

Roohollah Mofidi* and Peter Petré

Aspectual marking from a typologically uncommon origin: a quantitative account of the development of *hamē(w)* in Middle Persian

<https://doi.org/10.1515/flin-2022-2025>

Received January 8, 2021; accepted May 27, 2021; published online June 22, 2022

Abstract: This article follows the early grammaticalization path of a durative adverb to an aspect marker for a range of aspectual uses including durative, focalized, and habitual. In a usage-based approach, we investigate the behavior of *hamē(w)* in a Middle Persian database with respect to several grammatical variables, and we check its developmental path against the model proposed for the Romance languages by Bertinetto, Ebert and de Groot, aiming to uncover the amount of similarities/differences between those Romance languages and Persian. The results show that the lexical origin of the structure investigated locates Middle Persian in stage (iii) of the model, and the development documented deviates from the rest of the model in various respects. Furthermore, the usual priority of past tense in progressive developments is not observed: the adverb is originally present-oriented, but past tense usage increases as the grammaticalization proceeds.

Keywords: adverb; durative; grammaticalization; Middle Persian; progressive aspect

1 Introduction

The current paper analyses the early development of *hamē(w)*¹ in Middle Persian from an adverb meaning ‘always, continually’ into an aspectual marker of

1 In the texts and studies of Middle Persian, ‘front’ and ‘back’ low vowels are traditionally transcribed with the symbols *a* and *ā*, respectively. This is in line with other long vowels which are represented with the same ‘length’ diacritic: *ō*, *ū*, *ī*, and *ē*. Therefore, *hamē* is to be roughly shown as [hæme:] in IPA system, regardless of the exact phonetic characteristics of the first vowel.

***Corresponding author: Roohollah Mofidi**, English Department, Imam Khomeini International University, Qazvin, Iran, E-mail: mofidi@hum.ikiu.ac.ir. <https://orcid.org/0000-0002-9966-0580>
Peter Petré, Department of Linguistics, University of Antwerp, Antwerp, Belgium, E-mail: peter.petre@uantwerpen.be. <https://orcid.org/0000-0002-6503-1380>

(focalized) progressive and habitual aspect. After this period *hamē(w)* continued to develop into a quasi-general imperfectivizing prefix *mi-*, and has become a key element of Present-Day Persian (Farsi) grammar.

Grammatical markers of aspect are a core feature of many languages worldwide. It comes as no surprise, therefore, that they have been extensively studied from various perspectives, including, among others, typological, corpus-based, and historical ones. Not all languages have been looked at from all of these perspectives and with the same amount of detail, though. Specifically, usage-based historical work is actually quite uncommon for non-European languages. Also, Heine and Kuteva (2007: 87) claim adverbs to be “fairly uncommon as sources for verbal aspect markers”. They refer to American Sign Language (ASL) for an example of one such “occasional grammaticalizations” to express habituality.² In this paper we take a usage-based approach to one such case of an adverb developing into a marker of progressive as well as habitual aspect, in a non-European language. Our two main objectives are: (i) to test the historical and typological claims made by Bertinetto et al. (2000), when they discuss the diachronic development of progressive constructions in Romance, and to apply them to a non-European form with a different lexical origin; (ii) to provide a first quantitative account of the early grammaticalization of *hamē(w)* during Middle Persian, which has so far only been considered from a more traditional philological perspective.

Starting from a survey of a large number of European languages, as well as some non-European ones, Bertinetto et al. (2000) formulate what may be seen as a cluster of typological implicatures that appear to hold for most of the languages in their sample. They summarize their findings in a number of pathways of development. We focus on two, which recur in our case study, and we show that the pathways are not always the same. The differences may be accounted for by considering the different lexical origin of *hamē(w)*. More broadly, our analysis provides evidence for a more diversified range of grammaticalization paths towards imperfective marking than has been found so far.

Progressive aspect typically originates from a verb denoting location or position, or a related locative construction (Bertinetto et al.’s stage i), which have a literal meaning of ‘be at something’ or ‘in some state’. Out of this lexical origin, which has stative/durative function, a use as a general durative marker of the event encoded by the verb develops (their stages ii and iii), and further on, a focalized

² An anonymous reviewer, admitting the rarity of imperfective development out of adverbs, mentioned that the development of adverbs/adverbial elements to markers of telicity (and perfectivity) is not that uncommon. In fact, such a phenomenon is found in Early New Persian: *be-/bi-* is argued to be a perfective aspect marker, developed out of an adverb or a verbal particle (see Section 2; cf. Brunner 1977: 157–162; Windfuhr 1979: 94–96; also cf. MacKinnon 1977 for a comparison to Slavic and other Indo-European groups).

progressive usage ('be in the middle of something at a specific point in time', stage iv) and, finally, generally imperfective aspectual uses (stage V), in this particular order. One of the properties of the final stage of general imperfectivity they mention is the introduction of habitual uses. According to their theory, therefore, the order of acquisition of aspectual uses for *hamē(w)* is expected to be as follows (with the stages numbered within brackets):

- (1) (lexical use as adverb of duration) [i] > durative [ii & iii] > focalized [iv] > habitual (as part of 'pure imperfectivity') [v]

While some follow-up studies maintain that the situation may have been more complex for some languages (e.g. Killie 2008 on Old English), the overall picture drawn for stages iii to v (as represented in [1]) still seems to stand. The case of Middle Persian, however, differs from this model, in that the aspectual marker does not have a locative origin. More generally, progressives are "usually periphrastic in expression" (Bybee et al. 1994: 174), giving Persian an 'uncommon' status for its adverbial source of durative. What makes Persian similar to the Romance languages in Bertinetto (2000) and Bertinetto et al. (2000) is the initial status of the duratives; but it is as if Persian has started the whole path from stage iii in the model. Moreover, the available evidence casts doubts on the assumption that habituals come later in the grammaticalization process than focalized uses, which may again be explained on the basis of the original lexical meaning.

A second generalization Bertinetto et al. (2000) make relates to the tense usage of these aspectual markers: "on a broad typological scale, several languages allow a progressive/non-progressive opposition in the past tenses, but not in the present" (2000: 526). The implicit assumption that seems to be made here is that the development of specifically progressive aspectual functions is expected to proceed first in the past tense, and only then in the present tense (which for instance appears to be borne out for English, see e.g. Petré 2016):

- (2) past > past + present

Again, as we will see, this observation is not borne out by our data.

The second objective of this paper is to reconcile the unexpected findings regarding *hamē(w)* with what is expected from the perspective of grammaticalization theory. More broadly, the development of aspectual markers is a subcategory of the phenomenon of grammaticalization, the emergence of more grammatical, procedural functions out of originally lexical material, given a specific context or construction (cf. Hopper and Traugott 2003: 18, among others). A number of recurrent properties of grammaticalization have been described. The pathway of development seen in European languages by Bertinetto et al. (2000) seems to be well accountable in terms of some core concepts of

grammaticalization theory, such as that of bleaching or its more positive counterpart of semantic enrichment and context expansion (in this case from compatibility with activities to compatibility with [more punctual] achievements and accomplishments as well as, eventually, states).

A challenge specifically related to the case of Middle Persian is the scarceness of well-dated manuscripts that are temporally not far removed from their sources. Instead, many manuscripts only survive in rather late copies (sometimes several centuries later), and precise dating is challenging. This is of course not unlike the early stages of other languages. To avoid confirmation bias when faced with different dating options, we will take a minimalistic stance on dating in this paper. We will tentatively treat the distinction between the forms *hamēw* and *hamē* as reflecting a difference between early and late, as this is also in line with the evidence from external history (even though there is also a dialectal difference involved). In addition, we will treat the diagnostic of adjacency of the aspectual marker to the verb, closely related to the concept of ‘bondedness’ from grammaticalization theory, as indicative of degree of grammaticalization (cf. Lehmann 2015: 157–167). While this does not necessarily tell us much about relative dating (certain texts may be in a more archaic style, or the genre may bias the results towards more lexical uses), it should tell us something about logical order, which should be helpful in reconstructing the overall grammaticalization path.

The structure of our paper is as follows: Section 2 provides a sketch of the periods of Persian history, mainly concentrating on the Old Persian aspectual system and the changes leading to Middle Persian. Section 3 introduces some previous studies on the current topic. Our methodology and use of corpus data is explained in Section 4, which then continues with the analysis of the data on *hamē(w)* with respect to four factors: spelling, position, tense, and event type. Section 5, then, is devoted to interpreting the results in light of grammaticalization theory. Finally, Section 6 summarizes our findings, including the observation that *hamē* and *hamēw* should be treated separately. Our results show that *hamē*, but not *hamēw*, has started on its way to grammaticalization. Evidence is found in its position, which is generally adjacent to the verb, and the development of habitual and focalized functions beyond its original durative meaning, as well as an increased use in the past tense. While the observed grammaticalization path shares the priority of durative aspect with the findings in Bertinetto et al. (2000), it also differs from those in non-trivial ways: Middle Persian does not seem to have developed the focalized aspect before the habitual, and, possibly due to the adverbial origin of the grammaticalizing element, the initial orientation is towards the present tense rather than the past.

2 From Old to Middle Persian: development of the aspectual system

This section provides a brief outline of the history of Persian with particular attention to its aspectual system, by way of background information to the current case study. Persian is a language belonging to the Indo-Iranian branch of the Indo-European family. Both Iranian as a sub-branch and Persian as a language are traditionally divided into an Old, Middle and New period. The fall of the Achaemenian dynasty at the hands of Alexander III of Macedon (330 B.C.) marks the end of the Old Iranian Period, and the defeat of the Sasanian dynasty by Muslim Arabs (650 A.D.) that of the Middle Iranian Period. The earliest data available from Middle Persian are some inscriptions and what is known as the Manichaean texts. The oldest of both of these dates back to the third century A.D. The bulk of the data, however, comes from texts ranging from the sixth to tenth centuries, the latest ones probably being written when the language was no longer alive (Sundermann 1989: 138). Also, the surviving manuscripts of these texts are often of a much later date and not older than the fourteenth century, and some of them exhibit some linguistic features not common in Middle Persian (both lexically and grammatically), probably under the influence of the New Persian variety spoken by the scribe (see Tafazzoli 1997 for details). New Persian, itself, may have developed as early as 750 A.D. (Lazard 1989: 263), and, while it has obviously changed, is taken to last to the present day.

Old Persian was a richly inflected language, both in the nominal and verbal domains. These inflections, along with an extensive ablaut system, expressed a wide variety of nominal and verbal categories. As far as the verb is concerned, tense and mood were expressed by endings and augment, while aspect was expressed lexically by different stems of the verb (Skjærvø 2009: 86; also cf. Dahl 2011a, 2011b for detailed information on the functioning mechanism of these alternating stems in Indo-Iranian, including Vedic, Avestan, and Old Persian). The result was an aspecto-temporal system with a triple contrast of present (imperfect) versus aorist versus perfect, with innovations in its Indo-European inherited system, which made it different from older patterns found in Old Avestan, Vedic Sanskrit, and Ancient Greek (cf. Bubenik 1997 for these older patterns; Cowgill 1968 and Schmitt 1989: 76–80 for the innovations). In fact, Old Persian is recognized as being already well advanced toward the Middle Iranian state (Cowgill 1968: 259).

In its transition to Middle Persian, Old Persian went through a series of phonological changes and reductions (Maggi and Orsatti 2018: 19–20). These changes (especially the loss of final vowels and codas) have been assumed to be responsible for the loss of many forms and categories in the morphological system of Old Persian (Skjærvø 2009; Sundermann 1989). Related to this loss, analogical

levelling of surviving forms and morphological patterns also occurred (for an example in the verbal domain, cf. Sundermann 1989: 148–149). The overall result was that the old aspecto-temporal oppositions were lost, and Middle Persian only retained a present stem, which could be inflected with two paradigms of agreement markers (for distinguishing mood), and a past stem used in the participial form combining with inflected auxiliaries. At this stage, “the present indicative is the ‘unmarked tense’ and is used for events taking place in the present or soon in the future and for general statements” (Skjærvø 2009: 229).³ This aspectual ambiguity in the present tense along with the absence of a grammaticalized perfective-imperfective distinction in the past tense⁴ led to a generally underdetermined aspectual system. This situation provided an opportunity for a number of innovative aspectual uses of adverbs and particles appearing in the grammar of Middle Persian, including (i) the recruitment of the adverb *hamē(w)* ‘always, continually’ for durative aspect; and (ii) the use of the particle *bē* ‘out, away’ to express singularity, possibly perfective aspect (Skjærvø 2009: 240–241), as well as probably the beginnings of future and subjunctive-imperative marking (Josephson 2013; Jügel 2013).

3 Previous accounts of *hamē(w)*

The category of *hamē(w)* in Middle Persian, either as an adverb or a verbal particle, has been mentioned in several studies on the issue, with suggested original meanings as ‘always, continually, forever’. It has been proposed that *hamē(w)* comes from the reconstructed nominal **hama-aiwa-* ‘same duration, time’ in Old Iranian (cf. Bubenik 2019: 200 for a comparison with Sanskrit and Old Cypriot; Josephson 2016: 49 for information on the pioneers of the idea). Brunner (1977: 167) maintains that even among older, inscriptional examples, there are several examples which “seem to function as a particle indicating continuing action in the past”. In later works, in his view, “*hamē* continues to occur as an adverb [...]”. But more often it serves as a particle indicating the durative aspect of the action state” (Brunner 1977: 167).

³ An anonymous reviewer kindly reminded us that such an unmarked form is not an innovation of Middle Persian, since similar functions are associated with the present indicative already in Indo-Iranian. Dahl (2011a: 267) suggests that the present stem in Indo-Iranian had “a neutral aspectual meaning”, being “the unmarked or default category in the tense/aspect system”. More precisely, “the Present Indicative strongly tends to assume either a progressive or a habitual meaning” (Dahl 2011a: 275), and it is also compatible with future time reference, though there are several forms which can have futurate meanings (cf. Dahl 2011b for the future in Indo-Iranian).

⁴ We ignore here an unproductive imperfect inherited from Old Persian that only survived in early Middle Persian inscriptions (cf. Brunner 1977: 218).

The change in the original meaning of *hamē(w)* in preverbal environments has been described by some other studies or comments as well. Gignoux (1969) speaks of a durative, progressive, iterative meaning; Nyberg (1974: II/91) of perduration or iteration; Boyce (1977: 46) of continuous meaning; Skjærvø (2009: 239) of ongoing, progressive state or action (though still calling it an adverb); Bubenik and Ziamajidi (2018: 77) of “the notion of duration in the past”.

Josephson (2016) is the only work that provides a more systematic chronology of the semantics and contextual usage of *hamē(w)*. She takes *Wizīdagihā ī Zādspram* as belonging to the third quarter of the ninth century, based on some historical evidence, and uses the characteristics of *hamē* in this text as the criterion to classify similar texts (between 850 and 1000 A.D.) as ‘late Pahlavi’, as opposed to earlier texts in which *hamē* displays some older features. She believes that “in the earliest examples the adverb *hamē* in the sense of ‘continually’ interacts with the situational type or actionality expressed by the verb. It combines mainly with activity verbs ... It is not compatible with non-durative situations. At this stage its position is not fixed but it occurs frequently before the verb” (Josephson 2016: 61, with our emphasis on ‘frequently’).

She has observed that in these texts, *hamē* collocates with the present tense, and the examples of past tense occurrence are not frequent, concluding that “it is hard to judge whether this indicates that *hamē* + past tense spread from usage of the adverb with the present tense or if it simply means that such [narrative] texts are not extant” (Josephson 2016: 53).

In ‘late Pahlavi’, on the other hand, “*hamē* occurs regularly before the verb. It has lost some of its specific meaning and become more general” (Josephson 2016: 62, with our emphasis on ‘regularly’). Another change in the late usage is that it occurs in narrative texts to indicate background actions, and with achievement verbs to show that the action is going on at the moment. Josephson concludes that “[t]his usage belongs to the realm of imperfective aspect [...] [and] it sometimes appears to have become an aspectual particle” (Josephson 2016: 62). In her view, this tendency for being used in the narratives is confirmed by *Ardā Wīrāz Nāmag*, which represents a ‘post-Pahlavi’ stage, later than 1000 A.D. (referring to Gignoux 1969: 998–1002).

4 Data analysis: the changing behavior of *hamē(w)* in Middle Persian

The focus of this section is the changing distribution of *hamē(w)*. Section 4.1 first describes the data that were used. In the subsequent sections, four variables are discussed: structural position and form variation (Section 4.2), interaction with the category of tense in the clause (Section 4.3), and changes in the event type of the

verbs with which *hamē(w)* co-occurs (Section 4.4). These sections focus on the three major aspectual uses we find for *hamē(w)*, i.e. durative, focalized, and habitual. Their assignment is based on the observation that all instances of *hamē(w)* can essentially be treated as indicating that the event has some duration (durative aspect, ‘continually’), is repeated (habitual aspect, here broadly defined as also including iterative, frequentative and generic instances), or is ongoing at some reference time (progressive). Put differently, Sections 4.2 to 4.4 treat *hamē(w)* as a kind of polysemous particle single-handedly modifying the aspect of the clause. This is, however, a simplification, as the interpretation is a complex interplay of *hamē(w)*, the verb, and its context. Section 4.5 will attempt to propose a more general semantic analysis of the various functions of *hamē(w)*, which does justice to the holistic nature of aspectual semantics, and tries to integrate the meaning retention/shifts throughout the data, making use of the concept of pluractionality.

4.1 Data and methodology

Our source of data is the text database of Middle Persian available from TITUS (Gippert et al. 2016). At the time of our search, it included 11 longer texts plus some short ‘Pahlavi’ texts and a series of ‘Manichaean’ texts in this language. We searched for the words *hamēw* and *hamē* separately (in ‘all language varieties’, ‘all available texts’, and by ‘exact match’). Then, we exported all sentences from the Query Result Table into an excel file, where they were subjected to further analysis.

In the entire Middle Persian data, we found 273 instances of *hamē(w)*, with 16 being spelt as *hamēw* and the remaining 257 as *hamē*. The two main spellings of *hamē(w)* are not randomly distributed. No text includes both forms simultaneously. Moreover, all instances of *hamēw* occur in *Mēnōg ī xrad* and Manichaean texts. The difference may be accounted for either in terms of dialect differences or in terms of relative chronology. First, the two forms might be considered to reflect a dialectal alternation. If this holds true, then the lack of grammaticalization of *hamēw* shown in our results (cf. the following sections) means that in the relevant dialect no grammaticalization process took place. As already mentioned, in our corpus, *hamēw* is found in *Mēnōg ī xrad* and the Manichaean texts. We can add two more sources including this form from outside the corpus:

- (i) Manichaean texts written in Parthian (another Western Middle Iranian language), where we find at least five instances of *hamēw* (and only this form). Three of them are available in the Parthian part of the TITUS-corpus, and two are reported in Brunner (1977: 169). All the Manichaean manuscripts in both languages were found in Chinese Central Asia (cf. Skjærvø 2009: 196), which is essentially the most important pointer towards a dialect difference.

- (ii) However, five instances of (again exclusively) *hamēw* altogether were also found in at least two Sasanian inscriptions of Middle Persian on cliffs in Southern Iran, dating back to the late third and mid-fourth century (cf. Brunner 1977: 167–168 for four of them). While some amount of inter-dialectal exchange cannot be excluded, linking sources of such diverse geographic origin to a single dialectal origin seems far-fetched, though we cannot entirely reject the hypothesis.

Given what we know about the dating of the respective texts, and the Iranian location of the cliff inscriptions, the second hypothesis is more likely: *hamēw* is an older form than *hamē*, suggesting the phonological change *hamēw* > *hamē*. Some dating estimations and classifications proposed in previous studies can be adopted as supporting evidence for this hypothesis. For example, Manichaean texts have been suggested to be classified with the inscriptions as ‘Classical Middle Persian’ as opposed to ‘Late Middle Persian’, which includes most other texts (see Josephson 2013). Tafazzoli (1997: 198) concludes from some pieces of historical and linguistic evidence that *Mēnōg ī Xrad* has probably been written as early as the sixth century. Our working hypothesis therefore will be that the texts using the spelling *hamēw* are older than the nine texts that use the spelling *hamē*.

4.2 Position and spelling of *hamē(w)*

On the basis of evidence from the external history, we assume that *hamēw* is older than *hamē*. In line with Lehmann’s (2015) bondedness hypothesis, we also assume that adjacent use correlates with more grammaticalized use. In this section we take a first look at these two factors and their interplay with the three major aspectual uses we find for *hamē(w)*, i.e. durative, focalized, and habitual. Their assignment is based on the observation that all instances of *hamē(w)* can essentially be translated in one of three possible ways: ‘continually’, ‘always’, or (in combination with the main verb) by the English progressive construction.

We will first turn to the role played by the position of *hamē(w)* vis-à-vis the verb. Table 1 provides a classification of the relevant configurations.

Our main concern is with the first two groups, which represent the majority of uses (243 instances, which amounts to 89% of all instances), and where *hamē(w)* appears to be modifying the event structure at the clausal level. Group 3, including 20 instances, is a specialized use consisting of the fixed combination *hamē(w) ka*. *Ka* is a common temporal subordinator, and the combination *hamē(w) ka* is translated as corresponding to the meaning ‘whenever’ (MacKenzie 1971: 40), or, more literally, ‘always when’ (Nyberg 1974/II: 91). Whether the combination is a

Table 1: Function + position of *hamē(w)*.

Group	Function	Position with regard to verb	Example
1	Aspectual (durative, habitual, focalized)	Adjacent	<i>ka pad šab gyāg-ēw hamē šawēd</i> 'If one is going somewhere at night ...' (<i>Šāyist</i> , chapter 2, paragraph 72)
2	Aspectual (durative, [habitual])	Non-Adjacent	<i>zan ... hamē wāng ī ān kōdak ašnawēd.</i> 'The woman ... always hears that kid's shouting.' (<i>Ardā Virāz</i> , ch.64, sentence 6)
3	Semi-conjunction with subordinator <i>ka</i>	Non-Adjacent	<i>hamēw ka wārān wārīd</i> ' Whenever it rained, ...' (<i>Mēnōg</i> , ch.62, s.36)
4	Adjectival modification	Non-Adjacent	<i>hamē pārōzgar wahmān ī wahmānān</i> 'You Wahman, the son of Wahman, the ever-victorious !' (<i>Nāmag Nibēsišnīh</i> , s.5)

compositional collocational pattern or we are witnessing the makings of a complex subordinator is not our concern here. The last group consists of 10 instances of *hamē* modifying an adjective (five times *pārōzgar* 'victorious', twice *zīndag* 'alive', twice *sūd* 'useful',⁵ lit. profit', and once *kāmag xwadāy* 'mighty, lit. Lord of desire'), and the examples are all found in only two texts.

In these first two groups, *hamē(w)* can be seen as contributing to the aspectual interpretation of the clause.⁶ Common configurations of position and aspectual value are illustrated in (3) for *hamēw* and in (4) for *hamē*. Note that focalized use is not attested for *hamēw*. Cases that are ambiguous between a lexical adverbial and an aspectual reading receive a double gloss, such as (3b), where *hamē* could either be interpreted in its adverbial lexical meaning 'continually', or as a grammaticalized aspectual marker of duration. Cases where there is more than one aspectual reading available, such as (3c), are considered indeterminate with regard to their aspectual category. We also make a distinction between the sentence-level labels (DURATIVE, HABITUAL, FOCALIZED) in the headers to each example, and the lexeme-level labels (continually/DUR/HAB/FOC) in the glosses. While they typically overlap, the gloss DUR only applies to adjacent cases with durative interpretation, in agreement with our hypothesis that only adjacent cases are potentially instances of more grammaticalized aspectual functions.

5 In the reading of Rashed-Mohassel (2010: 282).

6 We are not concerned here with what is sometimes referred to as 'lexical aspect' or *Aktionsart* (see Sections 4.4 and 4.5 for a discussion of this issue).

(3) a. NON-ADJACENT DURATIVE

āginēn hamēw hāmōx ud hāmzōr baw-ānd.
 together continually same.mind and same.strength be-OPT.3PL
 ‘May they be always of the same mind (thought) and strength together.’
 (*Manichaean Reader*, y, ch.20, par.3)

b. ADJACENT DURATIVE

ān kēn ... tā frašegird hamēw paywand-ēd.
 that malice ... until restoration continually/DUR continue-3SG
 ‘That malice ... will be continued until the Restoration (at the end of the world).’
 (*Mēnōg ī Xrad*, ch.21, s.26)

c. ADJACENT DURATIVE OR HABITUAL (LABELLED ‘INDETERMINATE’)

u=š wak ud abārīg xrafstar az=iš abāz
 and=3SG frog and other reptile from=it away
hamēw dār-ēd.
 continually/DUR/HAB keep-3SG
 ‘And it keeps the frogs and other reptiles away from it.’
 (*Mēnōg ī Xrad*, ch.62, s.30)

(4) a. NON-ADJACENT DURATIVE

ka=t hamē ēdōn kām-ist kū=t
 because=2SG continually such wish-PST that=2SG
abāg xiyōn-ān kārezār kun-ē.
 with Khion-PL fight do-2SG
 ‘Because you always wished to fight against the Khions.’
 (*Ayādgār ī Zarirān*, s.85)

b. ADJACENT DURATIVE

u=š šāhpuhr nām nih-ād u=š hamē parwar-d
 and=3SG Shahpur name put-PST and=3SG continually/DUR raise-PST
tā ō dād ī 7-sālag mad.
 until to age GEN 7-year come.PST[3SG]
 ‘And he called him Shahpur, and raised him continually until he became 7 years old.’
 (*Kārnamag*, ch.16, s.2)

c. ADJACENT HABITUAL

čiyōn xānīg-ēw ī rōšn ud anāhōg kē bast ēst-ēd
 like fountain-INDF GEN bright and undefiled that bound stand-3SG
ud ō kār nē hamē šaw-ēd.
 and to work not continually/HAB go-3SG

‘[It] is like a bright undefiled fountain which is tied up and is never put to use.’

(*Šāyist nē Šāyist*, ch.20, par.3)

d. ADJACENT FOCALIZED

pad čē har kār ud kirbag ī kun-ēnd ēn
in which every work and virtue that do-3PL this
guftan abāy-ēd kū wināh wizārišn ud
say should-3SG that sin leave and
kirbag ruwān dōšāram rāy hamē kun-ēm.
virtue soul love for continually/FOC do-1SG

‘Whenever (people) do something or practice a virtue, they should say that “I’m doing this to leave sin and out of love for the soul’s virtue”.’

(*Handarz ī Dastwarān*, s.17)

Table 2 sets out the distribution of adjacent and non-adjacent uses in more detail for each text. The texts have been sorted according to spelling first and then a descending percentage of the non-adjacent position. Texts with fewer than five instances have been put in brackets.

As Table 2 shows, the form *hamē* appears significantly more frequently in adjacent position than *hamēw* does (86.1% compared to 53.8%). While the raw numbers are low for *hamēw*, a Fisher exact test reveals that the sample distribution is not a likely outcome under the null-condition ($p = 0.0074$), and the difference means something.

Table 2: Position of *hamē(w)* with respect to the verb.

Text name	<i>hamēw</i>				<i>hamē</i>			
	Non-ADJC		ADJC		Non-ADJC		ADJC	
	#	%	#	%	#	%	#	%
<i>Manichaeae Texts</i>	5	71.4	2	28.6				
<i>Mēnōg ī Xrad</i>	1	16.7	5	83.3				
<i>(Hōm-yašt Pāzand)</i>					1	100.0	0	0.0
<i>(Ayādgār ī Zarirān)</i>					1	100.0	0	0.0
<i>Nāmag Nibēsišnīh</i>					4	80.0	1	20.0
<i>Dēnkard V, VII</i>					15	78.9	4	21.1
<i>Šāyist nē Šāyist</i>					5	19.2	21	80.8
<i>Indian Bundahišn</i>					2	6.9	27	93.1
<i>Kārnāmag</i>					1	5.0	19	95.0
<i>Ardā Virāz Nāmag</i>					3	2.4	123	97.6
<i>(Handarz ī Dastwarān)</i>					0	0.0	3	100.0
Grand total	6	46.2	7	53.8	32	13.9	198	86.1

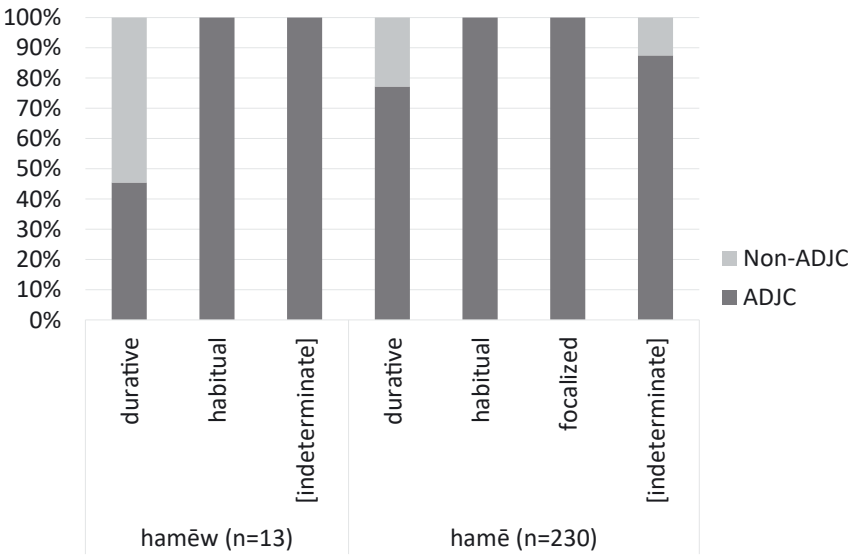


Figure 1: Distribution of aspectual functions over position and form of *hamē(w)*.

Figure 1 and Table 3 take the analysis one step further and provide an overview of the interaction between position, aspect, and form. To get at a more robust picture, texts displaying the same spelling have been aggregated.

Indeterminate instances of aspectual function generally involve ambiguity between durative and either habitual or focalized aspect, as illustrated in (3c) above. Even when allowing for cases like these, it appears that non-adjacent occurrences consistently carry a durative interpretation. Unambiguous focalized and

Table 3: Distribution of aspectual functions over position and form of *hamē(w)*.

Interpretation	<i>hamēw</i>				<i>hamē</i>				Grand total	
	Non-ADJC		ADJC		Non-ADJC		ADJC			
	#	%	#	%	#	%	#	%	#	%
DUR	6	100.0	5	71.4	31	96.8	105	53.0	148	60.9
HAB	0	0.0	1	14.3	1	3.2	64	32.3	64	26.3
FOC	0	0.0	0	0.0	0	0.0	22	11.2	22	9.2
[indeterminate]	0	0.0	1	14.3	0	0.0	7	3.5	9	3.6
Grand total	6	100	7	100	32	100	198	100	243	100

habitual interpretations are exclusively found with adjacent *hamē*. This confirms that durative aspect is the first natural extension out of a purely lexical adverbial sense of ‘continually, always’. It does not, however, provide any evidence on the relative ordering of habitual and focalized use. Some evidence (though not statistically significant) about this relative ordering is provided by the form variation. The (older) form *hamēw* is also generally durative (84.6% disregarding indeterminate cases), but habitual use is not excluded. Such a non-durative, habitual reading however is only possible for *hamēw* in two cases (out of all 95 possible instances of HAB and FOC combined). By contrast, adjacent *hamē* is unambiguously durative in only 53% of its instances, with a large remaining number being either habitual (32.3%) or focalized (11.2%).

4.3 Tense

The second variable we examine is that of tense. While we are primarily interested in reconstructing the aspectual development of *hamē(w)*, the investigation of how tense correlates with a number of other variables can be revealing. To this purpose we used tense as the outcome variable in a linear regression model (again using the data from groups 1 and 2 in Table 1), which tries to predict the value of that outcome variable on the basis of our correlating variables, referred to as predictor variables.⁷ We conflated future and present tense, which are both marked by present tense morphology.⁸ Categorical predictor variables fed into the model were the spelling of *hamē(w)*, its position, and the aspectual interpretation. With regard to the last one, we are not making any assumptions for now about the relative degree of grammaticalization of habitual and focalized uses. The genre of the text was added as an ordinal variable distinguishing the genres *non-narrative* > *mixed* > *narrative*. Table 4 below provides the coefficients with their estimates and *p*-values (present tense being the 1 outcome, past the 0 outcome). The important values in this table are the estimate (or coefficient) and $\Pr(>|t|)$ (or *p*-value). The estimate represents the strength of the correlation between our predictor variables and our outcome variable. A positive estimate

⁷ Data preparation and statistical analysis have been carried out with open-access software, R version 3.5.3 (R Core Team 2019).

⁸ Looking at future and present separately reveals that future tense is overrepresented with the form *hamēw* if occurring in non-adjacent position. Pearson residuals higher than two point to a significant deviation from the random distribution, and this value is reached in this context (in a correlation test distinguishing between *hamēw_adj*, *hamēw_nadj*, *hamē_adj* & *hamēw_nadj*). This overrepresentation may be a side effect of data scarcity (and concomitant genre biases), but investigating this falls outside the scope of this paper.

Table 4: Linear regression with tense as outcome variable.

F(6, 236) = 19.2, $p < 2.2e-16$, Adjusted R-squared = 0.311				
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.51852	0.04467	11.607	< 2e-16 ***
element: hamēw	0.16448	0.13757	1.196	0.2330
position: nadj	0.13895	0.08376	1.659	0.0984 .
aspect: foc	0.40201	0.09997	4.021	7.79e-05 ***
aspect: hab	-0.01969	0.06457	-0.305	0.7607
genre_ord.L	-0.37248	0.05916	-6.296	1.47e-09 ***
genre_ord.Q	-0.02277	0.06110	-0.373	0.7098

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

implies a correlation between the predictor value and present tense (with a hypothetical value of 1 meaning that they consistently co-occur), a negative estimate implies a correlation with past tense. The p -value is a measure of the significance of that correlation taking into account the Standard Error (which reflects how consistent the correlation of the predictor is). For convenience sake, R adds symbolic codes to different p -values, ranging from *** ('highly significant') to . (weakly significant).

From this Table it appears that the use of focalized aspect (as contrasted to durative and habitual aspect combined) is significantly correlated with the use of present tense, and that the correlation between them is stronger than that of any other variables. This is also obvious from the raw numbers: 19 out of 22 clear focalized uses are in the present tense. The second strongest correlation is, unsurprisingly, the one between a linear increase in narrativity (the factor *genre_ord.L*) and past tense. The spelling of *hamē(w)* or the use of habitual aspect (as contrasted with durative and focalized combined) do not significantly correlate with the choice of tense. As for position, Figure 2 suggests that there is a non-trivial difference between non-adjacent and adjacent. According to the linear model, the correlation between non-adjacent position and present tense is indeed significant, but only weakly so.

The reason why the correlation is only weakly significant is that the difference in distribution can largely be explained by the underlying variable of genre. Narrative texts apparently do not only prefer past tense, but also adjacent position. A time-based explanation would also tie in with the difference observed between *hamēw*, which prefers non-adjacent position and durative aspect, and *hamē*, which has all focalized uses and more adjacent instances.

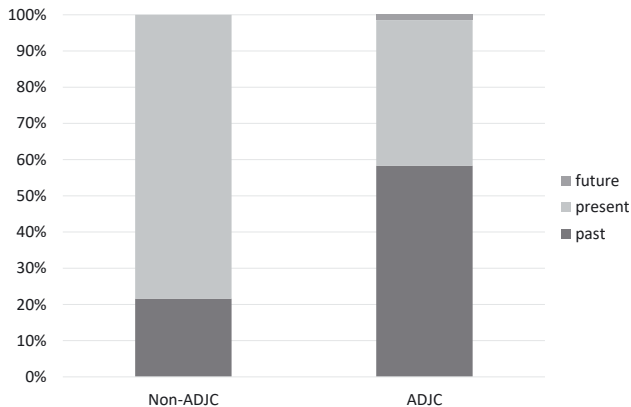


Figure 2: Tense and position.

4.4 Event type

A final variable that is expected to shed light on the aspectual development of *hamē(w)* relates to the semantic properties of the verbs that tend to select *hamē(w)*. With this end in view, we applied the traditional Vendler's (1957) classification of events into states, activities, accomplishments and achievements to our data, and looked for correlations with the position and aspectual interpretation of *hamē(w)*.⁹ As a durative adverb, *hamē(w)* should not normally be compatible with achievements, as these do not have any duration. States, activities, and even accomplishments all require some time to occur, and therefore are compatible with the lexical meaning of *hamē(w)*. Co-occurrence with achievements, then, might under certain circumstances be a sign of more advanced stages of grammaticalization.

Figure 3 visualizes how these variables interact.

A first observation to be made is the overall occurrence of achievements throughout. We hypothesized that these signal more advanced stages of grammaticalization, given their incompatibility with extended periods of time. However, in some cases where other time adverbials are present, achievements can arguably be reconciled with a lexical interpretation of *hamē(w)* without coercing a grammaticalized reading of the adverb itself. Example (5) is one such case.

⁹ Looking at the correlation with the spelling of *hamē(w)*, we see that states are slightly more prevalent with *hamē(w)*, but the difference is not statistically significant. It was pointed out in Section 3 that Josephson (2016) sees some correlation between early usage and activity verbs, but we do not currently have the data to provide robust statistical evidence for this. We also decided not to report in detail on the interaction with tense.

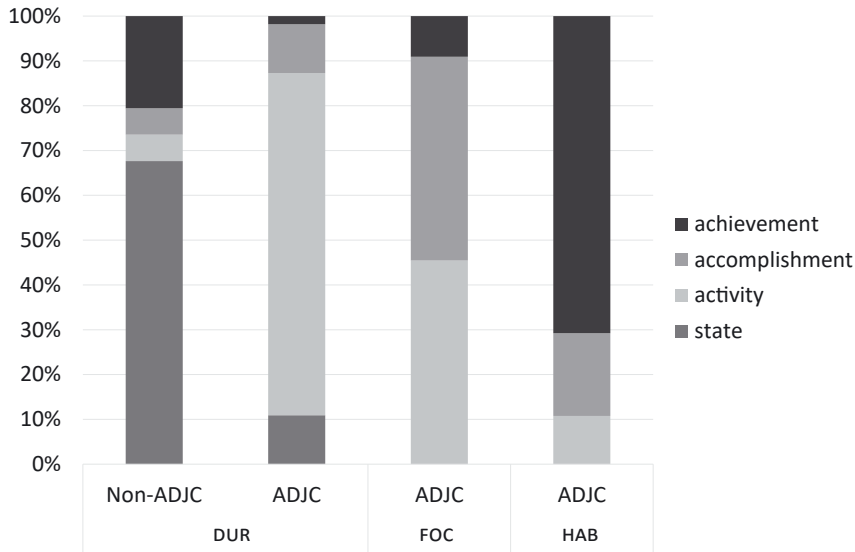


Figure 3: Correlation of event type with position and aspectual interpretation.

- (5) *paywastag hamē¹⁰ pad šab ka ardawān ī*
 non-stop continually at night when Ardawan GEN
bē-šudag-baxt be xuft, ān
 PFV-gone-fortune PFV/EMPH sleep.PST[3SG] that
kanizag pad nihān ō nazdik ardaxšir šud.
 maid in secret to beside Ardashir go.PST[3SG]
 ‘The whole time and without fail, at nights when the unfortunate Ardawan slept, the maid went to (see) Ardashir in secret.’
 (*Kārnamag*, ch.3, s.3)

A possible lexical reading of *hamē* in (5) would be ‘for a continued period of time, the maid visited Ardashir during the night’. While the habitual interpretation might be argued to be a case of coercion involving *hamē*, basically a pragmatic inference based on the incompatibility of ‘went to, visited’ (achievement) and ‘this whole time’ (duration), it is not necessarily the case that *hamē* itself triggers the

¹⁰ This word has been marked with an asterisk (*) in the TITUS-edition, showing reconstruction or correction. We did not include the symbol, following the transcription given in Faravashi (1975: 22).

habitual reading, because it would have emerged even without *hamē*, in this case to reconcile the event structure of achievement with the presence of an adverbial clause meaning ‘whenever Ardawan was asleep at night’. The adverbial idea of extended time is also reinforced by *paywastag* ‘non-stop, continually’. However, the import of such a second durative adverb is hard to assess, because its presence might also mean that the intended meaning of *hamē* itself is a different one (e.g., that of a habitual aspectual marker). Generally, instances like these might have served as bridging contexts for *hamē* to evolve into marking habituality aspectually, but because the habitual reading would be preserved when leaving out *hamē*, *hamē* is considered to fully preserve its lexical semantics of duration in them.

Most instances of achievements, however, combine with *hamē(w)* into a habitual or focalized reading, with no other contextual clues being present. An instance of an achievement in a focalized use is the final clause (containing the second instance of *hamē*) of (6):¹¹

- (6) *u=m dīd ruwān ī mard-ē kē=š kamāl pōst pahnāy*
 and=1SG see.PST soul GEN man-INDF that=3SG head skin width
hamē gīr-ēnd pad garān margīh
 continually/FOC take-3PL in painful death
hamē ōzan-ēnd.
 continually/FOC kill-3PL
 ‘I saw the soul of a man, the skin of whose head they were (continually) widening out, and with a painful death they were (*always) killing him.’¹²
 (*Ardā Virāz Nāmag*, ch.21, s.1)

In such cases the incompatibility with the lexical semantics of *hamē* is clear, showing that the adverb’s function has shifted to a more grammatical one. Habituals, in contrast, might be explained as a case of coercion, where the combination of duration with achievements automatically leads to the interpretation of repeated occurrence. However, according to Bertinetto et al. (2000), the occurrence of habitual uses (and achievement verbs within habitual uses) is a sign of ‘true imperfectivity’, the final stage in the progressive drift. The overall high occurrence of achievements in the category of habituals is, in any case, remarkable, but it is hard to see it as conclusive evidence, since the pragmatic implicature (without syntactic structure) of ‘always + achievement = repeated achievement’ is so widespread (as also observed by Bertinetto et al. 2000: 536).

¹¹ Here, the progressive aspect has coerced the achievement into an accomplishment.

¹² The original focalized verbs are in present tense, as the language does not display a ‘sequence of tense’ phenomenon in subordinate clauses.

Perhaps more important is the observation that the distribution in Figure 3 confirms the assumption that lack of adjacency generally correlates with a lower degree of grammaticalization. Within the non-adjacent durative instances, 65% combines with a state verb ([4b] is an instance of such a combination). With adjacent durative instances, this has decreased to a mere 11%, and this at the cost of activities, which take up 76% of all uses. An instance of a durative interpretation of an activity in adjacent position is (3b). Interestingly, the event type distribution of adjacent duratives is much closer to that of (adjacent) focalized uses. Perhaps, then, this indicates that adjacent durative activities provided a bridging context for the development of focalized uses.

4.5 Adverb–aspect–verb interactions: a semantic analysis

One question that is still open at this point is concerned with the semantic connections between the various uses in *hamē(w)*'s grammaticalization process. While Section 5 provides a thorough discussion of the possible range of (chrono) logical connections vis-à-vis our initial hypotheses, in this section we briefly explore the concept of pluractionality as a possible common denominator underlying these connections. In particular, we are concerned with event-external pluractionality, which (semantically defined) refers to an utterance signaling that the event expressed by the verb occurs more than once (cf. Bertinetto and Lenci 2012: 852 for definitions; Newman 2012: 191 for a short introduction to the investigations which treat pluractionality as a semantic concept). We leave aside the issue of event-internal pluractionality, which is essentially a subtype of *Aktion-sart*.¹³ Two semantic dimensions are at play in our case: the original meaning(s) of *hamē(w)*, and the semantic change(s) throughout the grammaticalization process. A crucial question is whether a single meaning thread can be detected throughout all types of usage introduced in Table 1. In an attempt to answer this question we examine to what extent, barring some exceptional cases, *hamē(w)* can be interpreted as expressing pluractionality, or some extension of pluractionality. To be more consistent with regard to pluractional labels, we have adopted part of the 'conceptual space' given by Mattioli (2019: 45; also cf. Mattioli 2020: 8–11 for precise, brief definitions),¹⁴ applying Mattioli's labels to our Middle Persian data. The results are displayed in Table 5.

¹³ Some of the examples labeled as 'continuative' in our data (for their distribution in time, cf. Mattioli 2020: 10) consist of sub-events, and they are thus good examples of event-internal pluractionality (e.g., *jū-d* 'chew-PST[3sg]', *kāh-ēd* 'decrease-3sg', and *dwār-ist* 'Evil walk-PST[3sg]').

¹⁴ Originally, in his works, these labels are meant to indicate morphological marking.

Table 5: The meaning of sentences including *hamē(w)* with respect to pluractionality.

Meaning	ADJC		Non-ADJC		Grand total	
	#	%	#	%	#	%
Iterative	53	25.9	1	2.6	54	22.2
Frequentative	10	4.9	4	10.5	14	5.8
Habitual	11	5.4	3	7.9	14	5.8
Generic	12	5.9	24	63.2	36	14.8
Continuative	89	43.4	4	10.5	93	38.3
Progressive	24	11.7	0	0.0	24	9.9
Iter/Prog	3	1.5	0	0.0	3	1.2
Cont/Prog	3	1.5	0	0.0	3	1.2
Non-pluractional	0	0.0	2	5.3	2	0.8
Grand total	205	100	38	100	243	100

A major difference between the pluractional labels in Table 5 and the aspectual labels in Figure 1 and Table 3 is that the former type of label approaches event interpretation holistically, taking into account how the verb's event type is modified by the semantics of all other elements in the context, while the latter type of label focuses solely on the impact of *hamē(w)* on the event interpretation. In other words, the pluractional meanings as assigned in Table 5 can be triggered by contextual elements other than *hamē(w)*. Such a holistic approach enables us to obtain a more fine-grained classification, which in turn may yield insights into the viability of a grammaticalization path starting out from a pluractional function. Instead of the general aspectual *HAB* category used in Table 1, for instance, different points on a continuum towards higher genericity are distinguished, viz. *iterative* > *frequentative* > *habitual* > *generic* (cf. Mattioli 2020: 20–21).¹⁵

Turning again to the discussion of the status of event types in the overall picture, *hamē(w)* with its *universal/omnial*¹⁶ quantificational meaning ('always') is appropriately compatible with all non-punctual events. Starting with states and activities, these events are 'unbounded' in nature, and some accounts in fact define them as "inherently pluractional" (van Geenhoven 2005: 117). Such events have

¹⁵ This different holistic approach also accounts for any seeming discrepancies between Table 1 and Table 5. For example, in the case of optative verbs, *hamē(w)* + verb in isolation may still be read as expressing (single-event) duration, but the optative triggers a frequentative pluractional sense, e.g. *hamē spōzād bēš ud gugānād zarīg* 'May [God] continually thrust sadness and destroy sorrow' (Nāmag Nibēsišnīh, s.19).

¹⁶ Cf. van der Auwera (1998: 87) for assigning "the temporal universal quantifier meaning" to 'always'; Moreno Cabrera (1998: 158) for the term *omnial*.

extensively been shown to be compatible with *for-adverbials* in the literature (e.g., van Geenhoven 2004), and this also seems to hold for ‘always’ as a “strong frequency adverb” (so-called by de Swart 1993 for its resemblance to strong nominal quantifiers like ‘every’). In the combination of atelic events with ‘always’, depending on the event type, the verb can either express a *generic* situation (‘always’), or a *continuation* of the event (‘continually’). As our etymological information about *hamē(w)* does not go beyond reconstruction (see Section 3), it cannot be decided whether ‘always’ developmentally precedes ‘continually’, or vice versa, and possibly both meanings were present from the start. The generic interpretation occurs with statives, in both adjacent and non-adjacent groups, and for adjectival modifications in which the adjectives are arguably stative predicates (cf. Dixon and Aikhenvald 2004; Jackson 2005; Lakoff 1966), and the continuous interpretation is found with activities.¹⁷

As for telic events, the combination of accomplishments and *for-adverbials* can give rise to a *continuous in time* reading (cf. Mittwoch 1991: 79; van Geenhoven 2004: 138),¹⁸ which can lead to the same interpretation with ‘always’ (cf. Figure 3 for the slight presence of accomplishments with durative aspectual interpretation). On the other hand, some accomplishments and achievements can trigger a multiple occurrence reading (again cf. Figure 3). As Table 5 shows, this reading amounts to 33.8% of all event types (the sum of iterative, frequentative, and habitual), and to these data we can add almost all instances of *hamē(w) ka* ‘whenever’ (group 3 in Table 1) which have frequentative/habitual meaning. More than any others, these cases exemplify the ‘plurality of the event’ in its strict sense. To account for these data, we can again extend the case of *for-adverbials* to ‘always’. Their combination with achievements triggers a frequentative interpretation, i.e. the verb receives a *separate in time* reading (cf. Mittwoch 1991: 79; van Geenhoven 2004: 138). Finally, certain telic events (accomplishments to a greater extent than achievements, cf. Figure 3) appear to give rise to a progressive reading (9.9% of the whole data in Table 5), which is no longer pluractional in the strict sense, but has been argued to develop out of pluractional context (Mattiola 2019: 34). In such cases the unbounded nature of *hamē* is compatible with, or provides, the required opportunity for the event to be “extended over time” (Mattiola 2020:

¹⁷ Appreciating an anonymous reviewer’s comment, we can add that this can be either as ‘sheer continuation’ or ‘continued effort to keep the action ongoing’. It seems that the choice depends on the amount of agentivity attributed to the verb’s subject.

¹⁸ Mittwoch’s (1991) description of the phenomenon requires a mass noun complement for both accomplishments and achievements, but van Geenhoven (2004: 139, 166) quotes some English examples with singular or no complements for both event types. In another article (van Geenhoven 2005: 110), she claims that “frequency marking on English accomplishment verbs in durative contexts is optional”.

10), and this continuativity opens a link for extension into the progressive (cf. Mattioli 2017: 133–134 for a short discussion of progressives in this regard). Essentially, then, all of the aspectual behavior shown by *hamē(w)* in Middle Persian might be accounted for as a single developmental pathway along the semantic space of pluractionality. However, while such a development is entirely possible, it is also somewhat speculative given the relative scarcity of well-dated data and the lack of robust typological data showing that this is a common source of progressives (cf. for instance Mattioli 2019: 60).

5 Discussion: grammaticalization and typological trends

In this section, we first discuss possible reconstructions of the grammaticalization pathway followed by *hamē(w)* in Middle Persian, after which the theoretical consequences of our reconstructed model will be discussed, especially with respect to the model introduced by Bertinetto et al. (2000) for the diachronic development of progressive in Romance languages.

The first issue to be addressed here is the structural position of *hamē(w)* in the clause. Turning back to Table 1, group 2 clearly belongs to the category of ‘adverbs’, but the category of members of group 1 cannot be determined with structural criteria: they can be adverbs, but their position next to the verb opens up classification into a range of possible verbal categories such as particle, clitic, derivational affix, or even inflectional affix. De Smet (2010: 75) calls this “the ambivalent character of grammaticalizing items in relation to the ‘source’ and ‘target’ of the grammaticalization process”. If we consider the *gradual* nature of a process like ‘decategorialization’, we can take the statistics presented in Table 2 as signalling the first important steps in this process. In the Table, there are three texts with high frequency of *hamē* in which this element appears adjacent to the verb in more than 93% of all instances (disregarding the bottom row text *Handarz ī dastwarān*, where the frequency of *hamē* is too low), and another text with more than 80% of adjacent instances. In these four texts, *hamē* seems to have lost most of its adverbial freedom to appear anywhere in the clause. This strongly suggests that these texts are further down the grammaticalization cline, as they illustrate ‘positional fixing’, that is “fixing of a lexical form in a specific potentially grammatical environment” (Hopper and Traugott 2003: 118, 127). This phenomenon can potentially pave the way for the gradual loss of syntactic independence. This particular case may point to a first stage of decategorialization from adverb to

verbal particle.¹⁹ However, there is as yet no sign of development of a bound morpheme in Middle Persian. This conclusion is in line with previous accounts outlined in Section 3.

Taking adjacency as a correlate of decategorialization and, hence, grammaticalization, we can also observe the following. Group 2, which contains non-adjacent instances with potential aspectual impact, shows no variation in interpretation. All 38 verbs may be interpreted duratively within the extended time interval provided by *hamē(w)*. In group 1, the adjacent group, more than half of the verbs (at least 53.7%) are also still interpreted duratively. However, the rest has to be interpreted either as being habitual or focalized. As far as duratives are concerned, it could be concluded that either *hamē(w)* is retaining its adverbial nature here, triggering the same durative interpretation, or else it reflects a more grammaticalized category (i.e. an aspectual particle) with the same contribution to the interpretation of the verb. Such a *meaning transfer mechanism* in early grammaticalization is in line with the hypothesis of gradualness of change (e.g., De Smet 2012), or “retention of earlier meaning” proposed by Bybee et al. (1994: 15–17), explained as “grammaticizing elements retain parts of their lexical meaning well into the grammaticization process” (1994: 148).²⁰ This persistence of the lexical meaning is also fully in line with the initial step of generalization (1) as stated in Section 1.

Things become hazier when it comes to deciding whether focalized progressives or habituals developed first out of duratives (cf. Table 3 for the numbers). Theoretically, the two simplest scenarios are the following.

First, the development may be DUR > FOC. In support of a direct link between durative and focalized use is the fact that they “have in common ... that they both view events as unbounded” (Killie 2008: 74). This link makes the semantic shift possible, from evaluating the event “relative to a larger interval of time” to viewing the event as “going on at a single point in time” (adopting the definitions in

¹⁹ An example of decategorialization of adverbs to verbal particles is found in some English verb particles “originating as spatial adverbs”, but developing increasingly “abstract meanings that typically modify the *aktionsart* of the verb they combine with, making the process designated by the verbal stem completive (*up, down, out, off*) or repetitive/ continuative (*along, about, around, on*)” (De Smet 2010: 84; cf. also the references therein). Another, probably more grammaticalized, example is the progressive particle in Cantonese, “where the particle is almost always used adjacent to the verb and while not entirely obligatory, is nevertheless extremely common” (Mair 2012: 805, who calls it a borderline case for inclusion in the category of progressive).

²⁰ Furthermore, there are supporting case observations such as Nicolle (2007) who concludes from the meaning content of the grammaticalized movement verbs in sentences such as ‘Let’s go find the paragraph marker’ and ‘Come wave goodbye’ that “grammaticalization need not involve semantic change” (2007: 49).

Bertinetto et al. 2000: 527 for durative and focalized, respectively).²¹ Initially, this new meaning may be seen as a kind of pragmatic extension of the durative one, but after some time the focalized interpretation is semanticized, becomes more frequent, and a ‘prototypical’ progressive construction emerges.

The second scenario would be a semantic generalization from durative to habitual (DUR > HAB) first. Bertinetto et al. (2000: 536) state that a durative progressive is “often available to a habitual interpretation, even without the facilitation of an appropriate syntactic structure”. Therefore, the extended time interval, shared by these two interpretations, could have worked as a common ground to give way to multiple occurrences of a continuing event.²² However, another option is to recast the whole second scenario as the generalization of durative progressive to a general imperfective (DUR > IMP), which also includes habituals. Such a hypothesis is in line with the observation that “[h]abitual meaning is more commonly included in the meaning of a more general gram, such as imperfective or present, than expressed separately” (Bybee et al. 1994: 159–160). What speaks against it, however, is that it is usually the focalized progressive (so FOC > IMP) that undergoes such a development into imperfective.

Returning to the data in Table 3, HAB seems to be more frequent than FOC. This might point to HAB having developed before FOC, at least if we assume that the next step in the development requires some degree of conventionalization (hence frequency) of the first step (as is commonly assumed in grammaticalization studies, e.g. Dahl 1985; Hopper and Traugott 2003: 82). However, there are various reasons why a comparison of global frequencies is inconclusive. To start, further inspection of the data shows that most of this imbalance is due to only one text: *Ardā Virāz Nāmag* which includes 51 instances of HAB (40.5%) and 12 of FOC (9.5%). Without this text, but including indeterminate labels, HAB and FOC occur almost as frequently (16 times [13.7%] and 13 times [11.1%] respectively). Still, HAB is slightly more evenly distributed (occurring in six texts, whereas FOC only occurs in 4), but the difference is in no way statistically significant. Then, there is also the habitual interpretation of almost all instances of *hamē(w) ka* (group 3) ‘whenever’. However, this is also inconclusive, as there is no logical connection with focalized uses. Overall, there is not much evidence to ascribe historical, let alone logical priority to it over FOC. The only evidence comes from the possibly older status of texts featuring spelling *hamēw* compared to *hamē*. If this hypothesis on their relative

²¹ For other explanations, Killie (2008) refers to Comrie (1976: 102–103) and de Groot (2007). See also Bybee et al. (1994: 170) for an explanation for meaning generalization from continuative to progressive.

²² In the formation of habituals, iteration of the event in the same situation (called iterative, subsumed here under habitual) may also have had a facilitating effect.

chronology is accepted, then the habitual interpretation of *hamēw* both in group 1 and group 3 could be seen as tentative positive evidence for the earlier development of *HAB*, though this would still not be evidence that *HAB* is a necessary step between *DUR* and *FOC*.

In conclusion, both of the scenarios *DUR* > *FOC* and *DUR* > *HAB* (and even the less plausible *DUR* > *IMP*) can be defended based on the data available in our corpus. Further investigation, especially into the New Persian Period, could possibly lend support to one of them. However, at least in the current data, no sign of *FOC* > *IMP* at the exclusion of *HAB* is found. Also, while habituals are considered later than *FOC* by Bertinetto, only arising when progressive aspect has generalized into imperfective aspect, the pathway *FOC* > *HAB* does not seem plausible, since focalized progressives are generally unavailable to the habitual reading (Bertinetto et al. 2000: 536). The developmental path of *HAB* > *FOC* cannot be verified, either (cf. Deo 2015: 25–26 for the non-attestation of *generic/HAB* > *PROG/IMP* path; Enke et al. 2016 and Mühlenbernd and Enke 2017 for verification of Deo’s claims; Bybee et al. 1994: 171 for their hesitation whether *HAB* > *IMP* is likely, indicating that they found no evidence for it). As a minimum final conclusion with regard to the generalization (1), it does not seem unlikely that both *HAB* and *FOC* developed separately from the ever dominant durative, with both of them also facilitating the later semantic generalization into *IMP*. Such a scenario could also be compatible with the role played by pluractionality as explored in Section 4.5. Based on the definition of desemanticization from Heine and Kuteva (2007: 34), these developments can be called “generalization in meaning content”.

Finally, the differences in the distribution of tense and event structure over non-adjacent and adjacent instances remain to be discussed. First, we see that in (the less grammaticalized) non-adjacent position there is still a strong preference for the present tense, with the past tense only representing 23.2% of all uses. The use of the past tense significantly increases for *hamē(w)* in (the more grammaticalized) adjacent position, to more than double (58.3%) (cf. Figure 2). This shift is largely mirrored if we only look at the durative uses (from 21.6 to 59.6%; not visible in Figure 2). This does not hold, however, for habitual and focalized uses. While habituals are past-oriented in adjacent position (70.7%), the focalized uses are overwhelmingly present-oriented (86.4%). To some degree this tense asymmetry may be explained by resorting to the default reading of each tense: “present tense is in fact the present imperfective” (Bybee et al. 1994: 175), which includes *HAB* and *PROG* as its interpretations. In contrast, the occurrence of the past tense by default signals a completed event.

Therefore, *hamē(w)* in the present can only increase the expressive power of the clause with regard to aspect, but its usage in the past disambiguates between the potential imperfective reading and the default perfective reading. This may explain

the higher share of the past tense in habitual use. Focalized uses do not seem to obey this general pattern, even though they could have in principle. The limited range of data in our corpus cannot motivate any reliable explanation, and the genre effect visible in Table 4 may partly explain away the unexpected distribution of tense. Future research into the behavior of the focalized use in New Persian will hopefully shed more light on the issue, but at least for now, our data do not support the second generalization made in the literature, that (focalized) progressive uses first develop in the past tense and only then in the present. The extremely low frequency of futurate uses in Middle Persian *hamē(w)* also signals that the stage of general imperfective aspect has not yet been attained (it is a common use of present imperfective in today's Persian). Again, further research into New Persian may yield evidence on when future appears on the imperfective scene.

A second difference concerns the event type of the verbs used with *hamē(w)*. The results summarized in Section 4.4 show that along with the shift to adjacent position, statives almost disappear from the scene, while achievements increase considerably, in particular in habitual uses. The largest increase may be seen in activities, especially in focalized progressives. Therefore, the predominant aspectual category at this stage appears to be durative/progressive (which is about the dynamic quality of the action), but it does not yet include 'continuous' aspect (describing a prolonged situation more generally). The extension into the continuous is considered to be a sign of further grammaticalization towards a general imperfective (cf. Bybee et al. 1994: 139; Mair 2012: 806), but this has not yet happened in our data of Middle Persian.

6 Conclusion

We argued that the different forms *hamēw* and *hamē* reflect two different stages of a diachronic development. Our analysis of how they are used is in line with this hypothesis. Generally, it is *hamē*, but not *hamēw*, that shows signs of being grammaticalized: it is used more consistently in adjacent position, it receives interpretations beyond its adverbial source, and it has a higher share of past tense usage. Yet, it is not clear whether the change *hamēw* > *hamē* is an integral part of the grammaticalization process, resulting from its expansion and increased frequency (i.e., a case of grammaticalization-induced phonetic 'erosion'), or *hamēw* simply went through this phonetic change while still at the lexical stage prior to its grammaticalization.

The first step in the grammaticalization process is normally that towards a durative progressive marker. There are two alternative pathways that could explain the rest of the development: a) as predicted in Bertinetto et al.'s (2000)

model, focalized progressive marking evolved out of this durative basis, eventually leading to pure imperfectivity, which then also includes habitual uses; or b) the focalized and habitual uses evolved separately out of the durative use. According to our data and analyses, it remains unclear which of these pathways is the one that has actually been taken by the grammar. We hope that future research on this marker in New Persian may shed more light on the issue. One promising avenue is the connection with the semantic space of pluractionality, and grammaticalization pathways through that space (cf. Mattioli 2019). Reappraising pluractional meaning of the adverb as a possible direct trigger of habituality is consistent with our conclusions and might eventually strengthen Bybee et al.'s (1994: 15–17) conclusions on semantic retention within grammaticalization. While we briefly discussed this alternative explanation in Section 4.5, for now we can only conclude that the first step of generalization (1) in Section 1 is supported by our data, but not the rest of it.

A stronger claim of the article with respect to theory of progressive marking is that the adverbial source of the marker seems to have played a role in the temporal orientation of the marker. Contrary to the typical priority of past tense in progressive developments, non-adjacent, adverbial *hamē(w)* has a clear preference for the present tense. As grammaticalization proceeds, past tense usage is increased in durative uses. Focalized uses, which are a further step on the grammaticalization cline, also appear to start in the present tense first, and seem not to have reached the next stage of extending into the past tense yet. While genre imbalances may in part explain away these distributions, overall, our data do not support generalization (2) in Section 1.

In sum, while there are clear similarities between the aspectual development of *hamē(w)* and the cross-linguistic tendencies described in, among others, Bertinetto et al. (2000), there are also non-trivial differences, which at the very least show that an adverbial, non-locational origin comes with its own grammaticalization pathway.

List of abbreviations

ADJC	adjacent
DUR	durative
EMPH	emphatic
FOC	focalized
GEN	genitive
HAB	habitual
IMP	imperfective

INDF	indefinite
Non-ADJC	non-adjacent
OPT	optative
PFV	perfective
PL	plural
PROG	progressive
PST	past
SG	singular

Acknowledgments: The authors would like to thank the two anonymous reviewers for their constructive and interesting criticism to an earlier version of this paper.

References

- Auwera, Johan van der. 1998. Phasal adverbials in the languages of Europe. In Johan van der Auwera (ed.), *Adverbial constructions in the languages of Europe*, 25–145. Berlin & New York: Mouton de Gruyter.
- Bertinetto, Pier Marco. 2000. The progressive in romance, as compared with English. In Östen Dahl (ed.), *Tense and aspect in the languages of Europe*, 559–604. Berlin & New York: Mouton de Gruyter.
- Bertinetto, Pier Marco, Karen H. Ebert & Casper de Groot. 2000. The progressive in Europe. In Östen Dahl (ed.), *Tense and aspect in the languages of Europe*, 517–558. Berlin & New York: Mouton de Gruyter.
- Bertinetto, Pier Marco & Alessandro Lenci. 2012. Habituality, pluractionality, and imperfectivity. In Robert I. Binnick (ed.), *The Oxford handbook of tense and aspect*, 852–880. Oxford: Oxford University Press.
- Boyce, Mary. 1977. *A word-list of Manichaean Middle Persian and Parthian* (Acta Iranica 9a). Leiden & Téhéran: Brill & Bibliothèque Pahlavi.
- Brunner, Christopher J. 1977. *A syntax of Western Middle Iranian*. Delmar & New York: Caravan Books.
- Bubenik, Vit. 1997. Development of Modern Iranian tense-aspect system. In John Hewson & Vit Bubenik (eds.), *Tense and aspect in Indo-European languages: Theory, typology, diachrony*, 304–313. Amsterdam & Philadelphia: John Benjamins.
- Bubenik, Vit. 2019. Grammaticalization and degrammati(calizati)on in the development of the Iranian verb system. In Lars Heltoft, Iván Igartua, Brian D. Joseph, Kirsten Jeppesen Kragh & Lene Schøsler (eds.), *Perspectives on language structure and language change*, 193–204. Amsterdam & Philadelphia: John Benjamins.
- Bubenik, Vit & Leila Ziamajidi. 2018. Restructuring of the Iranian tense/aspect/mood system. *Lingua Posnaniensis* 60(2). 73–98.
- Bybee, Joan, Revere Perkins & William Pagliuca. 1994. *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago & London: The University of Chicago Press.
- Comrie, Bernard. 1976. *Aspect*. Cambridge: Cambridge University Press.
- Cowgill, Warren. 1968. The aorists and perfects of old Persian. *Zeitschrift für vergleichende Sprachforschung* 82(2). 259–268.
- Dahl, Eystein. 2011a. Tense and aspect in Indo-Iranian, part 1: The present and aorist. *Language and Linguistics Compass* 5(5). 265–281.

- Dahl, Eystein. 2011b. Tense and aspect in Indo-Iranian, part 2: The perfect, futurate, participial and periphrastic categories. *Language and Linguistics Compass* 5(5). 282–296.
- Dahl, Östen. 1985. *Tense and aspect systems*. Oxford: Basil Blackwell.
- De Smet, Hendrik. 2010. Grammatical interference: Subject marker *for* and the phrasal verb particles *out* and *forth*. In Elizabeth C. Traugott & Graeme Trousdale (eds.), *Gradience, gradualness and grammaticalization*, 75–104. Amsterdam & Philadelphia: John Benjamins.
- De Smet, Hendrik. 2012. The course of actualization. *Language* 88(3). 601–633.
- Deo, Ashwini. 2015. The semantic and pragmatic underpinnings of grammaticalization paths: The progressive to imperfective shift. *Semantics & Pragmatics* 8. 1–52.
- Dixon, Robert M. W. & Alexandra Y. Aikhenvald (eds.). 2004. *Adjective classes: A cross-linguistic typology*. Oxford: Oxford University Press.
- Enke, Dankmar, Roland Mühlenbernd & Igor Yanovich. 2016. The emergence of the progressive to imperfective diachronic cycle in reinforcement-learning agents. In Sean G. Roberts, Christine Cuskley, Luke McCrohon, Lluís Barceló-Coblijn, Olga Feher & Tessa Verhoef (eds.), *The evolution of language: Proceedings of the 11th international conference (EVLANG11)*. <http://evolang.org/neworleans/papers/191.html> (accessed 16 July 2020).
- Faravashi, Bahram. 1975. *Kārnāme-ye Ardāshir-e Bābakān* [The book of the deeds of Ardāshir, Babak's son]. Tehran: Tehran University Press.
- Geenhoven, Veerle van. 2004. For-adverbials, frequentative aspect, and pluractionality. *Natural Language Semantics* 12. 135–190.
- Geenhoven, Veerle van. 2005. Atelicity, pluractionality, and adverbial quantification. In Henk J. Verkuyl, Henriette de Swart & Angeliek van Hout (eds.), *Perspectives on aspect*, 107–124. Dordrecht: Springer.
- Gignoux, Philippe. 1969. Notes sur la rédaction de l'Ardāy Virāz Nāmag : l'emploi de *hamē* et de *bē*. In Wolfgang Voigt (ed.), *XVII. Deutscher Orientalistentag. Zeitschrift der deutschen Morgenländischen Gesellschaft*, Supplement 1, 3. 998–1004.
- Gippert, Jost, Javier Martínez & Agnes Korn. 2016. *Thesaurus Indogermanischer Text- und Sprachmaterialien (TITUS)*. <http://titus.uni-frankfurt.de/indexe.htm> (accessed 16 July 2020).
- Groot, Casper de. 2007. The king is on Huntunge: On the relation between progressive and abensive in Old and Early Modern English. In Mike Hannay & Gerard J. Steen (eds.), *Structural-functional studies in English grammar: In honour of Lachlan Mackenzie*, 175–190. Amsterdam & Philadelphia: John Benjamins.
- Heine, Bernd & Tania Kuteva. 2007. *The genesis of grammar: A reconstruction*. Oxford: Oxford University Press.
- Hopper, Paul J. & Elizabeth C. Traugott. 2003. *Grammaticalization*. Cambridge: Cambridge University Press.
- Jackson, Eric M. 2005. *Resultatives, derived statives, and lexical semantic structure*. Los Angeles: University of California PhD dissertation.
- Josephson, Judith. 2013. The fate of the subjunctive in Late Middle Persian. In Folke Josephson & Ingmar Söhrman (eds.), *Diachronic and typological perspectives on verbs*, 57–77. Amsterdam & Philadelphia: John Benjamins.
- Josephson, Judith. 2016. The construction *hamē* + verb in Middle Persian. In Éva Á. Csátó, Lars Johanson, András Róna-Tas & Bo Utas (eds.), *Turks and Iranians: Interactions in language and history*, 49–64. Wiesbaden: Harrassowitz Verlag.
- Jügel, Thomas. 2013. The verbal particle *BE* in Middle Persian. *Münchener Studien zur Sprachwissenschaft* 67(1). 29–56.

- Killie, Kristin. 2008. From locative to durative to focalized? The English progressive and 'PROG imperfective drift'. In Maurizio Gotti, Marina Dossena & Richard Dury (eds.), *English historical linguistics 2006: Syntax and morphology*, vol. 1, 69–88. Amsterdam & Philadelphia: John Benjamins.
- Lakoff, George. 1966. Stative adjectives and verbs in English. In Anthony G. Oettinger (ed.), *Mathematical linguistics and automatic translation. Report NSF-17, I-1 – I-16*. Cambridge, MA: The Computation Laboratory, Harvard University.
- Lazard, Gilbert. 1989. Le Persan. In Rüdiger Schmitt (ed.), *Compendium Linguarum Iranicarum*, 263–293. Wiesbaden: Reichert.
- Lehmann, Christian. 2015. *Thoughts on grammaticalization*, 3rd edn. Berlin: Language Science Press.
- MacKenzie, David Neil. 1971. *A concise Pahlavi dictionary*. London: Oxford University Press.
- MacKinnon, Colin. 1977. The New Persian preverb *bi-*. *Journal of the American Oriental Society* 97(1). 8–26.
- Maggi, Mauro & Paola Orsatti. 2018. From Old to New Persian. In Anousha Sedighi & Pouneh Shabani-Jadidi (eds.), *The Oxford handbook of Persian linguistics*, 7–51. Oxford: Oxford University Press.
- Mair, Christian. 2012. Progressive and continuous aspect. In Robert I. Binnick (ed.), *The Oxford handbook of tense and aspect*, 803–827. Oxford: Oxford University Press.
- Mattiola, Simone. 2017. The conceptual space of pluractional constructions. *Lingue e Linguaggio* 16(1). 119–146.
- Mattiola, Simone. 2019. *Typology of pluractional constructions in the languages of the world*. Amsterdam & Philadelphia: John Benjamins.
- Mattiola, Simone. 2020. Pluractionality: A cross-linguistic perspective. *Language and Linguistics Compass* 14(3). 1–35.
- Mittwoch, Anita. 1991. In defence of Vendler's achievements. *Belgian Journal of Linguistics* 6. 71–85.
- Moreno Cabrera, Juan Carlos. 1998. Adverbial quantification in the languages of Europe: Theory and typology. In Johan van der Auwera (ed.), *Adverbial constructions in the languages of Europe*, 147–185. Berlin & New York: Mouton de Gruyter.
- Mühlenbernd, Roland & Dankmar Enke. 2017. The grammaticalization cycle of the progressive: A game-theoretic analysis. *Morphology* 27. 497–526.
- Nicolle, Steve. 2007. The grammaticalization of tense markers: A pragmatic reanalysis. In Louis de Saussure, Jacques Moeschler & Genoveva Puskas (eds.), *Tense, mood and aspect: Theoretical and descriptive issues*, 47–65. Amsterdam & New York: Rodopi.
- Newman, Paul. 2012. Pluractional verbs: An overview. In Patricia Cabredo Hofherr & Brenda Laca (eds.), *Verbal plurality and distributivity*, 185–209. Berlin & Boston: Mouton de Gruyter.
- Nyberg, Henrik S. 1974. *A manual of Pahlavi*. Wiesbaden: Harrassowitz.
- Petré, Peter. 2016. Grammaticalization by changing co-text frequencies, or why [BE *Ving*] became the 'progressive'. *English Language and Linguistics* 20. 31–54.
- R Core Team. 2019. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. <http://www.R-project.org/> (accessed 26 May 2020).
- Rashed-Mohassel, Mohammad-Taghi. 2010. *Dinkard-e haftom* [Dēnkard book 7]. Tehran: Institute for Humanities & Cultural Studies.
- Schmitt, Rüdiger. 1989. Andere Altiranische Dialekte. In Rüdiger Schmitt (ed.), *Compendium Linguarum Iranicarum*, 86–94. Wiesbaden: Reichert.
- Skjærvø, Prods Oktor. 2009. Middle West Iranian. In Gernot Windfuhr (ed.), *The Iranian languages*, 196–278. London & New York: Routledge.

- Sundermann, Werner. 1989. Mittelpersisch. In Rüdiger Schmitt (ed.), *Compendium Linguarum Iranicarum*, 138–164. Wiesbaden: Reichert.
- Swart, Henriette de. 1993. *Adverbs of quantification: A generalised quantifier approach*. New York: Garland.
- Tafazzoli, Ahmad. 1997. *Tārikh-e adabiyāt-e Irān-e pish az Eslām* [The history of Iranian pre-Islamic literature]. Tehran: Sokhan.
- Vendler, Zeno. 1957. Verbs and times. *The Philosophical Review* 66. 143–160.
- Windfuhr, Gernot L. 1979. *Persian grammar: History and state of its study*. The Hague: Mouton.