

Formularity

Our starting point in the investigation of Homer's machinery is *formularity*, which we can broadly define as the poet's reliance on prefabricated linguistic sequences in the composition of his verses. Few introductions to Homer will fail to mention the frequent recurrence of phrases like *long-suffering divine Odysseus* (42x in the poems) or *swift-footed Achilles* (30x),¹ and most (if not all) modern language translations of the epics will try to convey some of this repetitiveness as they render Homer's verses. Any reader of Homer will soon discover that this repetitiveness does not affect short phrases alone: whole clauses recur unchanged, from the atmospherically evocative *When early-born, rosy-fingered Dawn appeared* (22x) to the entertainingly irate "*What words escaped the fence of your teeth?*" (8x). And there are entire scenes, like duels or banquets, which often appear to be composed entirely, or almost entirely, of slight variations of the same handful of expressions.

Over the last century, formularity has acquired the status of perhaps the most notorious feature of Homer's style, and it has played a fundamental role in revealing the oral-traditional background of Homer's art. Formularity is also the feature of Homer's style on which the field is most divided, with scholars variously disagreeing on its definition, its function, and the extent to which it appears in the poems (50 percent? 90 percent?), and whether it can be used to demonstrate the orality of a text.

Because of the complex history of the term, I shall first give an overview of how the concept has evolved within Homeric studies, and the many issues it comprises (for a history of oral-formulaic theory outside of Homeric studies, see now Frog and Lamb 2022). Next, we will turn to

¹ These repeated expressions consisting of a proper name (e.g., Odysseus) and some "epithetic words" (e.g., *long-suffering, divine*) modifying it have been called *noun-epithet formulae* in Homeric scholarship since Parry's (1971: 17) seminal study.

linguistics and cognitive studies in order to find parallels for Homeric formularity in everyday language and cognition, and to establish whether there are any qualitative or quantitative differences between formularity in Homer and formularity in natural languages. Finally, we will tackle the practical questions of how best to describe Homeric formularity, and how to evaluate its meaning and antiquity.

1.1 The History of Homeric Formularity

1.1.1 Parry: Homer's Style as Traditional

Few scholars nowadays would doubt that formularity played a substantial role in the poet's technique. After all, formularity is very *visible* in the *Iliad* and the *Odyssey* as we have them. As Parry explained:

The easiest and best way of showing the place the formula holds in Homeric style will be to point out all of the expressions occurring in a given passage which are found elsewhere in the *Iliad* or the *Odyssey*, in such a way that, as one reads, one may see how the poet has used them to express his thought. (Parry 1971: 301)

Below, I reproduce the first twenty-five lines of *Iliad* 1 as given in Parry's *Homer and Homeric Style* (1971: 301–2), minus the heavy apparatus (the added translation is mine).² For several decades, this type of illustration was the only available evidence of the density of formulas in Homer, and played an important role in shaping the debate on Homeric style. Here, solid underlining identifies expressions that are found, unchanged, elsewhere in the poems (what Parry would call *formulas*). Broken underlining identifies expressions that appear to be slight variations of expressions found elsewhere in the poems (what Parry would call *formulaic expressions*).³

² Unless specified otherwise, the Greek text reproduced in this book reflects the *Thesaurus Linguae Graecae* (<http://stephanus.tlg.uci.edu>). For Homer, this means the editions of Allen (1931) for the *Iliad* and Von der Mühl (1962) for the *Odyssey*.

³ Note that, in theory, both criteria can be true at the same time, and thus an expression might appear to be a slight variation of an already-known formula *and* it might be repeated elsewhere in the poems. While this is not captured in Parry's analysis, it is captured in Lord's analysis of the first fifteen lines of the same passage (Lord 1960: 143), where almost every line is shown with a thorough broken underlining – suggesting that close to everything about the phraseology is traditional, even when expressions are not repeated verbatim.

(1)

Μῆνιν ἄειδε θεὰ Πηληϊάδεω
 Ἀχιλῆος (1)
 οὐλομένην, ἣ μυρὶ Ἀχαιοῖς ἄλγε'
 ἔθηκε,
 πολλὰς δ' ἰφθίμους ψυχὰς Ἀϊδὶ
 προΐαψεν
 ἥρώων, αὐτοὺς δὲ ἐλώρια τεῦχε
 κύνεσσιν
 οἰωνοῖσι τε πᾶσι, Διὸς δ' ἐτελείετο
 βουλή, (5)
 ἐξ οὗ δὴ τὰ πρῶτα διαστήτην
 ἐρίσαντε
 Ἀτρεΐδης τε ἄναξ ἀνδρῶν καὶ δῖος
 Ἀχιλλεύς.
 Τίς τάρ σφωε θεῶν ἑριδι ξυνέηκε
 μάχεσθαι;
 Λητοῦς καὶ Διὸς υἱός· ὁ γὰρ βασιλῆϊ
 χολωθείς
 νοῦσον ἀνὰ στρατὸν ὄρσε κακὴν,
 ὀλέκοντο δὲ
 λαοί, (10)
 οὐνεκα τὸν Χρῦσιν ἠτίμασεν
 ἄρητῆρα
 Ἀτρεΐδης· ὁ γὰρ ἦλθε θοὰς ἐπὶ νῆας
 Ἀχαιῶν
 λυσόμενος τε θύγατρα φέρων τ'
 ἀπερείσι' ἄποινα,
 στέμματ' ἔχων ἐν χερσὶν ἐκβόλου
 Ἀπόλλωνος
 χρυσεὶ ἀνὰ σκῆπτρῳ, καὶ λίσσετο
 πάντας Ἀχαιοὺς, (15)
 Ἀτρεΐδα δὲ μάλιστα δύω, κοσμήτορε
 λαῶν.
 Ἀτρεΐδαι τε καὶ ἄλλοι ἐϋκνήμιδες
 Ἀχαιοί,
 ὕμιν μὲν θεοὶ δοῖεν Ὀλύμπια δώματ'
 ἔχοντες
 ἐκπέρσαι Πριάμοιο πόλιν, εὖ δ'
 οἴκαδ' ἰκέσθαι.
 παῖδα δ' ἐμοὶ λύσατε φίλην, τὰ δ'
 ἄποινα δέχεσθαι, (20)
 ἀζόμενοι Διὸς υἱὸν ἐκβόλον
 Ἀπόλλωνα.

The wrath sing, o goddess, of Peleus'
 son, Achilles, (1)
 ruinous, which brought countless suffer-
 ings upon the Achaeans,
 and hurled down to Hades many excel-
 lent souls
 of heroes, and their bodies, it left them
 prey to the dogs
 and birds of all kinds, and so the will of
 Zeus was done, (5)
 from the time the two first began their
 stand-off,
 the son of Atreus, lord of men, and
 divine Achilles.
 But who was it among the gods who set
 them up to fight?
 It was the son of Zeus and Leto: for he
 was angry with the king,
 and he awoke a plague among the army,
 a terrible one, and the people were
 dying, (10)
 because the son of Atreus had disres-
 pected his priest, Khrúsēs.
 He had come to the fast ships of the
 Achaeans,
 wanting to free his daughter, bringing
 infinite gifts,
 holding in his hands the insignia of
 apollo the far-shooter,
 on his golden staff, and he implored all
 of the Achaeans, (15)
 and especially the two sons of Atreus,
 leaders of men:
 "Sons of Atreus and all other strong-
 greaved Achaeans,
 may the gods, who inhabit the houses of
 Olympus, grant you
 to take the city of Priam, and to return
 home unscathed.
 But free my daughter, and accept my
 gifts,
 appeasing the son of Zeus, Apollo the
 far-shooter."

<u>Ἐνθ' ἄλλοι μὲν πάντες ἐπευφήμησαν</u>	<u>And then all of the other Achaeans</u>
<u>Ἀχαιοὶ</u>	<u>called out in approval,</u>
αἰδεῖσθαί θ' ἱερῇα καὶ ἀγλαὰ δέχθαι	to show respect to the priest and to
ἄποινα·	accept the splendid gifts:
<u>ἀλλ' οὐκ Ἀτρεΐδῃ Ἀγαμέμνονι ἦνδανε</u>	<u>But this did not please the <i>thūmós</i> of</u>
<u>θυμῶ,</u>	<u>Agamemnon, the son of Atreus;</u>
ἀλλὰ κακῶς ἀφίει, <u>κρατερὸν δ' ἐπὶ</u>	instead, he sent him away badly, and
<u>μῦθον ἔτελλε· (25)</u>	<u>he gave him a harsh command: (25)</u>

In this sample, there is hardly a line without underlining, which means that there is hardly a line whose component expressions do not also appear somewhere else in our corpus. The message here is that the poet does not seem to be striving for originality. Rather, he seems to be putting verses together the way a child assembles a Lego castle: by snapping together prefabricated bricks (i.e., the underlined parts).

As confirmation of this theory, metrical blemishes may often be found at the junctures between bricks: sometimes, the poet will try to snap together pieces that don't perfectly fit (something that, admittedly, the Lego simile does not allow), and a small metrical bump will result. The classical study is Parry (1971: 201–21): for instance, in order to mention Telemachus in the second half of the line, the poet relied on the noun–epithet formula Ὀδυσσεύς φίλος υἱός “Odysseus’ dear son,” which is isometric (i.e., metrically equivalent) to many other famous noun–epithet formulas (βοὴν ἀγαθὸς Διομήδης “Diomedes good at the war-cry,” βοὴν ἀγαθὸς Μενέλαος “Menelaus, good at the war-cry,” πολὺτλας δῖος Ὀδυσσεύς “much-suffering divine Odysseus,” etc.), but begins with a vowel. While the latter formulas can happily follow a formulaic expression containing the verb ἤρᾱτο “s/he prayed,” the former cannot: when the poet, used to combining ἤρᾱτο with noun–epithet formulas of that shape, tries to snap the pieces together, a metrical bump (in this case, hiatus, i.e., the meeting of two vowels at a word boundary) results:⁴

- (2) δὴ τότε ἔπειτ' ἤρᾱτο βοὴν ἀγαθὸς Διομήδης (*Il.* 5.114)
and then Diomedes good at the war-cry prayed.
- (3) ὧς δ' αὖτως ἤρᾱτο Ὀδυσσεύς φίλος υἱός. (*Od.* 3.64)
thus in this manner Odysseus’ dear son prayed.

⁴ For an introduction to Homeric metrics (and metrical bumps therein), see Chapter 2.

These small bumps are a strong indication that the poet is composing by juxtaposing the bricks, and that he has come to rely on this strategy so much that, sometimes, he will disregard the meter to continue composing in this way.

But who made the bricks? Are they the poet's invention? And how could one go about establishing this either way? Parry observed that, in Homer, noun–epithet formulas, when considered together, appear to form a system which displays both *extension* and *economy* (or *thrif*t). For each task, the poet has just as many different-sized bricks as they need (*extension*), and virtually nothing more (*economy*). Extension and economy are exemplified in Parry's charts for noun–epithet formulas, one of which I partially reproduce in Table 1.1.⁵

Next, Parry turned to poets who wrote in the epic tradition and who attempted to imitate Homer's style, such as Virgil and Apollonius Rhodius. He showed that these poets behaved differently from Homer. While they did rely on some premade expressions (akin to Homer's noun–epithet formulas), there appeared to be no system in place: there were too many bricks for some tasks, and none for others, with no regard for economy or extension (Parry 1971: 24–36).

Parry argued that this difference could be explained by tradition: while Virgil and Apollonius made their own bricks (and just for a few tasks), Homer inherited them, in very large numbers, from the poets before him. It was the force of tradition which, over generations, and through a process similar to natural selection, strategically shaped the bricks into the elegant interlocking system that Homer had at his disposal.⁶ In other words, Parry concluded, Homer's technique was (mostly) traditional, while that of Virgil or Apollonius was (mostly) individual.⁷

⁵ One should notice how, even in Parry's formulation, economy is a strong tendency rather than an absolute law: even among noun–epithet formulas for main heroes, one finds equivalent formulas, though usually one of the variants is overwhelmingly more common than the other.

⁶ An attempt to more fully articulate this process of evolution and renewal in the technique is seen in Hainsworth (1978). Gray (1947) studies the system of epithets for metal weapons in an attempt to uncover how the tradition develops new phraseology for technological innovations.

⁷ Sale (1996) extends Parry's study by comparing extension and economy of formulas (as well as other criteria) in Homer and Quintus Smyrnaeus, demonstrating how even a very good literary imitator of Homer such as Quintus could not match the formal properties of Homer's formularity.

Table 1.1 *Parry's noun-epithet formulas for Odysseus and Achilles in the nominative (after Parry 1971: 39)*

	Between the feminine caesura and the end of the line	Between the hephthemimeral caesura and the end of the line	Between the bucolic diaeresis and end of the line
Odysseus	πολύτλας δῖος Ὀδυσσεύς "long-suffering divine Odysseus" (38)	πολύμητις Ὀδυσσεύς "Odysseus of many counsels" (81) πτολίπορθος Ὀδυσσεύς "Odysseus conqueror of cities" (4)	δῖος Ὀδυσσεύς "divine Odysseus" (60) ἔσθλ' Ὀδυσσεύς "good Odysseus" (3)
Achilles	ποδάρκης δῖος Ἀχιλλεύς "divine Achilles who runs to the rescue" (21)	πόδας ὠκύς Ἀχιλλεύς "swift-footed Achilles" (31) μεγέθυμος Ἀχιλλεύς "Achilles of the great <i>thúmos</i> " (1)	δῖος Ἀχιλλεύς "divine Achilles" (34) ὠκύς Ἀχιλλεύς "swift Achilles" (5)

1.1.2 *Homer's Orality and the Quantitative Study of Formulas*

The type of tradition that Parry had in mind came into focus in his later work, thanks to his experiences in the field. Between 1933 and 1935, Parry, accompanied by his student Albert Lord, traveled to then-Yugoslavia to record the performances of singers in the Islamic tradition of oral epic.⁸ There, he recognized many parallels between the technique of the poets he encountered (who were composing their songs *in performance*, and not simply reciting them from memory) and the formal features he had observed in Homer's diction (here, too, singers relied on formulas and formulaic expressions). He concluded that Homer's technique must also have been developed in the context of an oral tradition, and with the specific goal of supporting oral composition in performance.

Naturally, these observations raised several additional questions about Homer and his poems, which continued the centuries-old tradition of the Homeric question: was Homer (whatever we mean by the term) an oral poet or did he simply behave like one? If he was an oral poet, and he composed his poems orally, how did the poems come to be written down?⁹ Or was Homer perhaps an exceptional figure who used his training as an oral poet to compose his poems in writing, thus allowing them to survive?¹⁰

Answers to these questions were pursued, at first, empirically: Parry and Lord sought to prove that a skilled oral poet could compose a work of the length of a Homeric epic without the aid of writing.¹¹ They found Avdo Međedović, who, over the course of several days, could dictate a poem of the length and complexity of a Homeric epic.¹² They named him the Homer of the Balkans.

But a typological parallel was not enough: the next step was to identify a measurable feature in the Homeric poems that could speak to their oral composition. At this point, many scholars turned to *quantitative formula*

⁸ The series *Serbo-Croatian Heroic Songs* (Harvard University Press) presents some of the materials collected by Parry and Lord. The online portion of *The Milman Parry Collection of Oral Literature* is available at <https://library.harvard.edu/collections/milman-parry-collection-oral-literature>.

⁹ For the process of textualization in oral traditions, see Honko (2000).

¹⁰ Recent and specific attempts to answer all of these questions have been made by Skatfte Jensen (2011), who envisions the epics as oral-dictated texts, and Martin West (2011 and 2014), who envisions the authors of the *Iliad* and *Odyssey* as writing poets. An important recent contribution is Ready (2019).

¹¹ I.e., by relying on a traditional technique of oral composition in which poems are put together in performance by relying on traditional story patterns (*themes*) and linguistic expressions (*formulas*). For an introduction, see Lord (1960).

¹² The poem in question is *The Wedding of Smailagić Meho*, edited and translated by Lord in 1974.

analysis – that is, counting the density of formulas in the *Iliad* and the *Odyssey*. Thus, Lord argued:

There are ways of determining whether a style is oral or not, and I believe that quantitative formula analysis is one of them, perhaps the most reliable. (Lord 1968: 16)

The logic was appealingly simple: if prefabricated expressions (i.e., expressions that are not invented at the moment of performance, but that are arguably stored in the memory of the poet) are indicative of oral composition in performance, then a poem made overwhelmingly of prefabricated expressions ought to be considered an orally composed poem.¹³

Crucially, for many years, there were no automated ways of obtaining such counts, and there were no similar studies looking at prefabricated expressions in natural language. As a result, the few counts that were made were partial (such as example (1) above, comprising only twenty-five verses of the *Iliad*), and did not look at natural language for comparison. This led to a systematic overestimation of the portion of formularity in Homer (up to 90 percent, according to Parry and Lord – and based on example (1) above), and to a systematic underestimation of the extent of formularity in natural language (which was not usually discussed). Years after the beginning of the debate, Lord writes:

What is clearly needed most desperately is a moratorium on baseless speculation about formula quantity and in its stead active research in formula incidence and density, both in Homer and in oral poetry. (Lord 1968: 19)

Lord's wish was fulfilled when formulaic counts finally started to appear, either through painstaking hand-counting (e.g., Cantilena 1982), or, much later on, through computerized concordances (Pavese and Boschetti 2003).¹⁴

Yet, the question of the orality of Homer remained far from settled. In the absence of comparison with a wider selection of texts, and most of all

¹³ Naturally, actual practice (as we shall see below) was more complicated, since one had to grapple with many methodological questions: what expressions should one count, and how can we definitely tell that they are prefabricated? Some expressions are identical to some others (thus likely to be prefabricated), but others are only almost, not precisely, identical. Where to draw the line? And finally, are prefabricated expressions all that matters in establishing the orality of a text? An extensive attempt to answer this last question was made by Peabody (1975).

¹⁴ Existing quantitative formular analyses for the Greek epics include (partial analyses in italics): *Iliad* and *Odyssey* (Parry 1971, Lord 1960, 1968, Hainsworth 1968, Danek 1998, Pavese and Boschetti 2003), *Homeric Hymns* (Cantilena 1982), *Scutum* (Venti 1991), *Batrachomyomachia* (Camerotto 1992), Hesiod (*Minton* 1975, Pavese and Venti 2000).

with natural language, the counts were often interpreted in a circular fashion: whatever amounts of prefabricated expressions Homer showed (set at around 50 percent in Pavese and Boschetti 2003) were argued to be indicative of oral composition, and whatever smaller amounts later texts showed were argued to be indicative of a transition to writing.¹⁵ Few scholars hailed these results as conclusive – unsupported by agreement on a larger theory of formularity, the empirical findings largely fell by the wayside. Scholars could not agree, in fact, on what they had been measuring in the first place.

1.1.3 *Formulas and Their Flexibility*

For several decades, the exact definition of formula had been an arena of constant battle.¹⁶ To this day, most Homerists agree to disagree on the matter, or are content to adopt the term in a generic manner (i.e., to refer to any phraseology in Homer that appears to be repeated, and thus traditional). In practice, formulas proved hard to pin down for two main reasons: they are a hybrid phenomenon (in that they can be defined at the textual or at the psychological level) and a gradient one (in that each expression in Homer can be arranged on a scale from more formulaic to less formulaic).

Parry famously defined formula as follows:

An expression regularly used under the same metrical conditions to express an essential idea. (Parry 1971: 13)

The definition combines two very different elements: a textual entity (an expression found regularly under the same metrical conditions) and a psychological entity (the essential idea). Over the next few decades, and depending on the specific nature of their inquiry, scholars often ended up privileging one side of the definition or the other. Those interested in quantitative analysis needed a text-based definition of what could count as a formula or not, and were not in a position to focus on the *psychological reality* of the formula. Repetition in a text was considered a sufficient criterion for establishing the formulaic status of a sequence.¹⁷ Scholars

¹⁵ Never mind that, when measuring the formularity of a small corpus using Homer as reference, we are measuring how similar that corpus is to Homer (i.e., how much phraseology it shares with the *Iliad* and the *Odyssey*) more than anything else.

¹⁶ Edwards (1986, 1988) gives an impressively detailed account of the debate.

¹⁷ When dealing with living speakers, one can rely on a series of established psycholinguistic tests to verify whether a linguistic sequence (expression) is stored as a whole in memory or generated on the spot. This possibility is, of course, absent when discussing closed corpus languages. Schmitt et al. (2004) argue that corpus data on its own is a poor indicator of whether a linguistic sequence is stored

who were more interested in the process of oral composition and understanding the poet's technique were instead naturally drawn to a psychologically based understanding of the formula, which emphasized the gradience of formulaic phenomena and the remarkable flexibility of the system.

Already in *Homer and Homeric Style*, Parry was well aware that formulas exist along a continuum going from fixity to flexibility. Commenting on the *Iliad* passage quoted above in example (1), he noted:

I have put a solid line beneath those word-groups which are found elsewhere in the poems unchanged, and a broken line under phrases which are of the same type as others. In this case I have limited the type to include only those in which not only the metre and the parts of speech are the same, but in which also one important word or group of words is identical, as in the first example: μῆνιν ... Πηληϊάδεω Ἀχιλῆος and μῆνιν ... ἑκατηβόλου Ἀπόλλωνος. (Parry 1971: 301)

In other words, there are *formulas* (expressions that are found elsewhere in the poems unchanged), and there are *formulaic expressions* (expressions which are of the same type as others). This last category is left somewhat vague by Parry, but it seems like it could be easily extended to describe a vast amount of data.

In this direction, Russo (1966) coined the concept of the *structural formula* – a pattern of expression where the meter and the parts of speech (i.e., noun, verb, etc.) are the same, but no words or word groups are shared. Importantly, a structural formula has no essential idea: it is pure structure. As an example, the structural formula [–]Verb [–]Noun can be used to describe the two expressions τεῦχε κύνεσσιν “threw to the dogs” and δῶκεν ἑταίρῳ “gave to his/her companion,” where the only “idea” shared is that of having a finite verb followed by a noun in the dative (in this particular case, one could speak of a shared *argument structure* – i.e., a similarity at the level of syntax). While many scholars would agree that such patterns are to be found in our texts, many regard them as too generic and abstract to meaningfully qualify as formulas (see Kiparsky (1976: 89–90), who suggests that these phenomena are common in written poetry as well).

Even with respect to the canonical Parrian formula, the flexibility and creativity of Homeric diction was gradually vindicated. Hoekstra (1964) illustrated how poets could renew their hoard of formulas in order to accommodate linguistic innovations; specifically, he studied how recent

in the mind as a whole (i.e., whether it is a “psychological” formula), though the study is very limited in its scope and methods, and more research is needed to verify its results.

sound changes in the Ionic dialects may have had an impact on (and forced restructuring of) some formulaic systems, also trying to use this phenomenon as a way of dating the composition of the poems (which, in his view, would have to have happened shortly after these changes took place).

Going further, Bryan Hainsworth demonstrated how poets could adapt formulaic sequences to the needs of composition: formulas could be moved to other parts of the line, expanded, separated, or morphologically inflected (these modifications are illustrated in his 1968 monograph *The Flexibility of the Homeric Formula*). Hainsworth understood formularity as a living synchronic and diachronic system, in which *frequency of usage* determined what was fixed and what was flexible:

Highly schematized formula-types are then the consequence of ossification of more flexible systems at points of frequent use. (Hainsworth 1968: 113)

That is, the more a poet has to use a given expression, the more that expression is likely to become fixed.¹⁸ Hainsworth (1962, 1978) argued that frequency also played a role in establishing which formulas would stand the test of time, and which would be replaced by other, newer creations. We will see that this attention to frequency puts Hainsworth in tune with many contemporary approaches to formularity in language in general.

Kiparsky (1976) was a substantial theoretical step forward, in that it marked the introduction of tools from linguistic theory (namely, generative syntax) into the study of formularity.¹⁹ Kiparsky compares Homeric formulas to *bound phrases* in natural language – that is, idioms, such as *kick the bucket*, and fixed collocations, such as *foreseeable future*.²⁰ He further distinguishes between *fixed bound phrases* (which are effectively retrieved from memory, not generated, and can thus have odd syntactic behavior and noncompositional semantics) and *flexible bound phrases* (which are generated anew and should thus be syntactically and semantically well behaved). These would correspond, respectively, to fixed formulas and flexible formulas in Homer.²¹ In Kiparsky's model, flexible formulas in Homer ought to

¹⁸ One is reminded here of the famous dictum by John Du Bois, “grammars code best what speakers do most” (Du Bois 1985: 363).

¹⁹ Admittedly, some terminology inspired by generative grammar had already been introduced by Nagler 1967 (see below in section 1.1.4), but this did not amount to a true linguistic approach to the issue.

²⁰ The main difference between idioms and fixed collocations is that idioms are *non-compositional* in their meaning (i.e., their meaning does not transparently arise from the sum of their parts).

²¹ In replying to a comment by Calvert Watkins, Kiparsky admits that this strong bipartition is stipulative: “I cannot prove that they are exactly two categories. It might be that there is a continuum, for example: fixed formulas, flexible formulas, and all kinds of gradations of flexibility in between. And I don't see any way of settling the matter” (Kiparsky 1976: 114).

correspond to well-formed syntactic constituents (this restriction does not apply to fixed formulas). Kiparsky does away with any metrical requirements in his definition. As he puts it, the true essence of the formula is the *abstract bond* between the formula's components: for instance, the bond between ἄλγος "pain" and παθ- "suffer" (Kiparsky 1976: 86),²² or τεύχεα "weapons" and καλὰ "beautiful" (Kiparsky 1976: 87); it is not, then, surprising that a flexible formula would be split across multiple lines, as happens to τεύχεα καλὰ in *Iliad* 22.322–23, 19.10–11, and 18.82–84. One could summarize Kiparsky's flexible formula as a syntactic constituent that is filled by lexical items that have a strong tendency to co-occur with each other (i.e., lexical items which are collocates of each other), and Kiparsky's fixed formula as any linguistic sequence which is entirely retrieved from memory.

But the zenith of flexibility within the concept of formula was arguably reached by scholars operating within a historical perspective. Here, Watkins (1995), working on Indo-European poetics, argued that very ancient formulas, encapsulating important cultural *themes*, can be preserved over centuries while continually undergoing *lexical renewal*.²³ In Watkins' diachronic approach, everything that identifies a formula as such is its essential idea (its *theme* – e.g., the idea HERO SLAYS DRAGON), while the specific lexical items chosen to express this idea might change over time and space.²⁴

This stance has radical consequences: if we look back at Parry's definition, where a formula was intended as (1) a recurring fixed expression, (2) occurring under the same metrical conditions, and (3) expressing a given essential idea, we see how meter and fixity of expression have been

²² For these examples, see discussion in section 1.4.2 below.

²³ Note that, in Lord's *Singer of Tales*, formulas and themes were objects of radically different size and nature: formulas pertained to the diction, and themes to the narrative structure (Lord 1960: Chapter 4 fn. 1 recognized that themes in this sense corresponded to *motifs* as classified in the field of folklore studies). So a formula could be "swift-footed Achilles" (which does not correspond to any theme, unless we want to elevate the idea of a hero being fast to a theme in itself), while a theme (which Lord 1960: 68 defines as "a grouping of ideas regularly used in telling a tale in the formulaic style of traditional song") could be "the assembly," or "the recognition," or even "the return of the hero." In his definition, Watkins erases the distinction in scale between theme and formula, effectively focusing on some themes that can be expressed by a single formula (these are normally formulas centered on a finite verb).

²⁴ In a recent contribution, Kiparsky (2017: 156) embraced the idea that themes (in Watkins' sense), rather than formulas, should be seen as central to creativity in oral-traditional poetry. Themes allow for great flexibility; they can be exchanged and borrowed between traditions, and leave room for individual expression. In this definition, a theme can be as abstract as the idea of "magical growth or paradoxical disproportion" as it appears in several episodes of the *Kalevala*. Next to themes, Kiparsky recognizes the existence of structural formulas (à la Russo) as a means of formal organization of the diction.

essentially done away with, leaving the “essential idea” as the only peg on which to hang the entirety of oral-formulaic theory.

1.1.4 The Disappearance of the Formula

One risk of proceeding in this direction (i.e., taking the essential idea as the only defining element for a formula) is that of doing away with formularity entirely, either by denying that formularity is different from nonformularity, or by arguing that everything in Homer is formulaic by virtue of being there. At one extreme, those wishing to do away with formulas could cast them as an epiphenomenon: Nagler’s “generative” approach (1967) to the formula insisted that formulas were generated anew every time, and that they only appeared identical in every iteration because they satisfied identical constraints, in the way a calculator will always produce the same result for the same arithmetical operation. According to Nagler, formulas only existed as a *preverbal Gestalt* (i.e., Parry’s essential idea) in the poet’s mind: for instance, every time a poet tried to express the idea of Achilles after the hephthemimeral caesura, the string “swift-footed Achilles” was the only possible combination he could come up with. While Nagler’s explanation is a logical possibility (and might even be true for some repeated expressions in the poems), we now know, from the point of view of linguistic processing, that generating the same sequence anew over and over instead of simply retrieving it from memory is a very poor strategy for language production, and not the way human brains generally operate. This view also fails to explain how some formulas may preserve older linguistic features that are not part of the poet’s active grammar (if a poet generates each expression anew every time, wouldn’t those expressions always be linguistically up to date?).

Several decades later, Visser’s (1987, 1988) study of battle scenes argued that Homer’s process of composition proceeded by words, not formulas: core words (the *nucleus*) were placed in the line first, and the formulaic system only supplied metrical filling (the *periphery*), which was semantically vacuous, in order to help the poet complete his lines. According to Visser, this would make Homer similar to any “writing” poet wrestling with the strict demands of the hexametric line. There are several issues with Visser’s argument, which cannot be fully explored here; some basic limitations seem to stem from approaching language production without making reference to some foundational concepts of syntactic theory.²⁵ For instance, the idea that a writing poet (or a normal speaker) would “compose” by single words is in itself

²⁵ For an introduction to syntactic theory, see Adger (2003) or Carnie (2013).

problematic: language is organized and produced by *constituents* (e.g., noun phrases, verb phrases, etc.), not single words.²⁶ Visser's own concept of nucleus and periphery, moreover, largely overlaps with the basic syntactic concept of *headedness* – that is, the fact that constituents have heads (what Visser would call *nucleus*), as well as complements and adjuncts (what Visser would call *periphery*). But there are issues on the Greek side as well: Visser's study is limited to battle scenes where the poet is trying to express the idea “X killed Y” by fitting every element of the sentence (including the full names of the killer and the victim) in a single hexameter line. This is only one of the many possible options, and not even the most frequent (some killings are recounted over several lines, some in half a line; characters can be referred to by pronouns or ellipsis, and these options are actually the most frequent for subjects, largely because of discourse considerations). Most importantly, Visser overstates the freedom of word order in Greek, and assumes that the poet could arrange their words in any linear order needed to satisfy the meter without changing the meaning of the sentence. This is not the case: in the first place, as we have known for a long time, word order *within constituents* is not free in Greek (e.g., in a prepositional phrase, a preposition should come before its complement; definite articles do not follow the noun they modify, etc.); second, as we now know, after decades of research on syntax and information structure in Ancient Greek, different constituent orders (e.g., whether the subject and object precede or follow the finite verb) reflect different discourse configurations and result in different meanings.²⁷ Bozzone (2014: 219–22), for instance, shows that among the supposedly synonymous verbs of killing *ἐπέφνε* and *ἐνήρκετο* (both “s/he killed”), the former is used when the discourse is centered on the victims (“Who was killed next?”), while the latter is used when the discourse is centered on the attackers (“Who killed whom next?”). In other words, constituent orders in Greek are not freely interchangeable.

²⁶ The concept of syntactic constituent has been established in syntactic theory at least since the work of Bloomfield (1933). For a history of the issue, see Seuren (2015).

²⁷ In simple terms, Ancient Greek is one of the many *discourse-configurational languages* in which the status of a given referent as new vs. already known to the listener (in more technical terms, whether something is a *focus* or a *topic* respectively) will affect where in the sentence that referent will surface. A known pattern in Ancient Greek is, for instance, for the new information (*focus*) to be placed immediately before the finite verb. This pattern can be observed in the first line of the *Iliad*, where the most important piece of new information (the *focus*), is placed in the preverbal position: thus *μητιν δειδε* “THE WRATH sing.” For recent work on word order in Ancient Greek, see Dik (1995, 2007), Matić (2003), Goldstein (2014); see Bozzone (2014) for an application of these insights to the problem of word order in Homer's battle scenes.

At the other extreme, the definition of formularity was stretched to accommodate whatever blocks would fit in the system (prefabricated or not). While Russo's concept of structural formula was a step in this direction, Nagy's conception, whereby everything that is part of the tradition is formulaic (and vice versa: see, for instance, Nagy 2010), is perhaps now the most often quoted contribution. Nevertheless, erasing the difference between formularity and nonformularity has struck many as unhelpful: many scholars still feel that a well-worn and widespread formula like *swift-footed Achilles* is not quite the same as an isometric expression that only occurs once in our corpus, and might have been the lone invention of a single poet.

Among recent contributions, Bakker (1997: 186–87) argued that formulas are routinized bits of speech;²⁸ while this illuminates the process that creates the formula, it does not contribute to the question of how we can identify one in a text.

The main contributions summarized so far are presented in Table 1.2.

Table 1.2 *Some definitions of formula*

Parry (1971)	<ol style="list-style-type: none"> 1. Group of words: πόδας ὠκύς Ἀχιλλεύς “swift-footed Achilles” 2. Metrical conditions: ~ - - - - 3. Essential idea: Achilles
Russo (1963, 1966)	<p>Structural formula: [- ~]_V [~ - -]_N e.g.: τεῦχε κύνεσσιν “threw to the dogs,” δῶκεν ἑταίρῳ “gave to his companion”</p>
Nagler (1967)	<ol style="list-style-type: none"> 1. Preverbal Gestalt (true formula): idea of Achilles 2. Surface realization (not really a formula): πόδας ὠκύς Ἀχιλλεύς
Hainsworth (1968)	<ol style="list-style-type: none"> 1. Basic Formula: κρατερὰ δεσμά “strong chains” (the mutual expectation between the two words) 2. Modifications: (a) dislocation, (b) modification (i.e., inflection), (c) expansion, (d) separation <p>e.g.: expansion + modification: κρατερῶ ἐνὶ δεσμῶ “in strong chains” (<i>Il.</i> 5.386) separation + modification: δεσμοῖο ~ - κρατεροῦ. “strong [...] chain” (<i>Od.</i> 8.360)</p>
Kiparsky (1976)	<ol style="list-style-type: none"> 1. Fixed formula (a linguistic sequence stored in memory): Ἦμος δ’ ἠριγένεια φάνη ροδοδάκτυλος Ἥως “As soon as early-born rose-fingered Dawn appeared” 2. Flexible formula (a well-formed syntactic constituent): [[ἄλγος]_{NP} παθ-]_{VP} “pain [...] suffer”

²⁸ More recently, Bakker (2013: 159) defines formula as “a phrase that has been created in order to be uttered repeatedly or routinely.” This is part of a discussion on the theme of *interformularity* – i.e., whether we can take Homeric formulas as textually referential, to which we will return below.

Table 1.2 (cont.)

Visser (1987, 1988)	[πόδος ὠκύς] _{PERIPHERY} [Ἀχιλλεύς] _{NUCLEUS}
Watkins (1995: 302)	Theme: HERO [SLAY (*g ^{wb} en-) SERPENT (with WEAPON/with COMPANION)]
Conventionally, the word order is English. Some syntax is implied, though not expressed notationally (e.g., the sentence often exhibits the marked word order Verb-Object). The boxed portion constitutes the basic formula (the HERO is typically not realized overtly).	

Pulled in these opposite directions, the debate exhausted itself by failing to agree on its basic unit of measurement. Many felt that formularity was getting in the way of understanding the poetry, making Homer mechanical and abstract instead of clarifying his art. In what follows, I argue that the study of formularity in Homer was, in fact, suffering from what we can call “the disadvantage of the early start.”

1.2 Formularity in Language

1.2.1 The Disadvantage of the Early Start

We know now that formularity, or *idiomaticity* (i.e., relying on prefabricated expressions which might have conventionalized meaning, whatever their precise length or shape), is not at all a rare phenomenon in human language. On the contrary: it permeates many aspects of language usage and acquisition, and it has in fact attracted extensive study in many areas of linguistics over the past several decades.²⁹ But why, we might ask, did it take us so long to come to this realization?

The perceived exoticism of Homer’s traditional language is, in large part, a historical accident. The fact is that we are rather blind to the occurrence of formularity (i.e., repetition) in our daily lives. We notice it at the extremes (a memorized quote, a cliché, a plagiarized speech at a public event), but we don’t see it or look for it otherwise. In this respect, formularity in language is akin to the many other automatic behaviors that fill our everyday experience and assist us in completing any cognitively demanding task (as we shall see below): it runs quietly in the background, unnoticed.

²⁹ For an overview, see Bozzone (2010), with references.

Figure 1.1 Two pages of Schmidt's *Parallel Homer* (1885: 186–87)

³⁰ For a history of the concordance, see Haeselin (2019), with references. Other literary authors for which concordances had already been made in the nineteenth century are Shakespeare, Milton, and Pope (Higdon 2003: 57).

Tools that would enable a similar study of natural spoken and written English would not be compiled until the 1970s.³¹

There were theoretical reasons behind this disadvantage as well: for a good part of the twentieth century, a very influential theory in linguistics, Generative Grammar (as inaugurated by the work of Noam Chomsky in the 1950s), was concerned principally with the generative, creative potential of the human language faculty – not its formulaic bits. Idiomatcity in language was perceived as exceptional, and pushed to the margins.³² As a result, studies of Homeric formularity and studies of formularity in natural languages were “out of sync” for several decades. Paul Kiparsky, who (as we have seen above) was the first to attempt to reconcile the two areas within the generative framework, was well aware of this fact:

Formulaic diction has been extensively studied, but for the most part as a phenomenon *sui generis*. Noone has attempted to compare systematically the phrase patterns of oral poetry with those of ordinary language. (Kiparsky 1976: 1)

Decades later, we are in a much better position to carry out Kiparsky’s wish. Within generative theory, much more attention has been devoted to explaining the idiomatic and selectional restrictions that affect otherwise productive rules, and the role of the lexicon has been steadily expanded. Generative theory is not alone in this regard: other areas of linguistics have in fact also been busy exploring formularity for the past several decades, and providing us with theoretical insights and practical frameworks that can now be usefully applied to Homer. These areas are corpus linguistics, language acquisition studies, and usage-based linguistics – and it is into these areas that we shall venture next.

1.2.2 Formularity in Corpus Linguistics, Psycholinguistics, and Historical Linguistics

One of the first results of the development of corpus linguistics, since its beginnings in the 1970s, was the realization that idiomatcity was a much

³¹ For a history of corpus linguistics, see Facchinetti (2007).

³² “How is bound phraseology to be accounted for in the framework of a formal generative grammar? This is a question which has received regrettably little attention in linguistics recently. As might be expected, most of the excitement has for some time been around the new ways of investigating productive syntactic (and phonological) processes which generative grammar has opened up. The less productive regularities of language, notably morphology and phraseology, on which generative grammar does not throw nearly so much light, have been treated as sideshows, though interest in them is clearly beginning to revive” (Kiparsky 1976: 77).

broader phenomenon than previously acknowledged. Fixed linguistic expressions (termed *collocations*³³ in the field) seemed to account for a substantial percentage of the corpora, far from being relegated to the periphery. At the level of language production, scholars in this field started to doubt that syntax was as free as generative approaches assumed.³⁴ There was the so-called *puzzle of native-like selection*:

Native speakers do not exercise the creative potential of syntactic rules to anything like their full extent [...] indeed, if they did so they would not be accepted as exhibiting nativelike control of language. The fact is that only a small proportion of the total set of grammatical sentences are nativelike in form [...] in contrast to expressions that are grammatical but are judged to be “unidiomatic”, “odd” or “foreignisms”. (Pawley and Syder 1983: 193)

Language production seemed to involve large amounts of simple retrieval of stored sequences:

Speakers do at least as much remembering as they do putting together [...]. We are now in a position to recognize that idiomaticity is a vastly more pervasive phenomenon than we ever imagined, and vastly harder to separate from the pure freedom of syntax, if indeed any such fiery zone as pure syntax exists. (Bolinger 1976: 2–3)³⁵

What the field of psycholinguistics has established is that, while the brain does much that is creative, it also does a lot of simple retrieval. In fact, retrieval is often cheaper (i.e., less demanding) from the processing viewpoint:³⁶

The indications from neurophysiology and psychology are that, instead of storing a small number of primitives and organizing them in terms of

³³ “[Collocation] is a *psychological association* between words (rather than lemmas) up to four words apart and is *evidenced by their occurrence together* in corpora more often than is explicable in terms of random distribution” (Hoey 2005: 5, emphasis mine). Note the striking similarity to Parry’s definition of formula.

³⁴ A very readable history of the scholarship is Partington (1998), from which I derive many of the following quotations.

³⁵ In fact, in contemporary generative theories of syntax, pretty much every lexical item is specified for selectionality (in simple terms, almost every lexical item has preferences or requirements for what types of other elements it can combine with) – which puts a heavy burden on the lexicon and notably diminishes the realm of the “fiery zones of syntax” that Bolinger talks about.

³⁶ The amount to which our conceptions of human processing capacities are shaped by the development of information technology is instructive. In the early days of computers, storage was indeed expensive. Bill Gates is quoted as saying in the 1970s that computers in the future will need *little* storage capacity (and that would be a form of progress). The quote is allegedly: “No one will need more than 637 kb of memory for a personal computer” (http://en.wikiquote.org/wiki/Talk:Bill_Gates). The exact opposite has in fact occurred. Similarly, research on brain processing has moved from the view that processing is cheap and storage is expensive, to the view that storage is cheap and processing is expensive.

a (relatively) large number of rules, we store a large number of complex items which we manipulate with comparatively simple operations. The central nervous system is like a special kind of computer which has rapid access to items in a very large memory, but comparatively little ability to process these items when they have been taken out of memory. (Ladefoged 1972: 282)

The field of morphology has perhaps explored these topics to the greatest extent, especially when it comes to determining whether a speaker is generating a morphologically complex word anew (using a productive process in their language) or merely pulling it from memory. Let us take the English word *happiness*, for instance: is the speaker pulling it from memory, or deriving it from its base form, *happy*? And what about the word *bookishness* (memorized or generated)? This question can be tested in the psycholinguistic lab using *lexical decision tasks*, in which speakers are shown morphologically complex words and asked to decide whether they are real words in their language. These types of experiments consistently show that frequent words are more quickly recognized than infrequent ones; a widespread interpretation of this fact is that frequent words are stored, rather than assembled using grammatical processes (so, to answer our question, *happiness* is likely stored, and *bookishness* is likely generated).³⁷ In other words, frequency seems to decide, for each speaker, whether a linguistic string is more likely to be retrieved from memory or generated anew.

Frequency effects go beyond morphology, and can be observed in syntax as well. The process of *chunking*, for instance, happens when speakers begin to store a sequence of words (e.g., a phrase) as a single item, given its frequency of occurrence, and no longer generate it from scratch.³⁸ Everyday examples of chunked sequences include standardized greetings like *Thank you*, *How are you?* and *Bless you*, or frequent replies like *I don't know* (compare the informal spelling *dunno*).

Evidence for chunking can easily be found in the historical record, where chunked items can effectively become a single word, and lose any internal structure; often, erosion of phonetic material accompanies the fusion (see *dunno* above), as well as a semantic shift. A well-known example is the English collocation *going to*, which has now largely developed in the spoken language into *gonna* or even *'ma*³⁹ (along with the phonetic erosion, the meaning has shifted too, from an expression of physical movement to an expression of

³⁷ For one model, see Baayen and Schreuder (1995).

³⁸ For an introduction to chunking, see Bybee (2010: Chapter 3).

³⁹ For instance, the expression *I'ma let you finish* (where the first word can also be spelled as *I'mma*) can be used in some dialects of spoken American English as a more colloquial form of *I'm gonna let you finish*. For a short history of *I'ma*, see Whitman (2010).

tense), or, more simply, the development of the Old English expression *on slæpe* (Middle English *on sleep*) into Present-Day English (PDE) *asleep*. Chunking is in fact the first step in the process of *grammaticalization*, which is a way in which languages create new morphological material out of syntactic units.⁴⁰ According to Bybee (2002), chunking might even be at the root of the hierarchical organization that is pervasive in human language.

1.2.3 Measuring the Idiom Principle

To sum up, speakers seem to operate in at least two ways when producing language: they create some expressions from scratch (following the rules of their grammar), and they retrieve some from memory. The last strategy seems preferable with high-frequency items, so that a speaker can avoid computing the same task repeatedly. John Sinclair captured this duality in language processing – that is, computation vs. retrieval – in the principles of *idiom* and *open choice*:

The principle of idiom is that a language user has available to him a large number of preconstructed or semi-preconstructed phrases that constitute single choices, *even though they appear to be analyzable into segments*. (Sinclair 1991: 110, emphasis mine)⁴¹

The principle of open choice, on the other hand, entails that “at each point where a unit is completed (a word, phrase, clause), a large range of choice opens up and the only restraint is grammaticalness” (Sinclair 1991: 109).

A 2000 study by Erman and Warren sought to measure the extent to which the idiom principle was responsible for the creation of everyday spoken and written texts.⁴² To do this, the authors introduced the concept of the prefabricated unit, or *prefab*:

A prefab is a combination of at least two words favored by native speakers in preference to an alternative combination which could have been equivalent had there been no conventionalization. (Erman and Warren 2000: 31)

⁴⁰ See Bybee (2015: 117–39).

⁴¹ This last point touches on what Langacker has termed the *rulelist fallacy* (Langacker 1987): just because something can be rule-generated, it does not mean that it cannot be stored (i.e., listed) as well. In the framework of *Emergent Grammar* (Hopper 1987), rules “emerge” from storage, and are thus epiphenomenal.

⁴² Composition of the (admittedly small) corpus: seven extracts of 600 to 800 words from *The London–Lund Corpus of Spoken English* plus ten extracts of 100 to 400 words from the *Lancaster–Oslo–Bergen Corpus* (written English) plus two 400-word extracts from two versions of *Goldilocks*.

Note that prefabs are not just repeated word sequences: they are *conventionalized sequences*, in that they display restricted modificability (e.g., they cannot be negated, or pluralized, or undergo gradation, without losing in idiomaticity). Thus, the procedure for finding prefabs in a text has two steps: (1) finding all repeated sequences in a corpus, (2) running a restricted modificability test, to verify which sequences are conventionalized. For instance, a sequence like *black cat* (if repeated) would meet requirement (1), but not requirement (2), since it is not idiomatic and can be freely modified. A sequence like *black box* would meet both (1) and (2), since it is idiomatic, and cannot be modified (a *very black box* is something different entirely). Our definitions of Homeric formularity, which are simply based on repetition (and do not test for conventionalization), are thus less restrictive than the definition of prefab.⁴³ Still, the results of the study (as reported in Table 1.3) were rather striking: more than 50 percent of both spoken and written texts in the sample proved to be made up of prefabricated units, and while a difference was discernible in the amount of prefabs between spoken and written texts, it was rather narrow (a mere six percentage points).

The true difference between spoken and written corpora seemed to lie in the distribution of prefab types rather than in their quantity. For instance, Table 1.4 highlights the different proportions of *lexical prefabs* and *pragmatic prefabs* in written vs. spoken texts. As per Erman and Warren (2000: 38), “Lexical prefabs are semantic units in that they have reference and denote entities, properties, states, events, and situations of different kinds” (e.g., *intensive care, all over the place, here and there, a waste of time, on a clear night*). These appear to be nearly twice as frequent in written texts as in spoken ones. Pragmatic prefabs, on the other hand, “are functional in that they do not directly partake in the propositional content of the

Table 1.3 *Proportion of prefabs in the analyzed texts (after Erman and Warren 2000: 37)*

	Word slots	Filled with prefabs
Spoken	5,000	2,930 (58.6%)
Written	5,246	2,745 (52.3%)
	10,246	5,675 (55.4%)

⁴³ To be fair, it would be much harder to run a restricted modificability test on a repeated expression in a dead language, since we cannot rely on native-speaker intuition.

Table 1.4 *Distribution of prefab types (after Erman and Warren 2000: 37)*

	<i>Lexical</i>	<i>Grammatical</i>	<i>Pragmatic</i>	<i>Reducible</i>
<i>Spoken</i>	38.8%	20.5%	16.7%	24.0%
<i>Written</i>	71.5%	16.9%	2.4%	9.2%

utterance in question [. . .] Most of them are restricted to spoken language and some have functions which could be indicated by punctuation, paragraphing, or in other graphic ways in written texts” (Erman and Warren 2000: 43); examples include *and then*, *and finally*, *and of course*, *but anyway*, *the thing is that . . . you know*, *I mean*, *and so on*, *well I thought*, *as I said*. These, unsurprisingly, are decidedly rare in written texts (2.4 percent), and almost seven times more frequent (16.7 percent) in spoken ones.

These results⁴⁴ should cause any supporter of a quantitative formula analysis to question many of their basic assumptions. If formularity is caused by oral composition in performance and the strictness of the meter, why do we find it in natural language? And what causes it, then? Why is it so extensive in both writing and speaking? And why is writing apparently even more formulaic than speaking in some categories? Is Homer, then, not any more formulaic than natural language, and should we give up hope of demonstrating Homer’s orality?

In order to answer these questions, we need to develop a *general* account of formularity, one that combines insights from linguistics and oral-formulaic theory, and is grounded in what we know about human cognition. It is to this goal that we turn in the next section.

1.3 Formularity in Cognition

1.3.1 *Working Memory, Chunking, and Automation*

While we might pride ourselves on the complex achievements of the human brain, researchers in the cognitive fields have long known that human memory is, in many ways, heavily limited. This is especially true of

⁴⁴ Of course, Erman and Warren (2000) is just one study, on a limited corpus. Subsequent studies have also reported substantial numbers of formulaic sequences in natural-language corpora, if somewhat less than what was reported by Erman and Warren (a recent survey can be found in Read and Nation 2004). Additional research is certainly desirable.

the type of memory that we rely on, constantly, for all of our conscious endeavors: our working memory.⁴⁵ A long tradition in the field of cognitive psychology limits its capacity to just a handful of items, the exact number spanning from four to nine at the most (Miller 1956, Cowan 2001), and of course depending on the nature of the items. In fact, working memory functions as a bottleneck for much of what we do. The reason we are not particularly good at multitasking (despite what we might like to believe) lies precisely in our limited personal RAM.^{46,47}

Yet, this limitation does not stop us from achieving some rather complicated and impressive feats of attention management, like driving a car along a busy freeway, playing a musical instrument, and, perhaps most impressively of all, using human language. How can it be the case that we can carry out all of these resource-intensive activities (and sometimes even two of them at the same time), if our working memory is so limited?

All of the aforementioned complex activities have something in common: no one is able to do them well right away. They all require a long period of training, during which a lot of the component behaviors are performed over and over, until they become entirely automatic (one might say, second nature). Automation really is the key here: what is automatic can run in the background, without taking up space in our working memory. In other words, we can bypass the limitations of our working memory by bypassing working memory entirely.

We already mentioned the notion of *chunking* with regard to linguistic units. The same notion applies to all sorts of cognitive units, including units of human motor activities (*action units*: see Fenk-Oczlon and Fenk 2002: 221).⁴⁸

⁴⁵ Current models of human memory posit, at minimum, three components: sensory memory, working memory, and long-term memory (Baddeley, Eysenck, and Anderson 2009: 6). For further subdivisions of working memory, see Baddeley et al. (2009: Chapter 3).

⁴⁶ RAM, or *random access memory*, is a form of computer memory that is typically used to temporarily store working data, as opposed to a computer's hard drive, which is typically used for long-term data storage.

⁴⁷ There are many classic experiments that illustrate the limitations of our working memory; some of these may involve remembering word lists, or sequences of digits (the classic study is Miller 1956). Perhaps most memorably, the famous "invisible gorilla" experiment (Simons and Chabris 1999) illustrates how, when our working memory is busy with one task, we are effectively blind to much else that happens. In this type of experiment, participants are asked to keep track of some events (like how many times the ball was passed by the members of one team) while watching a short video (originally, this featured a ball game). Because they were busy with this task, participants typically completely missed an otherwise remarkable event in the video (in the original study, this entailed a person wearing a gorilla suit walking across the frame).

⁴⁸ Chunking can also be applied as a mnemonic strategy: it is well known that dividing information into small chunks makes it easier to remember. Everyday examples include the way in which the sixteen-digit sequences on credit cards are written out as four chunks of four digits each. Or the way telephone numbers are written out (and read out loud), which typically involves creating smaller

In fact, we might see linguistic units (especially units of speech like *intonation units* or *prosodic phrases*)⁴⁹ as a special case of more general action units in which all of our behaviors are broken down. While performing each of these units might be effortful, training and chunking can alleviate much of the cognitive load involved. When learning to play an instrument, then, we are building up our repertoire of chunked action units, just as oral poets build up their repertoire of chunked formulaic units, and just as speakers build up their repertoire of chunked linguistic units (see Wray and Perkins 2000). These chunks support the fluent execution of complex behaviors.

In fact, experiments have shown that *mastery* in many fields relies on the capacity to organize information into large chunks, both perceptually and in terms of recall. A classic study of chess players (Chase and Simon 1973) revealed that experienced players are able to handle much larger chunks of information (in this case, the details of a chess position) than novice players, resulting in more accurate recall and faster processing. When shown a chess position for just five seconds, experienced players were able to reconstruct it to a great level of accuracy, while novice players could not.⁵⁰ When looking at the chess board, experienced players did not see isolated pieces: they combined the information into large chunks, which they could then easily hold in their working memory. The same chunking ability assisted experienced players when making quick decisions about the next move: “What was once accomplished by slow, conscious deductive reasoning is now arrived at by fast, unconscious perceptual processing” (Chase and Simon 1973: 56). Any complex, repeated activity will come to rely on chunking.

Chunking similarly assists experienced oral-traditional singers in the impressive feat of being able to “faithfully” reproduce an unknown song that they have heard just once before.⁵¹ While for the untrained listener a song is made up of hundreds of unchunked details (and thus almost

sequences of two to four digits each. A recent popular introduction to mnemonic techniques is Foer (2011).

⁴⁹ In oral-traditional poetry, the basic unit of production is the *traditional word* (Greek *ἔπος*, Serbo-Croatian *reč*), which often corresponds to an entire line or half-line (see discussion in Foley 2002: Chapter 2).

⁵⁰ The difference between the two groups disappeared when an impossible position was shown (i.e., a random arrangement of pieces on a chess board): experienced chess players were good at chunking meaningful chess positions – not just anything.

⁵¹ Of course, *verbatim* reproduction is not the standard by which we measure such tasks. A song will count as “faithfully” reproduced when all the plot points are there, and when it is narrated in the same traditional style. See discussion in Lord (1960: Chapter 5). If these conditions are met, traditional singers will insist that two songs are the same even when their transcripts are quite different.

impossible to remember faithfully), a masterful singer will perceive the song as a combination of large, well-known chunks (provided the song belongs to a tradition with which they are familiar, of course), and will thus be able to easily reproduce it and make it their own.

What is interesting here is the link with creativity: one might think that automatic behaviors resulting from chunking are detrimental to the creative endeavor, but in fact the opposite is true. Just as it does for the experienced chess players, a greater reliance on chunking allows us to engage in more complex tasks more quickly. Freed from the low-level concern of verse-making, oral poets can focus on more complex narrative tasks. Formulas and prefabs in language corpora are just that: an (imprecise) measure of the automatic behaviors (chunks) poets and speakers have come to rely on when producing language. They are the trace of mastery.

1.3.2 *Formularity, Mastery, and Genre*

Of course, mastery might look different depending on the task you are trying to complete, and different circumstances can affect our reliance on automation (acting as dials, in a way, decreasing or increasing the amount of automation required for a given task). It is fully expected that we will find many formulaic sequences in both spoken and written language, since people tend to develop automatic behaviors for what they frequently do (and in our day and age, some of us write perhaps just as much as we speak).

At the same time, the *nature* of the formularity (e.g., the types of prefabs) that we find for each task will depend on the nature of the task itself: thus, in Erman and Warren's terminology, spoken language has more pragmatic prefabs (automatic behaviors for regulating interpersonal communication, like greetings) and written language has more lexical prefabs (automatic behaviors for describing objects and situations). This, of course, is simply a reflection of the types of linguistic tasks that speakers of PDE tend to do more often in one medium than another.

Different genres of language (whether spoken or written) are also likely to develop different types of prefabs: when it comes to professional language, waiters will develop different linguistic habits from lawyers, and might not recognize the prefabs used by the other group as such. The language of oral poetry is really just a specific type of professional language, with its own special types of prefabs, which may happen to be metrical. Within any linguistic community, speakers will share a large number of prefabs, but many others will be limited to given individuals or groups thereof.

In any given text, the amount of formularity will also depend on the individual and their experience: as an academic, I have come to rely on many automatic behaviors that support teaching and academic writing in English, but I am a lot less knowledgeable about verbal chunks that would be useful for describing a football game, or a ballet (or, sadly, for carrying out similar academic tasks in my native Italian or decidedly non-native German). Doing tasks in which we have relatively less training will result in fewer chunks (less mastery, fewer chunks).

Levels of formularity might also change for the same individual depending on the overall challenge of the task at hand: the higher the pressure on our cognitive resources (e.g., having to speak particularly quickly, or when particularly tired), the higher the likelihood of reliance on chunks. For instance, sportscasters tend to rely more on formulaic language when responding to events that are fast, unexpected, or important: “announcers tended to use more clichés when the game deviated from the expected outcome. Announcers also tended to use more clichés in games involving teams that were highly ranked” (Wanta and Meggett 1988: 87). Following this logic, a singer performing for a big audience, or in a high-stakes competition, might be more formulaic than when performing for a small, intimate crowd.

Finally, some genres might explicitly require verbal originality, and might thus encourage us to reduce our reliance on chunks. This is really what Horace is getting at in his *Ars Poetica*, as quoted in the Introduction: we expect originality (though not absurdity) from poets (even though, of course, the poets Horace is referring to are quite different from Homer), and known language chunks can sound weathered and worn; by creating new and effective word combinations, a poet can make language sound fresh again.

At the other end of the spectrum, genres in which *exact wording* is needed to obtain a given result might force us to rely extensively on chunks; we can think of the language of law, or the language of ritual, in which practitioners would be very wary of innovative wording even if they had the time to come up with it (*I proclaim you wife and husband* does not have the same effect as *I pronounce you man and wife*).

Is Homer more like Horace or more like a lawyer? This is a good question. In general, one might contrast the task of the epic poet (telling a traditional story in a traditional way) with that of the lyric poet (evoking traditional stories in a nontraditional way). The theory that an oral epic poet, when given more time to compose (by slowing down his delivery for the purposes of dictation or writing), would rely on fewer formulas actually rests on the assumption that verbal originality would be, in this context, a desirable goal.

It is unclear that this would always be the case.⁵² Some formulaic expressions are arguably there to achieve a given effect; they are rich in traditional associations, and the poet would be foolish to give them up.⁵³ At the same time, originality might be desirable for some areas of the tale. This is true for some types of storytelling that we experience today: when telling the story of *Little Red Riding Hood*, the words spoken between the little girl and the wolf dressed as Grandma (“What big eyes you have,” etc.) must remain the same – and especially the last crucial exchange: “What a big mouth you have”; “The better to eat you with, my dear!”⁵⁴ Ornamentation and improvisation are allowed (and in some cases even encouraged) in other areas of the story. For Homer, it is commonly observed that similes (certainly part of the ornamentation) are lower in formularity and higher in recent linguistic features (Shipp 1972: Chapter 6).⁵⁵

To sum up, the degree of formularity in a given text may be impacted by a multitude of factors, some having to do with the nature of the task, some with the skill level of the speaker, and some with the conditions of performance. If we find two texts, A and B, and the first is higher in formularity and the second is lower, the reasons could be many:

- a. We could be looking at the same speaker under different levels of cognitive strain (higher strain = more formularity), performing familiar vs. unfamiliar tasks (like, paradoxically, dictating a text at an unusually slow pace) or responding to different circumstances which encourage or discourage formularity.

⁵² Lord (1991: 43) observes that oral poets who write down their songs produce lower-quality texts because they don’t have mastery of the written medium, and yet they are moving away from the technique they know: “They become wordy and stilted to the point of being unconsciously mock heroic. The natural dignity of the traditional expressions is lost and what remains is a caricature. The literary technique takes several generations to mature.” With respect to Homer, Lord (1991: 45) makes the argument that the poet of the *Iliad* seems to have all the habits of an oral poet and none of the habits of somebody who is accustomed to writing.

⁵³ Foley (1991) has introduced the concept of *traditional referentiality* to clarify this property of traditional formulas. The expression “swift-footed Achilles” does not simply mean “Achilles”: it provides a link for the audience between the present performance and the hundreds of previous epic performances they have witnessed. It reminds them of all the traditional associations that come with Achilles.

⁵⁴ Note the archaic syntax in the wolf’s response. This is a case where formularity preserves an earlier stage of the language (see more in section 1.4.6 below).

⁵⁵ The usual interpretation of this fact is that similes (and other ornamentation) are the areas where individual poets feel more free to leave their mark, and thus are more open to linguistic (and thematic) innovation. If similes are conceived as pieces of *bravura*, one can imagine individual oral poets rehearsing them in advance, and perhaps even memorizing them in preparation for a performance. Finkelberg (2012) is an updated look at the appearance of more recent linguistic features in the speeches in the *Iliad*, another area that is generally seen as more open to linguistic innovation.

- b. We could be looking at two speakers with different levels of mastery of the same task (higher mastery = more formularity).
- c. We could be looking at speakers who have similar levels of mastery but are trying to achieve different goals (traditionality vs. originality, different genre aesthetics).
- d. Finally, we could be more informed about the types of formularity in text A than in text B (which might represent a subgenre about which we have little information).

While some of these differences might correlate with the spoken vs. written divide, not all do. The reason why Homer is more formulaic than, say, parts of the *Homeric Hymns* could be (b), (c), or (d). The reason why he is more formulaic than Virgil is arguably a version of (c). For this reason, quantitative formula analysis is far from a perfect tool, and one whose results should not be interpreted simplistically.

1.3.3 Collocational Measures in Homer and Other Corpora

With all this being said, we might still rightly wonder about the amount of formularity in Homer vis-à-vis the amount of “formularity” in natural language (spoken or written). Is there really no difference to be observed there? Is there any way to confirm the general intuition that Homer is more “repetitive” than normal speech? Even though a large number of prefabricated sequences are to be expected in many areas of human language, there might still be something quantitatively different about Homer.

In this section, I will discuss *collocational measures* that can be easily obtained using concordancing software and a digitized corpus, and which help us substantiate some of our intuitions about Homer being more formulaic than the norm. These collocational measures might not amount to a proof of orality, but they might allow us to isolate what exactly it is in Homer that we perceive as more automatic and more repetitive than other authors.

A concept similar to that of the prefab, but a lot more neutral, since it does not presume psychological reality or syntactic constituency (or even semantic contentfulness), is that of *collocation*. Collocations are text-based units formed by two or more orthographic words which tend to occur close to each other in a given corpus.⁵⁶ For instance, the words *foreseeable* and *future* constitute a collocation in English, since finding the first one in

⁵⁶ Words that are *collocates* (i.e., take part in a collocation) do not need to be immediately adjacent to each other, though many of the examples discussed below are. When looking for collocations in a text, one can specify a collocation window span (e.g., 5L 5R, meaning five words to the left

Table 1.5 *The ten most frequent two-word, three-word, four-word, and five-word collocations in the LOB corpus*

	2-word	3-word	4-word	5-word
#1	of the	one of the	the end of the	at the end of the
#2	in the	there was a	at the same time	and at the same time
#3	to the	out of the	in the case of	in the case of the
#4	on the	the end of	on the other hand	on the part of the
#5	and the	some of the	at the end of	the other side of the
#6	it is	part of the	for the first time	there is no doubt that
#7	for the	there is a	per cent of the	in the middle of the
#8	to be	it was a	i don t know	at the same time the
#9	at the	there is no	one of the most	as a result of the
#10	that the	i don t	as a result of	at the top of the

a text considerably increases the likelihood of finding the second one immediately afterwards, while the words *red* and *future* do not (since finding the first one does not make the occurrence of the second one any more likely).

Table 1.5 is a list of the ten most frequent two-word, three-word, four-word, and five-word collocations from the Lancaster–Oslo–Bergen (LOB) corpus of written English (the corpus also used by Erman and Warren 2000). One should note that most of these collocations appear to us to be substantially smaller and less contentful than a formula or a prefab as defined by Erman and Warren (2000). The average length of prefabs studied by Erman and Warren is three words for lexical prefabs, and about two words for grammatical, pragmatic, and reducible prefabs (Erman and Warren 2000: 40). Yet all of these prefab types constitute some kind of recognizable syntactic constituent (e.g., a noun phrase or an adjective phrase: see Erman and Warren 2000: Table 5), while most two-word and three-word collocations in our table do not. Similarly, most of the collocations in our table would not be units that we would recognize as candidates for Homeric formulas, in that many do not seem to express “a single essential idea.” In other

and five words to the right) within which the collocates can be sought. This is the case, for instance, for the Proximity text search tool provided by the Thesaurus Linguae Graecae (which was employed to obtain some of the data in section 1.4.2 below). For a short introduction to different approaches to identifying collocations in a text, see Gablasova, Brezina, and McEnery (2017).

words, while all prefabs (or formulas) are collocations, the reverse is not true.

The advantage of using collocations instead of prefabs (or formulas) is that collocations can be easily counted in an automated fashion, while the individuation of prefabs relies on manual analysis of each instance and the application of native speaker judgment, which makes the resulting measurements both more difficult to obtain and harder to replicate. What I hope to show below is that what bare, text-based collocational measures lack in sophistication (they are simple measures of repetitiveness, not of actual formularity), they make up for in efficacy.

For our comparison between Homer and natural language, it is enough to say that two-word collocations are extremely common in spoken and written language, and that they can be seen as prime evidence for the phenomenon of *chunking* mentioned above. On the other hand, what tends to be relatively less common in natural language corpora is an abundance of longer collocations – that is, collocations involving three or more words. Below, for instance, are some collocational data for the LOB corpus of written English, showing type and token frequencies⁵⁷ of collocations formed by two words, three words, four words, and five words respectively (see Figure 1.2).

By looking at the token frequency, we can see that two-word collocations are extremely common in the corpus, and that longer collocations are less and less so, as witnessed by the steeply declining slope of our token line. Looking at the relative position of type frequency and token frequency reveals something else: while two-word collocation types are likely to be repeated very frequently in our corpus (e.g., the most frequent two-word collocation type, *of the*, is repeated 9,009 times), this value steeply decreases as our collocations become longer (e.g., the most frequent five-word collocation type, *at the end of the*, is repeated only

⁵⁷ Within corpus linguistics, an important distinction is made between *tokens* and *types*. When calculating word frequencies for a given corpus, this is the difference between counting how many occurrences of a given word are found in that corpus (i.e., how many tokens of a given word are in that corpus), and how many different words (how many types of words) are found in that corpus. In the sentence *A cat sees another cat*, there are five word tokens, belonging to four word types (*a*, *cat*, *sees*, and *another*); the type *cat* has two tokens (i.e., it occurs twice), while all other types have one token each (they are all, within this short text, *hapax legomena*). In the following section, the type and token counts of collocations (as opposed to single words) will be discussed.

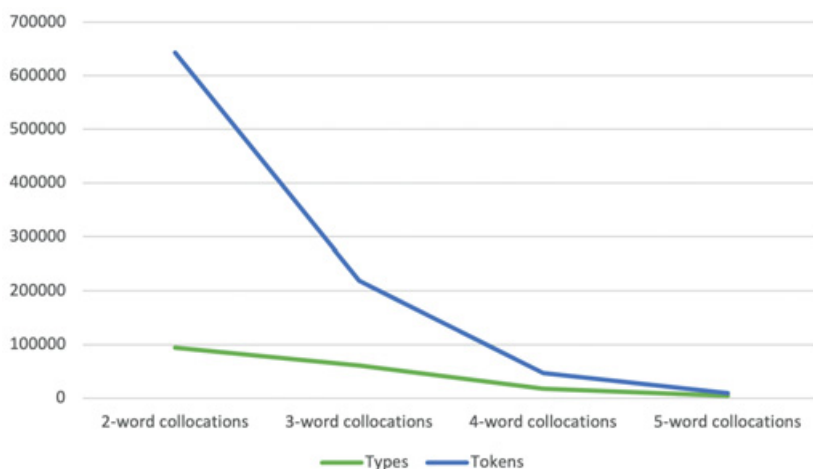


Figure 1.2 Type and token counts of two-, three-, four-, and five-word collocations in the LOB corpus of written English

twenty-eight times).⁵⁸ In other words, the longer the collocations become, the less they are repeated, and the more type and token lines tend to converge.

We can also observe this convergence in the sharp decrease in the proportion of collocation tokens of a given length that are repeated more than twice, and the corresponding increase in collocation tokens that are *singula iterata*. For two-word collocations, about 93 percent of all tokens are repeated more than twice – that is, only 7 percent of two-word collocations in our corpus are *singula iterata*. For five-word collocations, only 16 percent of all tokens are repeated more than twice – that is, 83 percent of five-word collocations are *singula iterata*.

While these tendencies could easily be replicated using many other modern corpora, one might object that the LOB corpus is very large (1,033,210 words), and in English, and as such might not provide an ideal comparandum for Homer. To this end, it would be ideal to study an Ancient Greek corpus of a length similar to that of the *Iliad* and the *Odyssey* combined (ca. 199,000 words). An ideal candidate in this sense is Herodotus (ca. 186,000 words), an author whom the ancients, for independent reasons, called ὁμηρικώτατος “the most Homer-like.” Figure 1.3

⁵⁸ Note that this decline is also to be expected because longer collocates can fully contain smaller collocates, just like *at the end of the* contains *of the*.

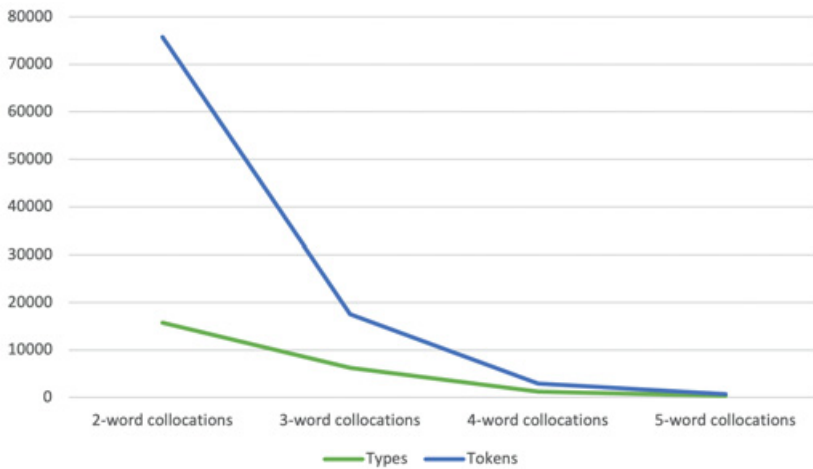


Figure 1.3 Type and token counts of two-, three-, four-, and five-word collocations in Herodotus

shows the type and token counts of collocations that we find in Herodotus, ranging from two to five words.⁵⁹

Overall, the situation in Herodotus seems to replicate the LOB situation, albeit on a much smaller scale (the *Historiae* are about one-tenth of the size of the LOB). Here, too, the number of collocations steadily diminishes with the length of the sequence, and our lines converge around the five-word collocation point. The fall in the ratio of token to type frequency appears somewhat sharper than what we saw in the LOB corpus: at the five-word collocation level, collocations repeated more than twice make up only 10 percent of our tokens, with *singula iterata* constituting 90 percent of the attested five-word collocations.

If we run the same counts using the Homeric corpus (Figure 1.4), we can see that the overall trend is similar: two-word collocations are the most frequent, and the number steadily decreases as we look at longer sequences. At the same time, the token and type lines gradually draw closer to each other.

⁵⁹ These counts were obtained by first extracting the complete texts of Herodotus and Homer using the Classical Language Toolkit (<http://cltk.org>) under Python 3, and feeding them through the software CasualConc (<https://sites.google.com/site/casualconc/home>), which then generated type and token lists of two-, three-, four-, and five-word collocations as requested (using the Word Count function). These lists were then exported into Microsoft Excel, where type and token counts were made for each list. The results were then visualized by entering the token and type counts in a separate table in Microsoft Excel, and generating a graph from the table. The same procedure was used for the counts below regarding Quintus Smyrnaeus.

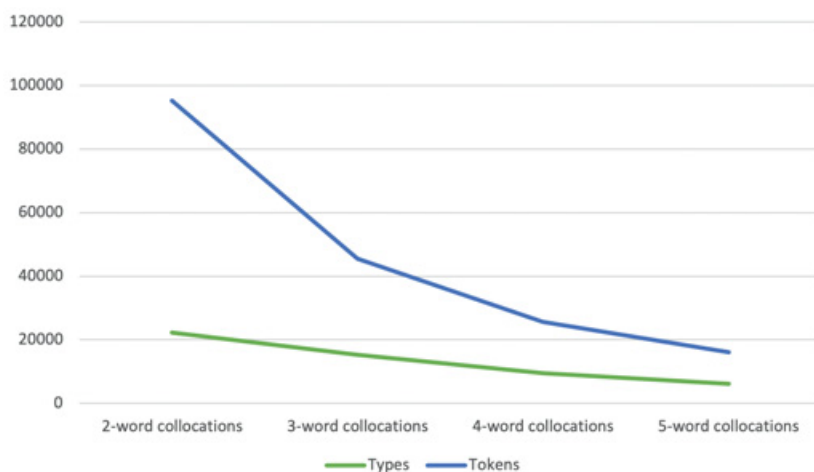


Figure 1.4 Type and token counts of two-, three-, four-, and five-word collocations in Homer

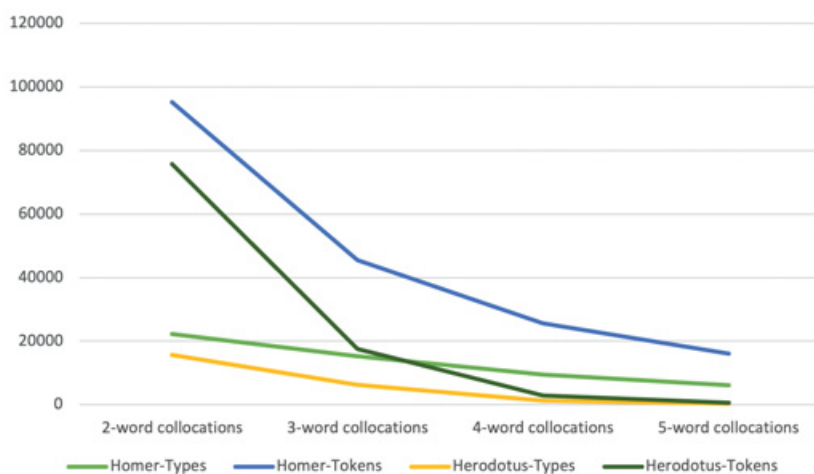


Figure 1.5 Type and token counts of two-, three-, four-, and five-word collocations in Homer vs. Herodotus

Yet, the similarities end here: if we compare Homer and Herodotus, we see that Homer registers significantly more collocations than Herodotus does *at all sizes*, and this is true in terms of both types and tokens (see Figure 1.5 above

for a direct comparison). Even at a superficial level, Homer does appear to be more repetitive than Herodotus.

A closer look at the data reveals even starker differences. The difference in number of collocations between Homer and Herodotus might seem small at the two-word level, though it is already statistically significant. But starting at the three-word level, it becomes noticeably larger. And remarkably, in Homer the token and type lines never touch: at the five-word stage, collocations repeated more than twice still make up 25 percent of all tokens, with *singula iterata* accounting for only 75 percent of the attested five-word collocations. A list of the ten most frequent two-word, three-word, four-word, and five-word collocations in Homer and Herodotus respectively is given in Tables 1.6 and 1.7. (Note that, for text processing reasons, apostrophe signs were removed from the corpus, and as such they do not appear in the tables.)

The next step is to compare our results concerning Homer and Herodotus with some other Greek hexametric poetry that is not suspected of being composed orally. This is to exclude the possibility that the “repetitiveness” that we observed in Homer might simply be due to the realities of composing hexametric poetry, regardless of orality. As it is traditional within the field of Homeric formularity, we can turn to authors like Apollonius Rhodius and Quintus Smyrnaeus – that is, poets who wrote, and who (to different extents) endeavored to imitate Homer’s style and language. Quintus offers a particularly apt comparison here, since several scholars have argued that his technique (while still being markedly distinct from Homer’s) comes closest to a genuine approximation of Homer’s oral style, including developing his own patterns of formularity.⁶⁰

Figures 1.6 and 1.7 contrast the results from Quintus Smyrnaeus (whose corpus counts ca. 61,000 words) with those from Homer and Herodotus respectively.⁶¹ The comparison is instructive: even at first sight, Quintus appears much more similar to Herodotus than to Homer. In Figure 1.7, the shapes and the slopes of the lines are almost identical for Quintus and Herodotus (the only exception being Herodotus having significantly more tokens of two-word collocations than Quintus; other than that, the figures overlap almost perfectly). If we contrast Quintus with Homer (Figure 1.6), on the other hand, we find the usual discrepancy observed above: the type

⁶⁰ See the discussion in Sale (1996) and, most recently, Bakker (2019).

⁶¹ The type and token counts for Homer and Herodotus have been scaled down in Figures 1.6 and 1.7 to match the smaller corpus size of Quintus. Specifically, the numbers for Homer have been divided by 3.3, while the numbers for Herodotus have been divided by 3.1.

Table 1.6 *The ten most frequent two-word, three-word, four-word, and five-word collocations in Homer*

	2-word	3-word	4-word	5-word
#1	τε καὶ	ἔπεα πτερόεντα προσηύδα	τὸν δ αὖτε προσέειπε	δ ἀπαμειβόμενος προσέφη πολύμητις Ὀδυσσεύς (47x)
#2	τὸν δ	δ ἀπαμειβόμενος προσέφη	τὸν δ ἡμείβεται	αὖ Τηλέμαχος πεπνυμένος ἀντίον ἦ᾽ οὐδ' αὖ (41x)
#3	δ ἄρα	δ αὖτε προσέειπε	τὸν δ ἀπαμειβόμενος προσέφη	δ αὖ Τηλέμαχος πεπνυμένος ἀντίον (41x)
#4	ὁ δ	τὸν δ αὖτε	δ ἀπαμειβόμενος προσέφη πολύμητις	ἔπος τ' ἔφατ' ἔκ τ (41x)
#5	δ ἄρ	ἀλλ' ὅτε δὴ	ἀπαμειβόμενος προσέφη πολύμητις Ὀδυσσεύς	καὶ μιν φωνήσας ἔπεα πτερόεντα (35x)
#6	δέ οἱ	δ ἡμείβεται	τ' ἔφατ' ἔκ τ	τὸν δ αὖ Τηλέμαχος πεπνυμένος (30x)
#7	οἱ δ	τὸν δ ἀπαμειβόμενος	Τηλέμαχος πεπνυμένος ἀντίον ἦ᾽ οὐδ' αὖ	μιν φωνήσας ἔπεα πτερόεντα προσηύδα (29x)
#8	οἱ δ	προσέφη πολύμητις Ὀδυσσεύς	αὖ Τηλέμαχος πεπνυμένος ἀντίον	ὥς ἔφαθ' οἱ δ ἄρα (28x)
#9	δ αὖτε	ἐς πατρίδα γαῖαν	ἔπος τ' ἔφατ' ἔκ	ὥς ἔφαθ' οἱ δ ἄρα (28x)
#10	δ ἐν	τὸν δ ἡμείβεται	δ αὖ Τηλέμαχος πεπνυμένος	τὸν δ ἀπαμειβόμενος προσέφη πολύμητις (27x)

and token lines in Quintus meet at around the four-word collocation mark, while in Homer they never do; and the token line in Homer is significantly higher throughout (i.e., Homer contains a lot more repeated sequences than Quintus, at any length).

To summarize our results, there *is* something measurably different about the text of Homer when it comes to collocational tendencies, especially when looking at longer collocational sequences. These longer collocations are notably more common in the text of Homer than they are in all other corpora

Table 1.7 *The ten most frequent two-word, three-word, four-word, and five-word collocations in Herodotus*

	2-word	3-word	4-word	5-word
#1	τε καί	καί δὴ καί	γῆν τε καί ὕδωρ	ἐμοὶ μὲν οὐ πιστὰ λέγοντες (5x)
#2	ἐς τὴν	Ὅ μὲν δὴ	ἐν δὲ δὴ καί	ὥς καί πρότερόν μοι εἴρηται (5x)
#3	δὲ καί	ἐπὶ τὴν Ἑλλάδα	στρατεύεσθαι ἐπὶ τὴν Ἑλλάδα	ἔτι καί ἐς ἐμὲ ἦν (4x)
#4	μὲν δὴ	Οἱ μὲν δὴ	ἔτι καί ἐς ἐμὲ	Ὅ δὲ εἶπε Ὡ βασιλεῦ (4x)
#5	μὲν νυν	ἐς τὴν Ἀσίην	ὥς καί πρότερόν μοι	ὥς καί πρότερόν μοι δεδήλωται (4x)
#6	ἐν τῇ	τοῦτον τὸν χρόνον	περὶ μὲν τῇσι κεφαλῇσι	Μετὰ δὲ οὐ πολλὸν χρόνον (3x)
#7	οἱ δὲ	τῶν ἡμεῖς ἴδμεν	τὸ δὲ ἀπὸ τούτου	Ταῦτα ὥς ἀπενειχθέντα ἤκουσαν οἱ (3x)
#8	ἐν τῷ	τῷ οὔνομα ἦν	Ὅ δὲ εἶπε Ὡ	Χρόνου δὲ οὐ πολλοῦ διελθόντος (3x)
#9	ἐκ τῆς	Ταῦτα μὲν νυν	καί δὴ καί ἐς	δι' ἀλλέων δέκα ἡμερέων ὁδοῦ (3x)
#10	ἐς τὸ	τῶν ἐν τῇ	τά τε ἄλλα καί	δὲ περὶ μὲν τῇσι κεφαλῇσι (3x)

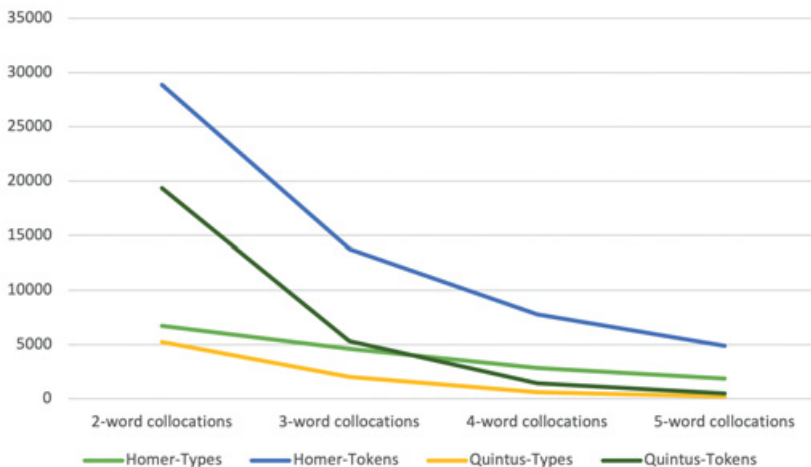


Figure 1.6 Type and token counts of two-, three-, four-, and five-word collocations in Homer (scaled down to match the corpus size of Quintus) vs. Quintus Smyrnaeus

Table 1.8 *The ten most frequent two-word, three-word, four-word, and five-word collocations in Quintus Smyrnaeus*

	2-word	3-word	4-word	5-word
#1	δέ οἱ	᾽Ως ἄρ ἔφη	᾽Ως φάτο τοῖ δ	φάμενον προσέειπεν Ἀχιλλέος ῥβριμος υἱός (5x)
#2	δ ἄρ	᾽Ως δ ὅτ	ἀμφὶ δ ἄρ αὐτῷ	ὁ δ ἄρ οὐ τι (5x)
#3	δ ἄρα	ἀμφὶ δ ἄρ	᾽Ως φάτο τὸν δ	᾽Ως φάμενον προσέειπεν Ἀχιλλέος ῥβριμος (5x)
#4	τε καὶ	Καὶ τὰ μὲν	Καὶ ῥ οἱ μὲν	Καὶ τὰ μὲν ὥς ὥρμαινε (4x)
#5	ὁ δ	ὁ δ ἄρ	Καὶ τὰ μὲν ὥς	τότ ἀρήιοι υἱέσ ἐυσθενέων Ἀργείων (4x)
#6	οὐ τι	ὥς οἱ γ	δ ἄρ οὐ τι	᾽Ως ἄρ ἔφη Τρώων τις (4x)
#7	ἀμφὶ δὲ	Ἄλλ ὅτε δὴ	δέ οἱ οὐ τι	δι ἡέρος ἄλλοτε δ αὐτε (3x)
#8	δέ μιν	δ ἄρ αὐτῷ	τοῖον ποτὶ μῦθον ἔειπε	δὴ τότε ἀρήιοι υἱέσ ἐυσθενέων (3x)
#9	οἱ δ	᾽Ως φάτο τοῖ	ἀλλά μιν οὐ τι	δὴ τότε πυρκαϊήν οἶνω σβέσαν (3x)
#10	δὲ καὶ	ἄλλοτε δ αὐτε	Καὶ νύ κε δὴ	καὶ ῥ ὀλοφυνδὸν αὔσε μέγ (3x)

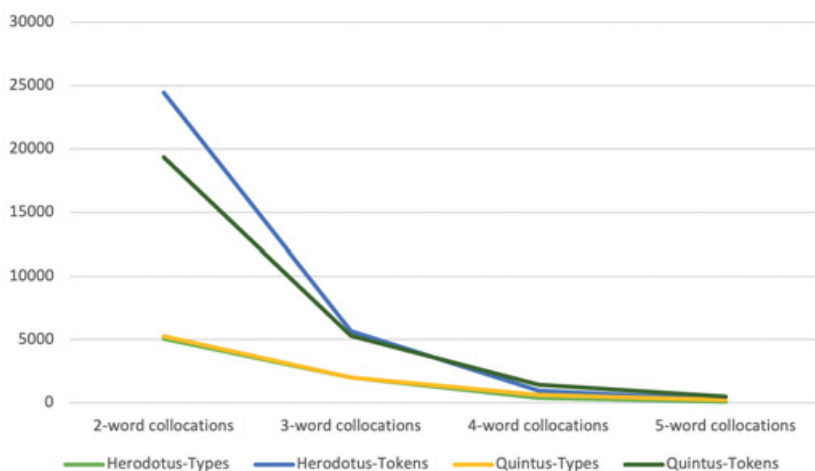


Figure 1.7 Type and token counts of two-, three-, four-, and five-word collocations in Herodotus (scaled down to match the corpus size of Quintus) vs. Quintus Smyrnaeus

under consideration. That is to say, the text of Homer *may* have a similar overall percentage (say, between 50 percent and 60 percent) of prefabricated sequences (as defined by Erman and Warren) or formularity (as defined by Pavese and Boschetti) as spoken and written natural language corpora do (both of these measures would only count a small subset of collocations as valid for their measurements), but we find significantly more *long collocations* in Homer than in other texts.

What does this mean? If we take collocational measures as a sign of *chunking* – that is, as indicating which sequences are likely to be stored vs. generated by speakers – we may say that Homer seems to operate with larger chunks than “normal speakers” do. His reservoir of collocations does not mostly stop at two- or three-word sequences, but keeps providing many options for four- and five-word sequences and beyond. Homer seems to be taking a natural tendency of human language (and cognition in general), and amplifying it. He is like an experienced chess player, who is able to handle much larger chunks (in this case, language chunks) than novice players can.

I would not venture to claim that in these collocational measures we have found a *direct and universal* proof of orality of composition. What we are seeing are the traces of mastery, and the results of extensive, likely years-long training to establish those longer chunks in the poet’s memory.⁶² Following Parry and Lord, I find persuasive the argument that only within an oral tradition would the conditions have arisen for this type of training to take place, and for this type of mastery to be desirable. But we cannot in principle exclude that types of written composition, under the correct conditions, could also yield similar collocational values; we simply have not come across any so far that do. Within the Greek tradition, the fact that Quintus Smyrnaeus specifically fails this test, is to me a strong indication that even literate poets who

⁶² An interesting question is whether, when analyzing a poem by a novice vs. expert oral poet, we would find more or fewer fixed runs – i.e., sequences of several lines that appear to be retrieved as a whole from memory, rather than generated anew during composition (note that this would not necessarily mean more formularity, just more verbatim repetition). Lord’s (1960: Chapter 2) account of the poet apprenticeship suggests that the capacity to use formulaic materials more flexibly is something that develops over time (this also parallels what children do during language acquisition – see Bozzone 2010), and that the apprentice poet is more reliant on “fixed” materials. My prediction is that an accomplished oral poet will know (and thus potentially employ) more different formulaic sequences, as well as fixed runs, than an apprentice poet, because they have had more years to acquire them. In other words, an accomplished poet will be able to choose whether they want to use more or fewer fixed materials in their compositions, depending on the performance conditions and requirements, while a novice poet might have no choice but to rely more on a smaller set of fixed runs and fixed formulas.

attempted to imitate Homer's style did not end up developing as many automatic behaviors as we find in Homer's oral technique.

1.4 A General Theory of Formularity

We have seen that formularity (in a very wide sense) is a common feature in human language and cognition, rooted in the well-understood psychological phenomenon of chunking.⁶³ We have also seen that Homer seems indeed to be more formulaic than natural language or other Greek authors; this is not in the sense that Homer necessarily relies on more formulaic sequences overall, but in the sense that he relies on formulaic sequences that are longer than the ones which normal language users employ (i.e., he uses larger chunks).

But how are we to treat formularity in Homer in practice? How are we to classify different "formulaic" sequences and phenomena, and how do we go about uncovering them in the texts? In what follows, I will sketch out a general theory of Homeric formularity and illustrate the different forms that formularity can take in the poet's diction. I will also provide examples of a formal notation that we can employ to describe such formulaic phenomena. To do so, I shall rely on concepts derived from usage-based linguistics, in particular from the frameworks of *Lexical Priming* (Hoey 2005) and *Construction Grammar* (Goldberg 2006). I will refrain from giving a single definition of the Homeric formula. One of the main problems in the formula debate was that scholars were trying to describe many distinct but similar and interrelated phenomena with just one or two terms. We shall go a different way: I will try to guide the reader through the maze of interconnected meanings and words that form the poetic language, using the concept of *collocation* as a heuristic tool to uncover formulaic phenomena.

1.4.1 *The Memory of the Poet*

Before we start mapping the poet's technique, however, we need to pick a basic cognitive model of how we think information is stored and organized in the poet's mind.⁶⁴ We will go with *connectionism* – that is, the theory that

⁶³ For more on formularity and chunking, see Pagán Cánovas (2020), which expands on Pagán Cánovas and Antović (2016), and whose outlook on formularity and cognition is largely compatible with mine.

⁶⁴ Of course, this is a scientific question that stretches far beyond the limits of this chapter. For our purposes here, we shall proceed with a very simplified model and discussion.

our mind is structured like a massive network which connects simple units working in parallel. The most influential implementation of connectionism in the twentieth century was *parallel distributed processing* (PDP), an approach developed in the 1970s by James L. McClelland, David E. Rumelhart, and the PDP Research Group (Rumelhart and McClelland 1986). Taking the human neural structure as an inspiration, PDP embraced a view of cognition in which complex processes emerged from a large number of simple microprocesses happening throughout the network (distributed vs. localized) at the same time (parallel vs. in sequence). In the 1980s, connectionism was pitted against computationalism (the idea that human cognition proceeds through explicit sequential operations on symbols – like high-level computer programming languages).⁶⁵ Today, connectionist models are alive and well in the form of neural networks employed in machine learning (for an introduction, see Graupe 2013). In linguistic theory, the frameworks of *Harmonic Grammar* (Legendre, Miyata, and Smolensky 1990) and *Optimality Theory* (Prince and Smolensky 1993) partly descend from connectionist models, and are still widely employed today.

From a connectionist perspective, different mental states result from different activation patterns of units (*nodes*) within the network, and the strength of connection between two given nodes is increased every time the nodes are activated at the same time. Let us say, then, that we have two words: *foregone* and *conclusion*.⁶⁶ The connection between the two words is strengthened each time they occur together. After a while, the activation of the word *foregone* is enough to activate the word *conclusion* as well. In the terminology used above, we can say that the two units have become chunked.⁶⁷ Another (more granular) way of expressing a similar concept is that of *priming*: in psycholinguistic experiments, we can see that the word *foregone* primes the word *conclusion* (i.e., showing the word *foregone* to a subject as a *stimulus* makes the subject faster at recognizing the *target*

⁶⁵ For a recent discussion of such problems, see Legendre and Smolensky (2006). A classic (and much more introductory) read on the architecture of the human mind, which introduces concepts of distributed cognition, is Minsky (1986).

⁶⁶ For the purposes of this illustration, we posit that each word would correspond to a node. Of course, a single word might in fact correspond to a given *pattern of activation* of many more nodes.

⁶⁷ Note that, of course, the same words can take part in several different chunks/collocations, and genre or context of usage are also a factor: while *foregone conclusion* might be recognized as a more generic (though formal) chunk of English, *foregone income* or *foregone earnings* are English collocations too, but in the specific context of accounting and finances (and, as such, they might be unfamiliar to some speakers).

word *conclusion* immediately after).⁶⁸ While chunking applies to units that have become so strongly associated as to work effectively as a single node, priming captures weaker associative effects between nodes.

To be more specific, there are actually two ways in which a given element (word, structure, etc.) can prime another: *absolute frequency* (long-term memory), and *recency* (working memory). Classic priming studies really reflect absolute frequency: the word *foregone* primes the word *conclusion* because the two are associated in a speaker's long-term memory, so that activating one activates the other as well. But recency matters too: several priming studies have focused on *syntactic priming* – that is, the tendency of speakers to reuse syntactic structures that have just been employed.⁶⁹ In English, for instance, speakers can choose between two equivalent structures when making a ditransitive clause: they can say *I gave Mary the book* or *I gave the book to Mary*, where the first sentence uses a dative construction, and the second a prepositional construction.⁷⁰ Studies have shown that the activation of a given syntactic structure in the working memory of a speaker prompts them to use it again shortly thereafter. This can happen across speakers or for the same speaker: thus, if somebody is asked a question using a prepositional construction (*To whom did you give the book?*), they are more likely to answer that question using a prepositional construction as well (*I gave the book/it to Mary*). Similarly, if somebody has already used a prepositional construction (*I gave one book to Mary*), they are more likely to use it again shortly after (*and I gave another one to Paul*).

More fine-grained studies (e.g., Gries 2005) have shown that the two types of priming can actually interact; for instance, if a given lexical item (e.g., the verb “give”) is strongly primed (long-term memory) to prefer a given construction, it is more likely to resist syntactic priming (working memory) for a different construction. This is to say that some primings are stronger than others, and some constructions are more likely to resist contextual adaptation.

⁶⁸ For a history of the concept of priming in psychological research and corpus linguistics, see Pace-Sigge (2013). Classical references for priming in psychology are Neely (1977, 1991) and Anderson (1983).

⁶⁹ See Gries (2005) with references.

⁷⁰ The sentence *I gave Mary the book* can also be described as a *double-object construction*. Many English verbs which take an indirect object allow for both a dative construction (i.e., a *double-object construction*) and a prepositional construction, though the conditions of usage might differ, and this is known in the literature as the *English dative alternation* problem or *dative shift*. A similar alternation has been described in many other languages (Indo-European and not). Literature on this problem is immense; an influential study is Hovav and Levin (2008). See most recently Goldberg (2019).

An easy-to-grasp theory of language that builds upon these insights is *Lexical Priming*, a framework developed by Michael Hoey (2005), which argues that linguistic knowledge (lexical *and* grammatical) is really a network of stronger and weaker primings that affect each word, morpheme, and even phonological sequence.⁷¹ What a speaker knows about their language, in other words, is really how likely one element (a word, a morpheme, a phonological sequence) is to co-occur (or not co-occur) with another, based on a lifetime's worth of language exposure (i.e., an ever-updating database that contains all of the speaker's linguistic experiences). When it comes to language production, this approach argues that words (i.e., lexical items) come first, and that grammatical structure emerges from each word's co-occurrence preferences, not vice versa (e.g., in a rule-based approach to language, where one would start from an empty grammatical structure and *then* proceed to fill it with lexical items).⁷²

How do these insights translate to the study of Homer? From a Lexical Priming perspective, we can describe the poetic language as a network of associations, in which ideas, word forms, syntactic constructions, and metrical positions can all come to be associated with each other, each association possessing a different strength. When the association is strongest, and when it affects all levels (ideas, word forms, syntactic constructions, and metrical positions), we find prototypical formulaic phenomena. But the technique is really composed by all the other, weaker associations as well. These associations provide the epic text with its texture, its familiar feel, its cohesion, and much of its poetic strength (i.e., the capacity, directly or indirectly, to evoke feelings in its audience). We will keep in mind that primings will exist at the long-term memory level as well as in the working memory level. Working memory primings will explain short-term repetitions. Long-term memory primings will explain long-distance repetitions.

In what follows, we shall look at some examples of such associations, moving from the base of the iceberg (i.e., weaker associations) to its tip (stereotypical formulaic phenomena).

⁷¹ A similar approach to understanding linguistic knowledge is fleshed out by Goldberg (2019), who discusses how words (Chapter 2) and syntactic constructions (Chapter 3) are learned and stored in the brain as "clusters of lossy [i.e., not fully specified] memory traces," resulting in a rich network. We will return to Goldberg and *Construction Grammar* in section 1.4.3 below.

⁷² This, one might add, is not too far from some aspects of the Minimalist program in generative syntax (Chomsky 1995), whereby each lexical item contains stored syntactic information (called features) which controls and constrains syntactic derivations (i.e., the process by which phrases and sentences are created).

1.4.2 *From Themes, to Conceptual Associations, to Collocations*

In Lord's definition, themes are "the groups of ideas regularly used in telling a tale in the formulaic style of traditional song" (Lord 1960: 68). Even more so than formulas, they are the building blocks of traditional storytelling. Themes exist at different sizes: some are as large as entire plot plans (i.e., the theme of the return song), some are the size of *motifs* (e.g., the ones described in the folkloric literature, such as Thompson's 1955 *Motif-Index of Traditional Folk-Literature*), and some of *type scenes* (a banquet scene, an arming scene). Yet, at the most basic level, a theme can also be the simple association between two ideas (i.e., a *conceptual association*), which a poet picks up as part of their training. Many of these traditional conceptual associations (or *mini-themes*) can be seen as the root of formulaic and non-formulaic diction alike. It is at this microscopic scale that Parry's essential idea and Lord's theme come to coincide, and it is here that our description of formulaic phenomena in Homer begins.

There are many expressions in Homer that express the traditional *conceptual association* (*mini-theme*) PAIN and SUFFER.⁷³ Most famously, the root παθ- "suffer" tends to occur in the vicinity of the root ἄλγ- "pain," both within and without recognizable formulaic patterns. We can call this the ἄλγ- + παθ- *collocation*.⁷⁴ To this family belong the well-established formula ἄλγεα πάσχω⁷⁵ "suffering pains" (9x in our poems)⁷⁵ as well as the

Table 1.9 *From conceptual association to collocations*

Conceptual association	PAIN + SUFFER
Collocation	ἄλγ- + παθ-
Formula	ἄλγεα πάσχω ⁷⁵ (9x)
Unique expressions	ἄλγεα πάσχουσιν (<i>Od.</i> 9.121), ἴν' ἄλγεα πολλὰ πάθοιμεν (<i>Od.</i> 9.53)

⁷³ Much like in Watkins' notation for the inherited formula, I use English words in all capitals to convey a concept (i.e., PAIN, the idea of pain) as opposed to a specific lexical realization thereof (e.g., the Greek root ἄλγ-, the English word *pain*).

⁷⁴ Note that Meusel (2020: 25–48), in the context of reconstructing Indo-European phraseology, introduces the distinction between collocations and formulas (along with the categories, which we shall not cover here, of idioms and part idioms).

⁷⁵ We designate this expression as a *formula* (*fixed formula*) because it recurs identically more than once in our poems. A *unique expression*, on the other hand, is an expression that occurs only once in our corpus.

unique expressions ἄλγεα πάσχουσιν “they suffer pains” (*Od.* 9.121) and ἵν’ ἄλγεα πολλὰ πάθοιμεν “so that we might suffer many pains” (*Od.* 9.53).

Other possible surface realizations of the same *conceptual association* (PAIN + SUFFER) are the *fixed formula* (part of a longer formulaic run) πρίν τι κακὸν παθέειν “before suffering something bad” (*Il.* 17.31–32, *Il.* 20.197–98), as well as the unique expressions αἰνὰ παθοῦσα “suffering pains” (*Il.* 22.431) and παθέειν τ’ ἀεκήλια ἔργα “to suffer shameful deeds” (*Il.* 18.77). Even the compound αἰνοπαθής “pain-suffering” (*Od.* 18.201), a *hapax* in the *Odyssey*, belongs to this conceptual association.⁷⁶ These last few expressions exemplify how traditional conceptual associations (mini-themes) underlie both formulaic and unique phraseology, and remain constant even when the diction changes. This is similar to Watkins’ insight, discussed above, that the surface form of an inherited PIE formula could undergo lexical renewal in the daughter languages.

Conceptual associations can give us a glimpse into the process of lexical renewal within the technique. For instance, the *conceptual association* between DARKNESS (=DEATH),⁷⁷ COVER, and EYES underlies an entire family of traditional expressions, among which some are clearly older, and some are clearly innovative. Here, while the concept of COVER is always expressed by the verb καλύπτω “conceal,” the concepts of DARKNESS (=DEATH) and EYES can be expressed by different lexical items (e.g., νύξ “night” vs. σκότος “darkness,” ὀφθαλμούς “eyes” vs. ὄσσε “two eyes”). See, for instance, the formulaic line τὸν δὲ κατ’ ὀφθαλμῶν ἔρεβεννὴ νύξ ἐκάλυψε “a dark night covered his eyes” (*Il.* 5.659, 13.580, 22.466) vs. the unique expression announcing Sarpedon’s death, which is split over two lines Ὡς ἄρα μιν εἰπόντα τέλος θανάτοιο κάλυπεν / ὀφθαλμούς ῥίνας θ’ “The edge of death covered his eyes and nose as he spoke” (*Il.* 16.502–3). The same conceptual association, this time realized with the archaic dual word form ὄσσε “two eyes” (a direct reflex of PIE **h₃ókʷ-id_h* “id.”), underlies the more archaic-looking formula τὸν δὲ σκότος ὄσσε κάλυψε(ν) “darkness covered his two eyes” (11x in the *Iliad*, all functioning as standard announcements of death) as well as the rare expressions ἀμφὶ δὲ ὄσσε κελαινὴ νύξ ἐκάλυψε(ν) “a dark night

⁷⁶ To be more precise, one could see these all as reflexes of the conceptual association (NEGATIVE EXPERIENCE + SUFFER), since κακὸν “something bad” and ἀεκήλια ἔργα “shameful deeds” are not necessarily physical or psychological pain in all of their readings.

⁷⁷ Various metaphors of death in the *Iliad* have been treated by Horn (2018), within the framework of conceptual metaphor theory (Lakoff and Johnson 1980, Lakoff and Turner 1989). For DEATH IS DARKNESS specifically, see Horn (2018: 368–71). Other recent applications of conceptual metaphor theory to Homer are Forte (2017) and Zanker (2019).

covered his two eyes” (*Il.* 5.310, 11.356, when Aeneas and Hector respectively are nearly killed by a projectile) and τὼ δὲ οἱ ὄσσε / νύξ ἐκάλυψε μέλαινα “a black night covered his two eyes” (*Il.* 14.438–39, when Hector is nearly killed by a stone thrown by Ajax and rescued by his companions). Interestingly, then, the fixed formula is seemingly used for typical business, while the unique expressions based on the traditional conceptual association are made to fit more atypical circumstances (an important character coming close to death but escaping it).

1.4.3 Enter Meter and Syntax: From Collocations to Constructions

So far, we have seen how *conceptual associations* can be expressed as *collocations* in our poems. In Parry’s (1971: 13) terms, we have covered the “repeated group of words” and the “essential idea” in the definition of formula. We haven’t, however, talked about meter (“same metrical conditions”), or syntax (à la Kiparsky). These last two criteria are necessary to describe phenomena that have been classified as *flexible formulas* (as per Hainsworth) or *formulaic expressions* in the Homeric literature so far. Let’s look at one example.

Another important formulaic complex belonging to the conceptual association PAIN + SUFFER is the *collocation* of the stems πηματ- “misery” and παθ- “suffer,” seen in the repeated line δῆμῳ ἔνι Τρώων, ὅθι πάσχετε πῆματ’ Ἀχαιοί “in the land of the Trojans, where you Achaeans suffered misery” (*Od.* 3.100, 4.243, 4.330), as well as, in the *Odyssey*, in the line-final expressions πῆματα πάσχων “suffering misery” (*Od.* 5.33, 17.444, 17.524),⁷⁸ πῆματα πάσχει “he suffers misery” (*Od.* 1.49), πῆματα πάσχειν “to suffer misery” (*Od.* 1.190), πῆματα πάσχεις “you suffer misery” (*Od.* 8.441), and πῆματα πάσχω “I suffer misery” (*Od.* 7.152). These last five expressions clearly belong together, and can be grouped into what Hainsworth would have called a *flexible formula*.

There is a way to notate this more precisely. To do so, we shall make use of the concept of *construction*, loosely derived from Construction Grammar. Within this framework, constructions are defined as “a learned

⁷⁸ Note that line-final expression πῆματα πάσχων “suffering pains” provides a useful metrical alternative to line-final ἄλγεα πάσχων “suffering pains,” as discussed above: the former starts with a consonant, the latter with a vowel. It is not true, however, that every expression combining πηματ- “misery” and παθ- “suffer” simply exists as a metrical alternative to expressions combining ἄλγ- “pain” and παθ- “suffer”: for one thing, the collocation πηματ- + παθ- appears overall later in attestation (it occurs only once in the *Iliad*, in a unique expression, and is otherwise limited to the *Odyssey*), and is significantly more flexible in usage than the ἄλγ- + παθ- collocation.

pairing of form and function” (Goldberg 2006: 4). Construction Grammar holds that, during language acquisition, children learn constructions as generalizations that emerge from encountering expressions that share similarities in form and meaning (Tomasello 2003, 2009: 75–79).⁷⁹ For instance, a child encountering the expressions *more milk*, *more juice*, and *more chocolate*, all sharing the function that they can be employed to ask for more of the item, will make the generalization that one can create expressions to request more food by combining the fixed part *more* + a variable slot containing a noun phrase expressing a food substance. We can notate this generalization as follows:

- (4) **more** [*food substance*]_{Noun Phrase}

This particular construction is made up of a fixed part (bolded) and a variable part (in brackets). Syntactic labels can be added to various parts of the construction as needed. Note that the notation only expresses the form of the construction. The function, here, would be “asking for more food.”

Coming back to Homer and to our examples above, a singer in training will figure out that, following the bucolic diaeresis, one can make a (finite or participial) verb phrase meaning “suffering misery” by combining the fixed sequence $\pi\acute{\eta}\mu\alpha\tau\alpha\ \pi\acute{\alpha}\sigma\chi\text{-}$ with an appropriate morphological ending for the verb (provided this ending corresponds to one heavy syllable). We can notate this generalization, which combines collocational information, syntactic information, and metrical information (something new to the concept of construction, which we need to add for Homer), as follows:

- (5) ^{5a}[$\pi\acute{\eta}\mu\alpha\tau\alpha\ \pi\acute{\alpha}\sigma\chi\text{-}$ -]_{Verb Phrase}

Here the variable slot in the construction is expressed by a metrical symbol (for a heavy syllable), and the brackets are used to encompass the entirety of the verb phrase. Metrical notation (5a) indicates that the construction starts with the first syllable of the fifth foot.⁸⁰ This is a *metrical construction*.

We can add even more material to the mix. In the *Odyssey*, in the same position in the line, we find the unique expressions $\pi\acute{\eta}\mu\alpha\ \pi\alpha\theta\acute{\omicron}\nu\tau\epsilon\varsigma$ “having suffered misery” (*Od.* 12.27) and $\pi\acute{\eta}\mu\alpha\ \pi\acute{\alpha}\theta\eta\sigma\iota$ “he will suffer

⁷⁹ Linguists of different persuasions will accept or deny this characterization of language acquisition. This point is immaterial to our current discussion, in which we simply borrow the concept of construction as a way to notate some generalizations that speakers might make about the linguistic data that they observe, and we apply it to the language of Homer.

⁸⁰ For the metrical terminology (after Janse 2003), see discussion in Chapter 2.

Table 1.10 *From themes to formulas*

	same ideas	same lexical item(s)	same syntax	same meter
Conceptual association (mini-theme)	✓			
Collocation	✓	✓		
Construction	✓	✓	✓	
Metrical construction (formula) ⁸¹	✓	✓	✓	✓
Structural formula (à la Russo)			✓	✓

misery” (*Od.* 7.195), which use the singular of the noun to make room for the trisyllabic forms of the verb (here seen in the aorist stem instead of the present stem). We could, as above, write a construction expressing the commonalities between these two expressions. But there is a larger pattern here: it seems like we are looking at a type of metrically localized collocational paradigm, whereby the collocation $\pi\eta\mu\alpha\tau-$ + $\pi\alpha\theta-$ is localized to a particular slot in the hexameter (5a–6b), and used in a specific syntactic structure (a verb phrase). All of this can be expressed by the following notation:

$$(6) \quad {}^{5a}[\pi\eta\mu\alpha\tau- + \pi\alpha\theta-]^{6b}_{\text{Verb Phrase}}$$

The possibilities covered so far are summarized in Table 1.10.

1.4.4 *From Phrase Constructions to Sentence Constructions*

Constructions exist at different sizes and levels of abstraction. So far, we have seen some small examples, mostly limited to a single syntactic phrase, and with very little variation allowed. These are the smaller chunks in the poets’ repertoire. Poets also had much larger units they could work with, which would help them to structure an entire line, or an even longer run. Many of these types of whole-line constructions exist around the most

⁸¹ In this group we put fixed formulas and (most) flexible formulas alike. Fixed formulas are instances of metrical constructions that recur without variation; flexible formulas are metrical constructions that allow for some variation. Admittedly, some of Hainsworth’s flexible formulas are independent of meter (e.g., in cases of separation and dislocation). These would better be described as collocations, not formulas.

frequent finite verbs in our poems, and have already been described in the literature, starting with Parry's seminal study of noun–epithet formulas and their usage (Parry 1971: 33–55).

For instance, there are 100+ lines in our poems that show formal and functional similarities to the examples below, all of which feature the verb form προσέφη in position 3b–4a (right after the masculine caesura in the third foot), and serve to introduce direct speech:⁸²

- (7) Τὸν δ' ἄρ' ὑπόδρα ἰδὼν προσέφη πόδας ὠκὺς Ἀχιλλεύς. (*Il.* 1.148)
To him, looking darkly, replied swift-footed Achilles.
- (8) Τὴν δὲ βαρὺ στενάχων προσέφη πόδας ὠκὺς Ἀχιλλεύς (*Il.* 1.364)
To her, sighing deeply, replied swift-footed Achilles.
- (9) Τὸν δ' ἐπιμειδήσας προσέφη κρείων Ἀγαμέμνων (*Il.* 4.356)
To him, smiling, replied Lord Agamemnon.
- (10) Τὴν δὲ μέγ' ὀχθήσας προσέφη νεφεληγερέτα Ζεὺς. (*Il.* 1.517)
To her, greatly enraged, replied cloud-gathering Zeus.

A constructional notation that would capture the similarities shared by these examples would be as follows:⁸³

- (11) [–]Object.Pronoun δ' [– – – –] Subject.Participial Phrase προσέφη [– – – –] Subject.Noun Phrase

These types of finite verb constructions have been studied by Bozzone (2014, forthcoming), who set out to establish which speech-introduction constructions appear to be gaining vs. losing ground in the technique as we move from the *Iliad* to the *Odyssey* (see section 1.4.6 below). These are only a particular kind of construction. At the most abstract level, the conceptual association they represent is identical to the argument structure of their main verb (e.g., SUBJECT + REPLY + OBJECT + IN A GIVEN MANNER). They are constructions for entire sentences. As such, they can work as a container for smaller constructions and collocations.

For instance, the participial phrase βαρὺ στενάχων “sighing deeply” seen in (8) is in itself a formula (attested 7x in the *Iliad*, in the line position 1c–3a, but not in the *Odyssey*), as well as an instance of the collocation βαρυ- “deep” + στεναχ- “to sigh,” which is seen in the line-final formula βαρέα στενάχοντα “sighing deeply” (4x in the *Iliad* and 4x in the *Odyssey*) and in the *singulum iteratum* βαρὺ δὲ στενάχοντος

⁸² On speech presentation in the Homeric poems (including speech introductions), see Beck (2012).

⁸³ For the metrical notation, see Chapter 2.

ἄκουσεν “he heard him sighing deeply” (*Od.* 8.95, 534), which is limited to *Odyssey* 8, and always describes Alkínoos heeding Odysseus’ crying. Altogether, they represent the conceptual association SIGH + DEEPLY.

The slot following the finite verb in the construction is taken up by a noun–epithet formula of the metrical shape 4b–6b; many of these formulas can be described as collocations and conceptual associations as well. For instance, the well-known formula πόδας ὠκύς Ἀχιλλεύς “swift-footed Achilles” reflects a more general collocation πεδ- “foot” + ὠκυ- “swift,” which is seen in the formulaic epithet for Iris (πόδας ὠκέα “swift-footed”), as well as in a unique epithet for the made-up hero Orsílokhos (πόδας ὠκύν “swift-footed”), part of Odysseus’ fanciful tale at *Odyssey* 13.260. The same collocation informs the compound adjective ποδῶκης “swift-footed,” used mostly for Achilles and horses. Together, these instances represent the conceptual association FEET + FAST, which occasionally can be realized with other lexical items. See, for instance, the epithet πόδας ταχύς “fast-footed,” used of Achilles (*Il.* 13.348, 17.709, 18.354, 18.358) and of Aeneas (*Il.* 13.482), as well as the metrical construction ^{3c}[ποσὶν ταχέεσσι]_{Dative.Noun Phrase} [διώκ- -]_{Verb Phrase} “to chase with fast feet,” which describes Achilles’ chase in *Iliad* 22 (8, 173, 230), as well as a lion’s in *Iliad* 8.339. This collocation also appears in *Odyssey* 13, right after the usage of πόδας ὠκύν mentioned above:

- (12) Ὀρσίλοχον πόδας ὠκύν, ὃς ἐν Κρήτῃ εὐρείῃ
ἀνέρας ἀλφειστὰς νίκᾳ ταχέεσσι πόδεσσιν (*Od.* 13.260–61)

Swift-footed Orsílokhos, who in vast Crete
defeated enterprising men with his fast feet.

This appears to be a simple example of *working memory priming* (see section 1.4.1): arguably, the usage of πόδας ὠκύν in the preceding line activated the conceptual association FEET + FAST in the working memory of the poet, who then used this association again (with a slight lexical change from ὠκύς to ταχύς to express the concept FAST) in the phrase ταχέεσσι πόδεσσιν.

We could, of course, try to write constructions for larger units as well. One could write constructions for complex sentences (perhaps specifying some embedded clauses) or even for longer stretches of discourse, taking up multiple verses. We know that poets had chunks of this size, which are visible in type scenes (e.g., banquet scenes or arming scenes). A construction for a larger narrative (or theme) would specify the general direction of events and perhaps some key fixed sentences/keywords that need to be uttered for the tale to be told correctly. For the purposes of this chapter, though, we shall stop at the sentence.

1.4.5 Constructions and the Poet's Mind

A well-meaning reader, looking at the algebraic-style notations in the preceding paragraph, might of course ask: do poets really have such objects in their minds? And how does it help us to write them up in this way? The answer comes in two parts.

First, it should go without saying that these are just notational devices. They are meant to represent a likely generalization that a poet in training might make if they were to use our *Iliad* and *Odyssey* as their learning data (given the nature of our data, this is really all we can hope for). In connectionist terms, these notations represent a given activation pattern resulting from the commonalities of many single instances, namely an abstraction or generalization. Among cognitive researchers, opinions differ as to whether these types of generalizations are stored in long-term memory as separate entities, or whether they are created on the spot based on the needs of the moment (e.g., a poet needs to create a new line containing the conceptual association PAIN + SUFFER, and several possible instances are activated in his mind), and are, as such, never independent from the instances they represent.⁸⁴

Second, as with science in general, the value of these models lies in what they can help us discover or explain that has not been noticed or understood before. We have no living singers belonging to the Homeric tradition, so we cannot directly probe what is in the poet's mind. But our theories can make predictions, and predictions can be tested. For instance, a connectionist view of the poet's mind would predict that we should find some priming effects between the elements that form a construction (or collocation, or conceptual association), testifying to their joint activation in the poet's mind. And we do encounter phenomena in the poems that seem to confirm this prediction.

A well-attested collocation in Homer is the combination of the adjective γλυκὺς "sweet" and the noun ἔμερος "desire." This is seen in the unique expression γλυκὺν ἔμερον ἔμβαλε θυμῷ "put a sweet desire in his chest" (*Il.* 3.139), as well as in the formula ὥς σεο νῦν ἔραμαι καὶ με γλυκὺς ἔμερος αἶρεϊ "like I desire you now and a sweet desire takes me" (*Il.* 3.446, 14.328). The latter line is also an example of a more abstract construction pattern,

⁸⁴ These are complex topics in human cognition in general, which also come up in the field of morphological processing and in the computational modeling thereof (e.g., to establish whether morphological rules are stored abstractly or generated on the spot based on stored exemplars that can be recalled online). An example of a rule-based approach is found in Albright and Hayes (2003) (on English past tenses); an example of online-generation of patterns based on stored exemplars is Keuleers (2008) (on English past tenses and Dutch noun plurals).

centered on the verb form αἶρεῖ “takes,” in which the verb takes a noun phrase containing an adjective + a noun expressing an emotion as its subject. Metrically, the verb is at the end of the line, and the subject immediately precedes it. The expression begins after the 3a caesura. Beyond the half-line καὶ με γλυκὺς ἥμερος αἶρεῖ “and a sweet desire takes me,” examples are μάλα γὰρ χλωρὸν δέος αἶρεῖ “for a green fear took (him/her)” (*Il.* 17.67) and μάλα γὰρ δριμύς χόλος αἶρεῖ “for a bitter *khólos* took (him/her)” (*Il.* 18.322). A constructional notation would be as follows:

- (13) ^{3b} — — — [— — — —]_{Subject.Noun Phrase} αἶρεῖ_V = EMOTION + TAKE.OVER

Now, something interesting seems to happen in the following verse:

- (14) σίτου τε γλυκεροῖο περὶ φρένας ἥμερος αἶρεῖ (*Il.* 11.89 = *Homeric Hymn to Apollo* 461)

a desire for sweet food took over his/their *phrénēs*.

There appears to be a sort of modification of the construction above, where the prepositional phrase περὶ φρένας lit. “around the *phrénēs*” replaces the adjective γλυκὺς that normally modifies the noun ἥμερος. Yet, somehow, the strength of the γλυκ- “sweet” + ἥμερο- “desire” collocation is intact: the displaced root γλυκ- “sweet” appears earlier in the line as a modifier of the noun σίτος “food.” Nowhere else in Homer is this word modified by the adjective “sweet,” suggesting that the occurrence of γλυκεροῖο here is likely due to the priming effect of ἥμερος. Thus, the collocation has been preserved, while the syntactic relation between the two items has been changed.

If this is true, it gives us a hint as to how the poet’s verse-making proceeded: here, they probably conceived of the end of the line before the beginning (since arguably the occurrence of γλυκεροῖο in the second foot was triggered by the presence of ἥμερος in the fifth foot). Following this model, the poet would start with a given conceptual association (a mini-theme), which would suggest some collocations, which only later would be constrained within a proper syntactic frame. Of course, this might not be the only way for a verse to come together. This model should also be further developed and then tested. But connectionism provides us with a viable starting point.

The concept of joint activation might also help us explain cases of seemingly odd formulaic usages, such as the known puzzle of Penelope’s “fat hand” (see Parry (1971: 151), references in Edwards (1988: 31–32), and most recently Vergados (2009)). At the beginning of book 21, Athena inspires Penelope to retrieve Odysseus’ bow and put it in front of the suitors (an element that is key to the rest of the plot). As Penelope makes her way to the storage room, she picks up the key for it:

- (15) εἴλετο δὲ κληῖδ' εὐκαμπέα χειρὶ παχείῃ (*Od.* 21.6)

she took a well-curved key with her thick hand.

The usage of the adjective παχείῃ “fat, thick” here has attracted scrutiny, in that it seems like an odd attribute for Penelope. In fact, this verse reflects a formula for the collocation χειρ- “hand” + ἔλ- “take” + παχυ- “fat, thick,” which is common in the *Iliad* and the *Odyssey*.

- (16) ἀλλ' ἀναχασσάμενος λίθον εἴλετο χειρὶ παχείῃ (*Il.* 7.264)

but drawing back he picked up a boulder with his thick hand.

- (17) ἥ δ' ἀναχασσαμένη λίθον εἴλετο χειρὶ παχείῃ (*Il.* 21.403)

and, drawing back, she picked up a boulder with her thick hand.

- (18) δόρυ δ' εἴλετο χειρὶ παχείῃ. (*Il.* 10.31)

and he picked up a spear with his thick hand.

- (19) ὧς ἄρα φωνήσας ξίφος εἴλετο χειρὶ παχείῃ (*Od.* 22.326)

thus he spoke, and he picked up a sword with his thick hand.

This formula is normally used with martial connotations, and the subjects tend to be male and strong. Is the usage here in *Odyssey* 21 simply awkward, or consciously humorous? Perhaps. There is, however, one exception to the generalization above: in example (17), the formula refers to Athena, as she picks up a boulder to use as a weapon against Ares. The attribution here seems unobjectionable. So what could explain the odd usage in *Odyssey* 21? In a connectionist model, we could think about which elements in the passage could have conspired to “activate” the expression χειρὶ παχείῃ “with a thick hand” in the poet’s mind. We could envision the spread of activation in two ways: the context (the preparation for what will ultimately become a fight) brought up an εἴλετο construction which is normally used for arming scenes (this is in line with Foley’s argument that Penelope here is entering a heroic mode). This εἴλετο construction, in turn, combined with the recent mention of Athena, brought up the prepositional phrase χειρὶ παχείῃ. In other words, an attribute that would be appropriate for Athena in this construction was contextually reassigned to Penelope, just like (in example (14) above) an attribute of “desire” was contextually reassigned to “food” (cf. Foley 1999: 202–21, contra Wyatt 1978).

While the last two examples may look like “errors” in the workings of oral composition, the spread of activation through the network

actually has the fundamental role of contributing to discourse cohesion: it helps the text hold together, fulfilling the audiences' expectations.

1.4.6 *Formulas and Diachrony*

A viewpoint that might not interest most readers of Homer directly, but might have an important role in answering the Homeric question, is: to what extent can we use formulas as a window onto the history of the poetic tradition? We have mentioned before the idea that formulas (conceptual associations) can undergo lexical renewal. We are also familiar with the fact that formulas can sometimes preserve very old linguistic features, thus offering us a glance at what could be chunks of poetry that are hundreds of years old (we will discuss this more in Chapters 2 and 3).

Just like our own native language, Homer's trove of expressions is composed of a mixture of very old and very new material. How can we tell archaic expressions apart from innovative ones? The classical method is that of checking whether an expression happens to preserve a clear archaism that is guaranteed by the meter (several examples will be discussed in Chapter 2). This method, however, will only work on a handful of truly old expressions, thus helping us identify only a small subset of everything that is actually old in the language. Another method (Bozzone 2014, 2022) is that of using the *flexibility* of an expression to gauge its antiquity. In general terms, truly archaic expressions tend to survive only in fixed forms, while newer, living expressions tend to display flexibility. This has to do with Kiparsky's dichotomy discussed above in section 1.1.3: what is retrieved from memory as such (e.g., *fixed formulas*) tends to be unchangeable, while what is still actively generated (e.g., *flexible formulas*) can change. If an expression reflects an older stage of the grammar (one that moreover would be at odds with the synchronic grammar of the poet), it is likely to be pulled from memory as a chunk.

Another way to express this concept is that young expressions have company (in the form of other, similar expressions created by the synchronic grammar), while older expressions do not (they are, in other words, the lone survivors of an earlier era). We can thus look at a given expression and its "family" to establish whether it is isolated or not, and then make inferences as to whether it is old or new in the technique. While more precise quantitative measures can be employed to this effect (see discussion in Bozzone 2022 and forthcoming), approximation will often be sufficient. Let us look at some examples.

Bozzone (2010, 2016b) uses the example of two equivalent noun–epithet formulas for Hera, θεὰ λευκώλενος Ἥρη “white-armed goddess Hera” and βοῶπις πότνια Ἥρη “cow-eyed queen Hera,” in which the flexibility of each expression and their combinatory possibilities clearly identify one as archaic and fossilized (the latter) and one as more recent and still alive and well in the language (the former).⁸⁵ This analysis is confirmed, on the linguistic level, by two archaisms that are preserved in βοῶπις πότνια Ἥρη, namely the hiatus between πότνια and Ἥρη (which would have originated after the lenition of initial *s- in the word for Hera) and the apparent violation of Wernicke’s law in the last syllable of βοῶπις (which would have been absent at an earlier stage of the language).⁸⁶

We can use this method in constructions other than noun–epithet formulas, to verify whether an expression was alive in the poet’s language or not. For instance, there are two similar ways, in the *Iliad*, to figuratively announce the death of a warrior.⁸⁷ The first set of expressions reflects the conceptual association TAKE + EARTH + WITH TEETH, the second reflects the association TAKE + EARTH + WITH PALM. Formulas reflecting these conceptual associations are as follows:

- (20) οἱ μὲν ἔπειθ’ ἅμα πάντες ὀδᾶς ἔλον ἄσπετον οὐδᾶς (*Od.* 22.269)

and then they all took the infinite earth with their teeth.

- (21) ὃ δ’ ἐν κονίῃσι πεσὼν ἔλε γαῖαν ἀγοστῶ. (*Il.* 13.508)

and falling in the dust he took the earth in his palm.

While, on the surface, these two expressions might appear similar to one another (one seems specialized for the plural, one for the singular), a closer inspection reveals that one is a fossil, and the other one is part of the living language of the poet. The expression ἔλε γαῖαν ἀγοστῶ “s/he took the earth with his/her palm” is relatively high frequency (5x in the *Iliad*), and never displays any flexibility. It also always occurs within the same type of verse construction, with a bisyllabic finite verb starting the line (in enjambement), followed by a syntactic break:

⁸⁵ These two equivalent noun–epithet formulas have received much attention in the literature. A history of the debate is given in Beck (1986). See also more recently Beck (2005: 129–30).

⁸⁶ Wernicke’s law is a dispreference for a syllable of the shape CVC in the contracted biceps of the fourth foot (see Chapter 2 for this terminology), when the biceps is filled by the last syllable of a word (conversely, a sequence CVC is preferred). βοῶπις, with short -ι-, violates this law, but an earlier < *βοώπις (with long vowel resulting from PIE *-ih₂-) does not. See Cassio (2016b).

⁸⁷ For similar metonymic descriptions of death in the *Iliad*, see Horn (2018: 363–68).

- (22) ῥηφυσ'· ὃ δ' ἐν κονίησι πεσὼν ἔλε γαῖαν ἀγοστῶ. (*Il.* 13.508)
[the bronze] pulled out [his innards]. And having fallen in the dust he took the earth with his palm.
- (23) ῥηφυσ'· ὃ δ' ἐν κονίησι πεσὼν ἔλε γαῖαν ἀγοστῶ. (*Il.* 17.315)
[the bronze] pulled out [his innards]. And having fallen in the dust he took the earth with his palm.
- (24) ἔσχεν· ὃ δ' ἐν κονίησι πεσὼν ἔλε γαῖαν ἀγοστῶ. (*Il.* 13.520)
[the heavy spear] pierced him [through his shoulder]. And having fallen in the dust he took the earth with his palm.
- (25) ἔσχεν· ὃ δ' ἐν κονίησι πεσὼν ἔλε γαῖαν ἀγοστῶ. (*Il.* 14.452)
[the heavy spear] pierced him [through his shoulder]. And having fallen in the dust he took the earth with his palm.
- (26) νύξεν· ὃ δ' ἐν κονίησι πεσὼν ἔλε γαῖαν ἀγοστῶ (*Il.* 11.425)
[he] hit him. And having fallen in the dust he took the earth with his palm.

There are no other occurrences of the lexical item ἀγοστός in the poems (in fact, its meaning, “palm of the hand,” is entirely inferred from these occurrences; in Theocritus, it is used with the meaning “arm”). This conceptual association, in other words, only has one fixed surface realization.

On the other hand, the conceptual association TAKE + EARTH + WITH TEETH knows many incarnations. Next to the more regulated formulaic usages (which are common to the *Iliad* and the *Odyssey*), for example:

- (27) οἱ μὲν ἔπειθ' ἅμα πάντες ὀδᾶξ ἔλον ἄσπετον οὐδας (*Od.* 22.269)
and then they all took the infinite earth with their teeth.
- (28) τὼ κ' οὐ τόσσοι Ἀχαιοὶ ὀδᾶξ ἔλον ἄσπετον οὐδας (*Il.* 19.61)
then not so many Achaeans would have taken the infinite earth with their teeth.
- (29) Ἑκτορος ἐν παλάμῃσιν ὀδᾶξ ἔλον ἄσπετον οὐδας. (*Il.* 24.738)
at the hands of Hector they took the infinite earth with their teeth.

one also finds simple collocations (indicated with wavy underlining):

- (30) φῶτες ὀδᾶξ ἔλον οὐδας ἐμῶ ὑπὸ δουρὶ δαμέντες. (*Il.* 11.749)
the men took the earth with their teeth, tamed by my spear.

as well as combinations in which EARTH and TAKE are expressed by different lexical items:

- (31) πρηνέες ἐν κονίῃσιν ὁδᾶξ λαζοίατο γαῖαν. (*Il.* 2.418)
face-first in the dust, they seized the earth with their teeth.
- (32) γαῖαν ὁδᾶξ εἶλον πρὶν Ἴλιον εἰσαφικέσθαι. (*Il.* 22.17)
they would have taken the earth with their teeth before ever making it back to Ilion.

From these distributional properties alone, we can safely infer that the conceptual association TAKE + EARTH + WITH TEETH was still lively in the poet's language, while TAKE + EARTH + WITH PALM was an isolated fossil. As confirmation, the expression γαῖαν δ' ὁδᾶξ ἐλόντες "having taken the earth with their teeth" is still found in Euripides (*Phoenissae* 1423), while TAKE + EARTH + WITH PALM simply disappears from the later record.

Bozzone (2014, 2022, forthcoming) proposed to use the concept of *linguistic productivity*⁸⁸ as a way to assign each formulaic expression in Homer to a different "life stage" within the diction. The idea is that each expression goes through a life cycle that is marked by productivity (i.e., flexibility) changes, and that productivity measures can help us establish the relative "age" of an expression. Furthermore, looking at how the productivity of given expressions changes between two texts (e.g., the *Iliad* and the *Odyssey*) can help us to create a relative chronology of Greek epic. For instance, Bozzone (forthcoming) looks at speech-introduction constructions in the poems, and shows that all constructions seem to "age" between the *Iliad* and the *Odyssey* (while new constructions are introduced as well), which agrees with the general consensus that the *Odyssey* was composed at a later point in time (for a recent discussion, see Andersen and Haug 2012).

1.5 Conclusion: What Are Formulas and What Can We Do with Them?

The discussion above has covered much ground. Following a review of the history of the study of formularity in Homer (section 1.1), sections 1.2 and 1.3 argued that formularity (in the broad sense) is a general and widespread feature of human language and cognition, ultimately rooted in the limitations of our working memory and in the strategies that our mind adopts in order to overcome these limitations. Homeric formularity is, then, just a special case within this general tendency to rely on *chunks* when carrying out a cognitively demanding task.

⁸⁸ For an introduction to productivity in morphology, see Bauer (2001). For the role of morphological productivity in historical linguistics, see Sandell (2015: 8–32).

While the overall reliance on chunks in Homer is similar to what happens in normal language processing, the extent to which Homer relies on *large linguistic chunks* (i.e., stored verbal sequences that are more than two words long) does seem to set him apart from ancient prose authors like Herodotus, literate hexametric poets like Quintus Smyrnaeus, and modern corpora of spoken and written English. Rather than constituting direct proof of orality of composition, I have argued that this phenomenon points to a high level of *mastery* on the part of the poet, and specifically to the accumulation, likely over the course of a long period of training, of many automated behaviors (i.e., chunks) of increasing size that can support the task of composition. Within the landscape of Archaic Greece, it is persuasive that the conditions that would make such mastery necessary or desirable would only arise within the context of an oral tradition.

Section 1.4 of this chapter has sketched a general theory of formularity in Homer, rooted in cognitive and linguistic frameworks, and has proposed a fine-grained terminology for distinguishing different types of formulaic phenomena in Homer, moving from the very abstract (themes and conceptual associations) to the very concrete (syntactic and metrical constructions – i.e., traditional Parry’s formulas). While these definitions have been tailored to the Homeric poems, they should also be applicable to other oral traditions (as well as literary and nonliterary texts in general).

Going forward, one might ask what else can be done after a given formulaic phenomenon has been identified as such – that is, what do we do after we have described a construction, conceptual association, or simple collocation in Homer, and perhaps provided some formal notation for it. For this task, we can take inspiration from the practice of lexicography and literary analysis.

The first step, arguably, would be to establish what the *meaning* and *narrative function* of that phenomenon are. Here, Foley’s suggestion that formulas in oral traditions should be regarded as “bigger words” can be employed fruitfully (Foley 2002: 14).⁸⁹ In order to establish the meaning of a word (or, in our case, of a formulaic phenomenon) in a closed corpus, one typically starts by collecting and studying all occurrences of that word in the corpus.⁹⁰ The same can be done for a formulaic phenomenon in Homer, in order to establish its basic meaning as well as its *traditional referentiality*.⁹¹ Let us say we are studying the word *cat* in a closed corpus: in

⁸⁹ See Foley (2002: 18): “if the guslar thinks and composes in terms of *reči*, then we must strive to listen and read in terms of *reči*.” *Reči* is the plural of Serbo-Croatian *reč* “word, traditional word”, discussed in fn. 49.

⁹⁰ This is in line with J. R. Firth’s (1957: 11) maxim: “You shall know a word by the company it keeps!”

⁹¹ In Foley’s (1999) terms, this would mean: what basic traditional associations does the word evoke for an audience that is thoroughly familiar with the specific tradition?

this first step, we would collect all of the occurrences, and describe in general what the basic meaning of the word in the corpus appears to be.

Second, we would study individual usages of our “word”/formulaic phenomenon, in order to see how it works specifically in a given passage. There are at least two aspects to this study: first, we might want to see how the specific context of usage in a given passage selects a specific aspect of the semantic and referential potential of the formulaic phenomenon/“word.” In our study of the word *cat*, we might find that the word denotes something different when used in a discussion of house pets vs. in a description of sub-Saharan mammals. Similarly, the same formulaic phenomenon in Homer might express a different meaning when used in different contexts (put another way, the context of usage will select or suggest a specific reading, and a specific referentiality, within all the ones that are possible).

Finally, we might employ some basic techniques of literary analysis in order to ask whether that “word”/formulaic phenomenon is being employed for some special effect (e.g., for intertextual referentiality),⁹² or whether it is tied to a specific narrative or stylistic function. Is the word *cat* an important keyword in our text, does it take part in foreshadowing or does it appear to refer to another passage or text? The same can be asked for a formulaic phenomenon.

After these primary facts have been established, we can ask some further questions concerning the status of this item within the poetic technique: is this “word”/formulaic phenomenon traditional or innovative? Does it belong to an archaic or recent layer of the technique? Can we trace its evolution over the course of time? Some examples of how to answer these questions have been sketched in section 1.4.6 above.

⁹² For the concept of *intraformularity* – i.e., the capacity of formulas to refer to other specific passages in a song or to another song altogether – see Bakker (2013: 157–69).