaritischen Texten.' In Wagner, A. (ed.), *Parallelismus membrorum*, 133–146. Göttingen, Vandenhoeck & Ruprecht.
- Gzella, H. 2007b. 'Some penciled notes on Ugaritic lexicography.' *Bibliotheca Orientalis* 60(5–6): 527–567.
- Hackstein, O. 2010. 'The Greek of epic.' In Bakker, E.J. (ed.), *A companion to the ancient Greek language*, 401–423. Hoboken, Wiley.
- Hall, T.A. 1999. 'The phonological word: a review.' In Hall, T.A. and Kleinhenz, U. (eds), *Studies on the phonological word*, 1–22. Amsterdam, John Benjamins.
- Hamilton, G.J. 2002. 'W. F. Albright and early alphabetic epigraphy.' *Near Eastern Archaeology* 65(1): 35–42.
- Haring, B. 2020. 'Ancient Egypt and the earliest known stages of alphabetic writing.' In Boyes, P.J. and Steele, P.M. (eds), *Understanding relations between scripts II: Early alphabets*, 53–67. Oxford, Oxbow Books.
- Harris, Z.S. 1936. *A grammar of the Phoenician language*. New Haven, Connecticut, American Oriental Society.

- Haspelmath, M. 2011. 'The indeterminacy of word segmentation and the nature of morphology and syntax.' *Folia Linguistica* 45(1): 31–80.
- Hawley, R., Pardee, D. and Roche-Hawley, C. 2016. 'The scribal culture of Ugarit.' *Journal of Ancient Near Eastern History* 2(2): 229–267.
- Hayes, B. 1989. 'The prosodic hierarchy in meter.' In Kiparsky, P. and Youmans, G. (eds), *Phonetics and phonology*. Vol. 1, 201–260. San Diego & London, Academic.
- Herdner, A. 1963. *Corpus des tablettes en cunéiformes alphabétiques découvertes à Ras Shamra-Ugarit de 1929 à 1939*. Paris, Imprimerie Nationale.
- Hestrin, R. 1987. 'The Lachish Ewer and the 'Ašerah.' *Israel Exploration Journal* 37(4): 212–223.
- Hoekstra, A. 1978. 'Metrical lengthening and epic diction.' *Mnemosyne* (1): 1–26.
- Hoftijzer, J. 1982. 'Quodlibet Ugariticum.' In van Driel, G. (ed.), *Zikir šumim: Assyriological studies presented to F.R. Kraus on the occasion of his seventieth birthday*, 121–127. Leiden, Brill.
- Horwitz, W.J. 1971. 'Graphematic representation of word boundary: The small vertical wedge in Ugaritic.' Yale University doctoral dissertation.
- Huehnergard, J. 2012. *An introduction to Ugaritic*. Peabody, MA, Hendrickson.
- Hummel, H.D. 1957. 'Enclitic mem in early Northwest Semitic, especially Hebrew.' *Journal of Biblical literature* 76(2): 85–107.
- Inkelas, S. 1989. 'Prosodic constituency in the lexicon.' Stanford University doctoral dissertation.
- Irvine, J.A.D. 1997. 'Euripides' Ion l. 1 and Pap. Herc. 1088 2 a reconsidered.' *Zeitschrift für Papyrologie und Epigraphik* 117: 1–8.
- Jameson, M.H. 1974. 'A treasury of Athena in the Argolid (IG IV, 559).' In Bradeen D.W. and McGregor, M.F. (eds), *ΦΟΡΟΣ: Tribute to Benjamin Dean Meritt*, 67–75. Locust Valley, NY, J.J. Augustin.
- Janis, N. 1987. 'A grammar of the Biblical accents.' Harvard University doctoral dissertation.
- Kaiser, R. 1887. *De inscriptionum graecarum interpunctione*. Berlin, Gustav Schade.
- Kantor, B.P. 2017. 'The second column (Secunda) of Origen's Hexapla in light of Greek pronunciation.' The University of Texas at Austin doctoral dissertation.
- Kautsch, E. 1910. *Gesenius' Hebrew grammar*. 2nd edn. Oxford, Clarendon.
- Khan, G. 1987. 'Vowel length and syllable structure in the Tiberian tradition of Biblical Hebrew.' *Journal of Semitic Studies* 32(1): 23.
- Khan, G. 2001. 'The Hebrew Bible.' In Rogerson, J. (ed.), *The Oxford illustrated history of the Bible*, 60–109. Oxford, Oxford University Press.
- Khan, G. 2013. 'Biblical Hebrew.' In Khan, G., Bolozky, S., Fassberg, S., Rendsburg, G.A., Rubin, A.D., Schwarzwald, O.R. and Zewi, T. (eds), *Encyclopedia of Hebrew language and linguistics*. Leiden, Brill.
- Khan, G. 2020. *The Tiberian pronunciation tradition of Biblical Hebrew* (Semitic Languages and Cultures 1). Cambridge, University of Cambridge & Open Book.
- KJV: *Authorized Version, 1769 Blayney Edition*. 1769. Oxford: Cambridge University Press; BibleWorks, v.9.
- Klavans, J.L. 2019[1995]. *On clitics and cliticization: The interaction of morphology, phonology, and syntax*. London, Routledge.
- Knapp, A.B. 1983. 'An Alashiyan merchant at Ugarit.' *Tel Aviv* 10(1): 38–45.
- Knox, B.M.W. 1968. 'Silent reading in antiquity.' *Greek, Roman and Byzantine studies* 9(4): 421–435.
- Knox, B.M.W. and Easterling, P.E. 1985. 'Books and readers in the Greek world.' In Easteling, P.E. and Knox, B.M.W. (eds), *The Cambridge history of classical literature: Greek literature*. Vol. 1, 1–41. Cambridge, Cambridge University Press.
- Koller, A. 2021. The alphabetic revolution, writing systems, and scribal training in Ancient Israel. In Hornkohl, A. and Khan G. (eds), *New perspectives in Biblical and Rabbinic Hebrew*, 1–27. Cambridge, University of Cambridge & Open Book.
- Krahmalkov, C.R. 2001. *A Phoenician-Punic grammar*. Leiden, Brill.

- Krahn, A.E. 2014. 'A new paradigm for punctuation.' University of Wisconsin-Milwaukee doctoral dissertation.
- Kuryłowicz, J. 1959. 'The accentuation of the verb in Indo-European and in Hebrew.' *WORD* 15(1): 123-129.
- Larfeld, W. 1914. *Griechische Epigraphik*. 3rd edn. München, C.H. Beck.
- Lehmann, R.G. 2005. 'Space-syntax and metre in the inscription of Yehawmilk, King of Byblos.' In Al-Ghul, O. (ed.), *Proceedings of the Yarmouk second annual colloquium on epigraphy and ancient writings, 7th-9th October, 2003*. Irbid, Jordan, Department of Epigraphy, Faculty of Archaeology & Anthropology, Yarmouk University.
- Lehmann, R.G. 2016. 'On word division and grammar as it is applied to the second Nerab inscription.' *Eretz-Israel: Archaeological, Historical and Geographical Studies* 32: 37-47.
- Lehmann, R.G. 2020. 'Much ado about an implement! – the Phoenicianising of early alphabetic.' In Boyes, P.J. and Steele, P.M. (eds), *Understanding relations between scripts II: Early alphabets*, 69-90. Oxford, Oxbow Books.
- Leibovitch, J. 1965. 'Le tesson de Tell Nagila.' *Le Museon* 78: 229-230.
- Lewis, D. (ed.) 1981. *Inscriptiones Graecae I: Inscriptiones Atticae Euclidis anno anteriores*. 3rd edn. Fasc. 1. *Decreta et tabulae magistratum* (nos. 1-500). Berlin.
- Lewis, D. and Jeffery, L. (eds.) 1994. *Inscriptiones Graecae I: Inscriptiones Atticae Euclidis anno anteriores*. 3rd edn. Fasc. 2. *Dedications. Catalogi. Termini. Tituli sepulcrales. Varia. Tituli Attici extra Atticam reperti. Addenda* (nos. 501-1517). Berlin. <http://telota.bbaw.de/ig/digitale-edition/inschrift/IG%201%C2%B3%201084>. Accessed 22nd October 2021.
- Lidzbarski, M. 1898. *Handbuch der nordsemitischen Epigraphik nebst ausgewählten Inschriften*. Weimar, Felber. <https://archive.org/details/32882013305316-handbuchdernord>. Accessed 19th October 2021.
- Linder, E. 1970. 'The maritime texts of Ugarit: a study in late bronze age shipping.' Brandeis University doctoral dissertation.
- Lipiński, E. 1971. 'Etymological and exegetical notes on the Meša^c inscription.' *Orientalia* 40(3): 325-340.
- Luraghi, S. 2013. 'Clitics.' In Giannakis, G. (ed.), *Encyclopedia of ancient Greek language and linguistics*. Leiden, Brill.
- Mabie, F.J. 2004. 'Ancient Near Eastern scribes and the mark(s) they left: A catalog and analysis of scribal auxiliary marks in the Amarna corpus and in the cuneiform alphabetic texts of Ugarit and Ras Ibn Hani.' UCLA doctoral dissertation.
- Macdonell, A.A. 1927. *A Sanskrit grammar for students*. 3rd edn. Oxford, Oxford University Press.
- Margalit, B. 1975. 'Studia Ugaritica, I.' *Ugarit Forschungen* (7): 289-313.
- Matthews, P.H. 1991. *Morphology*. 2nd edn. Cambridge & New York, Cambridge University Press.
- Mead, A. 1992. 'Review of the development of Multidimensional Scaling methods.' *The Statistician* 41(1): 27-39.
- Meiggs, R. and Lewis, D. (eds) 1969. *A selection of Greek historical inscriptions to the end of the fifth century B.C.* Oxford, Clarendon.
- Millard, A.R. 1970. "'Scriptio continua" in early Hebrew.' *Journal of Semitic Studies* 15(1): 2-15.
- Millard, A.R. 1986. 'The infancy of the alphabet.' *World Archaeology* 17(3): 390-398.
- Millard, A.R. 2012a. 'Scripts and their uses in the 12th-10th centuries.' In Galil, G., Gilboa, A., Maeir, A.M. and Kahn, D. (eds), *The Ancient Near East in the 12th-10th centuries BCE. Culture and history. Proceedings of the international conference held at the University of Haifa, 2-5 May, 2010*, 405-412. Münster.
- Millard, A.R. 2012b. 'The alphabet.' In Gzella, H. (ed.), *Languages from the world of the Bible*, 14-27. Boston, De Gruyter.

- Millard, A.R. and Bordreuil, P. 1982. 'A statue from Syria with Assyrian and Aramaic inscriptions.' *The Biblical Archaeologist* 45(3): 135–141.
- Miller, P.H. 1992a. *Clitics and constituents in Phrase Structure Grammar*. <https://books.google.co.uk/books?id=-UFUankRck0C&printsec=frontcover>. Accessed 2021.
- Miller, P.H. 1992b. 'Postlexical cliticization vs. affixation: Coordination criteria.' *Papers from the 28th regional meeting of the Chicago Linguistic Society*, 382–396. Chicago, Chicago Linguistic Society.
- Morpurgo Davies, A. 1987. 'Folk-linguistics and the Greek word.' In Cardona, G. and Zide, N.H. (eds), *Festschrift for Henry Hoenigswald: On the occasion of his seventieth birthday*, 263–280. Tübingen, Narr.
- Munro, D.B. and Allen, T.W. (eds) 1920. *Homeri opera in five volumes*. Oxford: Oxford University Press. <http://www.perseus.tufts.edu/hopper/>. Accessed 19th October 2021.
- Naveh, J. 1973a. 'Some Semitic epigraphical considerations on the antiquity of the Greek alphabet.' *American Journal of Archaeology* 77(1): 1–8.
- Naveh, J. 1973b. 'Word division in West Semitic writing.' *Israel Exploration Journal* 23(4): 206–208.
- Neef, M. 2015. 'Writing systems as modular objects.' *Open Linguistics* 1(1): 708–721.
- Nespor, M. and Vogel, M. 2007. *Prosodic phonology: With a new foreword* (Studies in generative grammar 28). Berlin, Mouton de Gruyter.
- Noorlander, P. 2012. 'Samʿalian in its Northwest Semitic setting.' *Orientalia* 81(3): 202–238.
- Nunberg, G. 1990. *The linguistics of punctuation*. Stanford, Center for the Study of Language and Information.
- Packard, J.L. 2000. *The morphology of Chinese: A linguistic and cognitive approach*. Cambridge, Cambridge University Press.
- Page, D.L. (ed.) 1972. *Aeschylus septem quae supersunt tragoedias*. Oxford, Clarendon.
- Pardee, D. 1975. 'The preposition in Ugaritic.' *Ugarit-Forschungen* (7): 329–378.
- Pardee, D. 1981. 'Ugaritic and Hebrew metrics.' In Young, G.D. (ed.), *Ugarit in retrospect: Fifty years of Ugarit and Ugaritic*, 113–130. Winona Lake, IN, Eisenbrauns.
- Pardee, D. 1997. 'Ugaritic.' In Hetzron, R. (ed.), *The Semitic languages*, 131–144. New York & London, Routledge.
- Pardee, D. 1998. 'Remarks on J.T.'s "Epigraphische Anmerkungen".' *Aula Orientalis* 16: 85–102.
- Pardee, D. 2003a. 'The Baʿlu myth.' In Hallo, W.W. and Younger, K.L. (eds), *The context of Scripture*. Vol. 1. *Canonical compositions from the Biblical world*, 241–274. Leiden, Brill.
- Pardee, D. 2003b. 'Ugaritic letters.' In Hallo, W.W. and Younger, K.L. (eds), *The context of Scripture*. Vol. 3. *Archival documents from the Biblical world*, 87–116. Leiden, Brill.
- Pardee, D. 2003–2004a. 'Review of Josef Tropper, "Ugaritisch. Kurzgefasste Grammatik".' *Archiv für Orientforschung* (50): 412–417.
- Pardee, D. 2003–2004b. 'Review of Josef Tropper, Ugaritische Grammatik (online version)'. *Archiv für Orientforschung* (50): 1–404. https://orientalistik.univie.ac.at/fileadmin/user_upload/i_orientalistik/Zeitschriften/Rezension_Tropper_AOAT273.pdf. Last accessed 26th October 2021.
- Pardee, D. 2008. 'Ugaritic.' In Woodard, R.D. (ed.), *The ancient languages of Syria-Palestine and Arabia*, 5–35. Cambridge, Cambridge University Press.
- Pardee, D. 2012. *The Ugaritic texts and the origins of West Semitic literary composition*. New York, Oxford University Press.
- Park, S.J. 2020. *The fundamentals of Hebrew accents: Divisions and exegetical roles beyond syntax*. Cambridge, Cambridge University Press.
- Parry, D.W. and Qimron, E. 1999. *The Great Isaiah Scroll (1QIsa^a): A new edition*. Leiden, Brill.
- Pat-El, N. and Wilson-Wright, A. 2015. 'Deir ʿAllā as a Canaanite dialect: A vindication of Hackett.' In Hutton, J.M. and Rubin, A.D. (eds), *Epigraphy, philology, & the Hebrew Bible: Methodological perspectives on philological and comparative study of the Hebrew Bible in honor of Jo Ann Hackett*, 13–23. Atlanta, SBL Press.

- Peek, W. (ed.) 1957. *Attische Grabschriften*. Vol. 2 (Abhandlungen der deutschen Akademie der Wissenschaften zu Berlin. Klasse für Sprachen, Literatur und Kunst 3). Berlin.
- Pico, M. 2019. 'A nascent definiteness marker in Yokot'an Maya.' In Aguilar-Guevara, A., Pozas Loyo, J. and Vázquez-Rojas Maldonado, V. (eds), *Definiteness across languages*, 153–200. Berlin, Language Science Press.
- Pighi, G.B. 1959. *Aristoxeni rhythmica*. Bologna, Patron. <http://stephanus.tlg.uci.edu/>. Accessed July and August 2021.
- Pitard, W.T. 2012. 'Nodding scribe and heavy thumb: The scribal errors in CAT 1.14 III 52-V 8.' *Puzzling Out the Past*. Vol. 55, 135–153.
- Pope, M. 1951. 'Ugaritic enclitic -m.' *Journal of Cuneiform Studies* 5(4): 123–128.
- Price, J.D. 2010. *The syntax of Masoretic accents in the Hebrew Bible*. 2nd edn. http://www.jamesdprice.com/images/21_Syntax_of_Accents_rev_ed..pdf. Accessed 30th April 2021.
- Prince, A. and Smolensky, P. 2002. *Optimality Theory: Constraint interaction in Generative Grammar*. Rutgers University. <https://doi.org/10.7282/T34M92MV>
- Prince, A.S. 1975. 'The phonology and morphology of Tiberian Hebrew.' MIT doctoral dissertation.
- Probert, P. 2003. *A new short guide to the accentuation of ancient Greek*. Bristol, Bristol Classical Press.
- Probert, P. 2006. *Ancient Greek accentuation: Synchronic patterns, frequency effects, and prehistory*. Oxford, Oxford University Press.
- Probert, P. and Dickey, E. 2015. 'The 'ϕηδιέστας' inscription from archaic Argos (SEG 11:314).' *Journal of Hellenic Studies* (135): 110–131.
- Puech, E. 1986. 'The Canaanite inscriptions of Lachish and their religious background.' *Tel Aviv* 13(1): 13–25.
- Ravid, D. 2012. *Spelling Morphology: The psycholinguistics of Hebrew spelling*. New York, Springer.
- Redford, D. 2000. 'Scribe and speaker.' In Zvi, E.B. and Floyd, M.H. (eds), *Writings and speech in Israelite and ancient Near Eastern prophecy*, 145–218. Atlanta, GA, Society of Biblical Literature.
- Rendsburg, G.A. and Schniedewind, W.M. 2010. 'The Siloam Tunnel inscription.' *Israel Exploration Journal* 60(2): 188–203.
- Revell, E.J. 1971. 'The oldest evidence for the Hebrew accent system.' *Bulletin of the John Rylands Library* 54(1): 214–222.
- Revell, E.J. 1976. 'Biblical punctuation and chant in the Second Temple period.' *Journal for the Study of Judaism* 7(2): 181–198.
- Revell, E.J. 2015. *The pausal system: Divisions in the Hebrew Biblical text as marked by vowelings and stress position*. Sheffield, Sheffield Phoenix Press.
- Revell, E.J. 2016. 'Terminal markers in the Masoretic Text.' *Journal of Semitic Studies* 61(1): 67–84.
- Revithiadou, A. 2013. 'Clitic Group.' In Giannakis, G. (ed.), *Encyclopedia of ancient Greek language and linguistics*. Leiden, Brill.
- Richardson, M.E.J. 1973. 'Ugaritic spelling errors.' *Tyndale Bulletin* (24): 3–20.
- Richey, M. 2019. 'Inscriptions.' In Doak, B.R. and López-Ruiz, C. (eds), *The Oxford handbook of the Phoenician and Punic Mediterranean*, 223–240.
- Robertson, A.F. 1994. 'Word dividers, spot markers and clause markers in Old Assyrian, Ugaritic, and Egyptian texts: Sources for understanding the use of the red ink points in the two Akkadian literary texts, *Adapa* and *Ereshkigal*, found in Egypt.' New York University doctoral dissertation.
- Robertson, A.F. 1999. 'Non-word divider Use of the small vertical wedge in *Yarih* and *Nikkal* and in Akkadian text written in alphabetic cuneiform.' In Chazan, R., Hallo, W.W. and Schiffmann, L.H. (eds), *Ki Baruch hu: Ancient Near Eastern, Biblical, and Judaic studies in honor of Baruch A. Levine*, 89–109. Winona Lake, IN, Eisenbrauns.
- Roehl, H. 1882. *Inscriptiones Graecae antiquissimae: praeter Atticas in Attica repertas*. Berlin, Georg Reimer.
- Röllig, W. 1983. 'On the origin of the Phoenicians.' *Berytus* 31: 79–93.

- RSV: *Revised Standard Version of the Bible*. 2nd edn. 1971. New York: Division of Christian Education of the National Council of the Churches of Christ in the United States of America; BibleWorks, v.9.
- Saenger, P.H. 1997. *Space between words: The origin of silent reading*. Stanford, CA, Stanford University Press.
- Sandbank, A., Walden, Z. and Zeiler, E. 1995. *Young children's writing*. Jerusalem, Israeli Ministry of Education.
- Sapir, E. 1921. *Language: An introduction to the study of speech*. Oxford, Oxford University Press.
- Sass, B. 2017. 'The emergence of monumental West Semitic alphabetic writing, with an emphasis on Byblos.' *Semitica* 59: 109–141.
- Sasson, V. 1982. 'The Siloam Tunnel inscription.' *Palestine Exploration Quarterly* 114(2): 111–117.
- Sauzet, P. 1989. 'L'accent du grec ancien et les relations entre structure métrique et représentation autosegmentale.' *Langages* 24(95): 81–113.
- Schniedewind, W.M. 2004. *How the Bible became a book: The textualization of ancient Israel*. New York, Cambridge University Press.
- Schwyzler, E. 1923. *Dialectorum Graecarum exempla epigraphica potiora*. Leipzig, Hirzel.
- Segert, S. 1961. 'Die Sprache der Moabitischen Königsinschrift.' *Archiv Orientalní* 29: 197–267.
- Segert, S. 1984. *A basic grammar of the Ugaritic language with selected texts and glossary*. Berkeley, CA, University of California Press.
- Selkirk, E. 1986. 'On derived domains in sentence phonology.' *Phonology* 3: 371–405.
- Selkirk, E. 1996. 'The prosodic structure of function words.' In Morgan, J.L. and Demuth, K. (eds), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, 187–214. New York, NY, Psychology Press.
- Selkirk, E. 2011. 'The syntax–phonology interface.' In Goldsmith, J., Riggle, J., and Yu, A. (eds), *The Handbook of Phonological Theory*. 2nd edn, 435–484. Oxford, Blackwell.
- Shattuck-Hufnagel, S. and Turk, A. 1996. 'A prosody tutorial for investigators of auditory sentence processing.' *Journal of Psycholinguistic Research* 25(2): 193–247.
- Sihler, A.L. 1995. *New comparative grammar of Greek and Latin*. Oxford, Oxford University Press.
- Sivan, D. 2001. *A grammar of the Ugaritic language*. Atlanta, Society of Biblical Literature.
- Sommerstein, A.H. 2009. *Oresteia: Agamemnon. Libation-Bearers. Eumenides* (Loeb Classical Library 146). Cambridge, MA, Harvard University Press.
- Sproat, R. 2000. *A computational theory of writing systems*. Cambridge, Cambridge University Press.
- Steele, P.M. 2011. *A linguistic history of Cyprus: the non-Greek languages, and their relations with Greek, c.1600–300 BC*. Cambridge, Cambridge University Press.
- Steele, P.M. 2020. 'The development of Greek alphabets: Fluctuations and standardisations.' In Boyes, P.J. and Steele, P.M. (eds), *Understanding relations between scripts II: Early alphabets*, 125–149. Oxford, Oxbow Books.
- Steiner, R.C. 2016. 'Phonemic spelling and *scriptio continua* for sandhi phenomena and glottal stop deletion.' *Journal of Near Eastern Studies* 75(2): 311–334.
- Steriade, D. 1988. 'Greek accent.' *Linguistic Inquiry* 19(2): 271–314.
- Stuart, D. 1976. *Studies in early Hebrew meter*. Missoula, MT, Scholars Press.
- Suchard, B.D. 2019. *The development of the Biblical Hebrew vowels: Including a concise historical morphology* (Studies in Semitic Languages and Linguistics 99). Leiden, Brill.
- Swiggers, P. 1982. 'The Aramaic inscription of Kilamuwa.' *Orientalia* 51(2): 249–253.
- Thomas, R. 1992. *Literacy and orality in ancient Greece*. Cambridge, New York & Melbourne, Cambridge University Press.
- Thomas, R. 2009. 'Writing, reading, public and private 'literacies': Functional literacy and democratic literacy in Greece.' In Johnson, W.A. and Parker, H.N. (eds), *Ancient literacies the culture of reading in Greece and Rome*, 13–45. Oxford, Oxford University Press.

- Thompson, R. 2010. 'Mycenaean Greek.' In Bakker, E.J. (ed.), *A companion to the ancient Greek language*, 189–199. Hoboken, Wiley.
- Threatte, L. 1978. 'Unmetrical spellings in Attic inscriptions.' *California Studies in Classical Antiquity* 10: 169.
- Threatte, L. 2015[1980]. *The grammar of Attic inscriptions*. Vol. 1. *Phonology*. Berlin & Boston, De Gruyter.
- Tov, E. 2015. 'The coincidental textual nature of the collections of ancient Scriptures.' *Textual criticism of the Hebrew Bible, Qumran, Septuagint: Collected essays*. Vol. 3, 20–35 Boston, Brill.
- Tropper, J. 1994a. 'Die enklitische Partikel -y im Ugaritischen.' *Ugarit Forschungen* 26: 473–482.
- Tropper, J. 1994b. 'Zur Grammatik der ugaritischen Omina.' *Ugarit Forschungen* 26: 457–472.
- Tropper, J. 2000. *Ugaritische Grammatik*. 1st edn (Alter Orient und Altes Testament 273). Münster, Ugarit-Verlag.
- Tropper, J. 2001. 'Die Herausbildung des bestimmten Artikels in Semitischen.' *Journal of Semitic Studies* 46(1): 1–31.
- Tropper, J. 2012. *Ugaritische Grammatik*. 2nd edn (Alter Orient und Altes Testament 273). Münster, Ugarit-Verlag.
- Truckenbrodt, H. 2007. 'The syntax–phonology interface.' In de Lacy, P. (ed.), *The Cambridge handbook of phonology*, 435–456. Cambridge, Cambridge University Press.
- Tugendhaft, A. 2018. *Baal and the politics of poetry*. London, Routledge.
- Ulrich, E. 2017. *The Dead Sea scrolls and the developmental composition of the Bible* (Supplements to Vetus Testamentum 169). Leiden & Boston, Brill.
- Vajda, E.J. 2005. 'Distinguishing between referential and grammatical function in morphological typology.' In Frajzyngier, Z., Hodges, A. and Rood, D.S. (eds), *Linguistic diversity and language theories*, 399–422. Amsterdam, John Benjamins.
- van den Branden, A. 1966. 'Le tessou de Tell Nagila.' *Melto* 2(1): 131.
- van der Merwe, C.H. J., Naudé, J.A. and Kroeze, J.H. 2017. *A Biblical Hebrew reference grammar*. 2nd edn. London, T&T Clark.
- Van Effenterre, H. 1946. 'Inscriptions archaïques crétoises.' *Bulletin de correspondance hellénique* 70: 588–606.
- Van Effenterre, H. 1961. 'Pierres inscrites de Dréros.' *Bulletin de correspondance hellénique* 85: 544–568.
- Vis, J. 2013. 'Prosodic word.' In Giannakis, G. (ed.), *Encyclopedia of ancient Greek language and linguistics*. Leiden, Brill. doi: https://doi.org/10.1163/2214-448X_eagll_COM_00000295
- Waal, W. 2018. 'On the "Phoenician letters".' *Aegean Studies* (1): 83–125.
- Waal, W. 2020. 'Mother or sister? Rethinking the origins of the Greek alphabet and its relation to the other 'western' alphabets.' In Boyes, P.J. and Steele, P.M. (eds), *Understanding relations between scripts II: Early alphabets*, 111–124. Oxford, Oxbow Books.
- Wachter, R. 1999. 'Evidence for phrase structure analysis in some Archaic Greek inscriptions.' In Cassio, A.C. (ed.), *Katà diálekton. Atti del III colloquio internazionale di dialettologia greca*, 365–382. Naples.
- Wachter, R. 2010. 'Inscriptions.' In Bakker, E.J. (ed.), *A companion to the ancient Greek language*, 47–61. Hoboken, Wiley.
- Wansbrough, J. 1983. 'Metra Ugaritica.' *Bulletin of the School of Oriental and African Studies* 46(2): 221–234.
- Watkins, C. 1976. 'Observations on the 'Nestor's Cup' inscription.' *Harvard Studies in Classical Philology* (80): 25–40.
- Watson, W.G.E. 1992. 'Final -m in Ugaritic.' *Aula Orientalis* 10: 223–252.
- Weingarten, Rüdiger 2011. 'Comparative graphematics.' *Written Language and Literacy* 14(1): 12–38.
- West, M.L. 1971. *Iambi et elegi Graeci*. Vol. 1. Oxford, Clarendon. <http://stephanus.tlg.uci.edu/> Accessed July and August 2021.
- West, M.L. 1972. *Iambi et elegi Graeci*. Vol. 2. Oxford, Clarendon.

- West, M.L. 1981. 'The singing of Homer and the modes of early Greek music.' *The Journal of Hellenic Studies* 101: 113–129.
- Whitley, J. 2021. 'Why $\mu\epsilon$? Personhood and agency in the earliest Greek inscriptions (800–550 BC).' In Boyes, P., Steele, P. and Elvira Astoreca, N. (eds), *The social and cultural contexts of historical writing practices*, 269–287. Oxford, Oxbow Books.
- Wingo, E.O. 1972. *Latin punctuation in the Classical Age* (Janua Linguarum. Series Practica 133). The Hague & Paris, Mouton.
- Wintner, S. 2000. 'Definiteness in the Hebrew noun phrase.' *Journal of Linguistics* 36(2): 319–363.
- Woodard, R.D. 2020. 'Vowel representation in the archaic Greek and old Aramaic scripts: A comparative orthographic and phonological examination.' In Boyes, P.J. and Steele, P.M. (eds), *Understanding relations between scripts II: Early alphabets*, 91–107. Oxford, Oxbow Books.
- Woodhead, A.G. 1968. 'SEG 23-530. Dreros. De venatione praescriptio, aet. fere eiusdem.' In Chaniotis, A.T.E.N. and Papazarkas, C.S. (eds.), *Supplementum Epigraphicum Graecum*. Leiden, Brill.
- Wyatt, N. 2005. 'Epic in Ugaritic literature.' *A companion to ancient epic*, 246–254.
- Xella, P. 2017. 'Phoenician inscriptions in Palestine.' In Hübner, U. and Niehr, H. (eds), *Sprachen in Palästina im 2. und 1. Jahrtausend v.Chr.*, 153–169. Wiesbaden, Harrassowitz.
- Yeivin, I. 1980. *Introduction to the Tiberian Masorah*. Missoula, Scholars Press.
- Young, G.D. 1950. 'Ugaritic prosody.' *Journal of Near Eastern Studies* 9(3): 124–133.
- Zec, D. and Inkelas, S. 1990. 'Prosodically constrained syntax.' In Inkelas, S. and Zec, D. (eds), *The phonology-syntax connection*, 365–378. Chicago & London, University of Chicago Press.
- Zernecke, A.E. 2013. 'The Lady of the Titles.' *Die Welt des Orients* 43(2): 226–242.
- Zevit, Z. 1980. *Matres lectionis in ancient Hebrew epigraphs* (ASOR Monograph Series 2). Cambridge, American Schools of Oriental Research.
- Zwicky, A.M. 1985. 'Clitics and particles.' *Language* 61(2): 283–305.
- Zwicky, A.M. and Pullum, G.K. 1983. 'Cliticization vs. inflection.' *Language* 59(3): 502–513.

