



Universiteit  
Leiden  
The Netherlands

## On the etymology of Persian yak 'one'

Bernard, C.B.A.S.

### Citation

Bernard, C. B. A. S. (2019). On the etymology of Persian yak 'one'. *Wékwos*, 4, 41-55.  
Retrieved from <https://hdl.handle.net/1887/3198863>

Version: Publisher's Version

License: [Licensed under Article 25fa Copyright Act/Law \(Amendment Taverne\)](#)

Downloaded from: <https://hdl.handle.net/1887/3198863>

**Note:** To cite this publication please use the final published version (if applicable).

Chams BERNARD  
Leiden University

## On the etymology of Persian *yak* ‘one’<sup>1</sup>

*À ma grand-mère Yagouta Chérif (1931-2019)*

ABSTRACT.—In this article I propose a new derivation of Persian *yak* ‘one’ and other related words. Based on a series of analogical, phonetic and semantic changes, I will propose that *yak* does not go back to *\*aiya-ka-* as previously suggested but, ultimately, to *\*aiyaja-ka-*.

### 1. Introduction<sup>2</sup>

The Persian and Manichean Middle Persian word *yak* ‘one’ (contemporary Iranian Persian *yek*) has a difficult derivation. It is impossible to derive it from *\*aiya-ka-* by regular sound laws, since the expected result would be Persian †*ēk*.<sup>3</sup>

Two different proposals have been put forward to derive *yak* from *\*aiya-ka-*: the first is a metathesis of *\*aiya-ka-* to *\*iayā-ka-*: it is implicitly done by HORN (1893:252) and BARTHOLOMAE (1895-1901:111). HÜBSCHMANN (1895:151) HORN, (1898-1901:44) and HASANDUST (2014: 2947), posit that *\*ai* became *ya* in two examples: *yak* from *\*aiya-ka-* and *yax* ‘ice’ from *\*aixa*. The second proposal is made by KLINGENSCHMITT (2000), whose theory explains the different forms of the development of *\*aiya* in late Old and Middle Iranian. As part of this theory, he makes Book Pahlavi <’ywk’><sup>4</sup>, MMP <yk> and NP *yak* go back to “*\*ēāhk-* < *\*ēhák-* (Gen. *\*ēhákē* < *\*aiyá-kahja*)”.

Both proposals will be discussed in this article (Sections 2 and 3). An alternative etymology of *yak* will be proposed in multiple steps (Sections 4, 5 and 6). I will then examine the use and distribution of *yak* and *yakē* (Section 7).

---

<sup>1</sup> I would like to thank the following people for their scientific help and their help with the article in general: Christopher and Kate BELLAMY, Gilles BERNARD, Johnny CHEUNG, Michaël PEYROT, Agnes KORN and Nicholas SIMS-WILLIAMS. I would also like to heartfully thank Professor KÜMMEL, Professor LUBOTSKY as well as Dr. GARNIER and Stefan NORBRUIS for having generously taken the time to discuss with me a number of the problems mentioned in this article. All mistakes and misconceptions remain my own.

<sup>2</sup> Abbreviations used: Av. = Avestan, Bal. = Balochi, Gav. = Gavruni, Lat. = Latin, JP = Judæo-Persian, MMP = Manichean Middle Persian, MP = Middle Persian, NP = New Persian, OIr. = Old Iranian, Ved. = Vedic. If not otherwise specified, the New Persian forms cited are Classical.

<sup>3</sup> Such a form is attested in Kurdish, especially in the conservative Iraqi and Yezidi Kurmanji dialects, see BLAU (1975:41; 214). I thank Johnny CHEUNG for bringing this fact to my attention.

<sup>4</sup> Which, in fact, should be <’ywk’>, the last character being a stroke, see FIG. 1. In KLINGENSCHMITT (2000) the final stroke is noted as aleph <’>.

Further, the New Persian form *yak* and its variants will be discussed (Section 8). Finally, other similar New Iranian forms will be presented (Section 9), before the conclusion (Section 10).

## 2. A metathesis that never happened

The metathesis of *\*ai* to *\*ia* is not presented as a sound law but as occurring sporadically by HÜBSCHMANN,<sup>5</sup> HORN and HASANDUST. Indeed, it has many more counterexamples than examples: only *yax* and *yak* stand as examples, whereas *\*ajta-* became *ēd*, and *\*naiya-ka-(ka)-* has yielded *nēk*. This metathesis of *\*aiya-ka-* to *\*iaya-ka-* should have happened before the monophthongization of *ai* to *ē*, thus even before Middle Persian (which only knows the monophthongized reflexes of OIr. *\*ai*). This would yield early Proto-MP *\*\*yōk* (*\*iaya-ka-* should have first become late Proto-Persian *\*\*iōkə*, as in Proto-Bactrian, cf. Bactrian ωγο ‘one’) and *\*yax*, thus MP and NP †*jōk* and †*jax* (HÜBSCHMANN, 1895:150-151).<sup>6</sup> In fact, it seems that the word for ‘ice’ in a number of Iranian languages does not go back to *\*ajxa-* but to *\*ajaxa*. See, for example, Sogdian (<yxn(w)>, /yəxnú/), perhaps also Choresmian (<yyx>) and Sistāni (*āx* < *\*jax* < *\*ajax*).<sup>7</sup> The Iranian word for ‘ice’ must thus be treated in a separate study.<sup>8</sup>

From all these points, it seems clear that the alleged metathesis of *\*ai* to *\*ia* never took place in the ancestor language of Middle and New Persian. It is nevertheless possible that it occurred in Proto-Pashto, cf. GEIGER (1893:206), yielding *yaw*, *yav*, *yō*.<sup>9</sup> This would be, according to GEIGER, l.c., the only example of this sound law.

<sup>5</sup> “In *yak* ‘ein’ und *yax* ‘Eis’ geht *ya-* auf *ai-* zurück.” HÜBSCHMANN (1895:151).

<sup>6</sup> There are four cases of the apparent preservation of word initial #*y* in NP (see HÜBSCHMANN, 1895:151). The conditions are unclear, but in these four words, the #*y* is followed by a long vowel: see the doublet *yōγ* ~ *juγ* ‘Joch’ (but Early NP *ayōγ*). It is thus possible that one dialect preserved word initial *y* before long vowels. The possibility that these words were loaned from a particular language or dialect is increased by their material nature. One of them, *yāwar* ‘Stössel’ (cf. Av. *yāvarəna-* ‘idem’) cannot be inherited: *yār* is expected (and also found). Another one, *yāsaman*, is possibly borrowed back from Arabic or another Semitic language. It is to be excluded, in my view, that *yak* and *yax* belong to this group of words. The words related to *yazad*, *yazdān* ‘God, God(s)’, such as *yaziš* ‘prayer, worship’ are loanwords from Avestan (through Pahlavi).

<sup>7</sup> This implies that the Sistāni sound law #*ya* > *ā* (and #*yā* > *ā*) is more recent than #*y* > *j* (seen in, for example, *jvō* ‘young’ < *\*yuān*). There are good arguments for this, such as *ala* ‘yalla’ < *yalla* (XOMAK, 2000:42), and *ārā* ‘power, might, energy’ (<ārâ> in XOMAK’s transcription) < *yārā* (XOMAK, 2000:3).

<sup>8</sup> I thank Professor LUBOTSKY for helping me with this word. Dr. Lubomír NOVÁK (p.c.) claims that NP *yax* ‘ice’ is a borrowing from an Eastern Iranian language. This would fit very well with the phonetics of the NP word.

<sup>9</sup> The change is made plausible by the fact that OIr. *\*au* > Proto-Pashto *\*ua* > Pashto *va*, except in front of a nasal (GEIGER, 1893:206). Yet, interestingly, we find two variants for the adjective



FIG.1 &lt;'ywk'&gt; in Pahlavi (CPD:233)

### 3. Gert KLINGENSCHMITT's proposal

KLINGENSCHMITT (2000) explains both the preservation of the suffix *-ak* in New Persian and the etymologies of a number of Persian words ending in *-ak* (thus apparently contradicting the sound law MP *-ag, -ak* > NP *-a*). For a number of cases he explains the preservation of the *-k* by a double suffixation *\*-á-ka-ka-* > *\*-ákka-* in an unattested stage of the language. However, for a number of other cases, namely, those going back to *\*-aiju-*, he employs a different strategy. Here he suggests that Old Iranian trisyllabic words show the effect “eines paradigmatischen Akzentwechsels”. For instance he considers <n'ywk'> ‘good (adj.)’ (sic, cf. fn 4) < *\*naijuaka-* < *\*najibaka-* as possibly representing either /nēk/ (as in New Persian) or /nyak/.<sup>10</sup> Departing from *\*najibaka-*, he proposes the following derivation: “In einem ursprünglichen Paradigma Sg. Nom. *naijuakəh*, Gen. *naijuakahja* trat durch Schwund des intersonantischen *ju* und andere Veränderungen zunächst eine Weiterentwicklung zu *\*né'aki*, Gen. *nē'ákē* und in der Folge durch Einschub eines hiattilgenden *h* zu *\*néhaki*, *\*néhákē* ein. Daraus entstand vor Aufkommen der Lenierung von intersonantischem *k* ein Paradigma Sg. Nom. *\*néhki*, Gen. *\*néhákē* > *\*nyáhkē*.” According to KLINGENSCHMITT (2000:212-13), different phonetic solutions could have yielded the two dialectal variants *\*néhk* (> *nēk*) and *\*nyahk* > *nya(h)k*. He similarly explains Book Pahlavi /hyc/ ‘irgendein (lat. *ullus*)’ as *\*aiju-* + *čit* > *\*éha-* + *či* > *\*éhči* > *hēc* and MMP, New Persian *yak* ‘ein’ from *\*éhk-* < *\*éhák-* (Gen. *\*éhákē* < *\*aijuakahja*).<sup>11</sup>

The demonstration by KLINGENSCHMITT (2000) touches upon various complicated and intricate matters. It is ingenious and has the advantage of explaining the preservation of the *-k* in all described words,<sup>12</sup> as well as the variants in Pāzand, not to mention the preservation of the *yod* in *yak*. I must

‘alone’: *ēv-ājai* (*ibid*:177) and *yav-ājai*. Since *ēv* alone apparently does not exist in Pashto, *ēv-ājai* should be the older, inherited form or, on the contrary, borrowed.

<sup>10</sup> KLINGENSCHMITT (2000:212) bases himself on the Pāzand readings <niiak>, <niak>, <niiakh> and on a reading in Hāfez in Šīrāzī <nkhw> ~ New Persian *nēkō* ‘goodness’.

<sup>11</sup> As Johnny CHEUNG (p.c.) highlights, Old Persian *\*-či* yields MP *-z*.

<sup>12</sup> Nevertheless, it seems unclear whether *\*-hk#* really yields *-k#* in New Persian. Its possible preservation in Šīrāzī needs a more thorough study: perhaps <h> was used to indicate a short vowel. In the Pāzand form <niiakh>, it could be a way to indicate /k:/. It is indeed surprising that it is never found in *yak*, for instance. Note that the reading <nōk> /ne:k/ is also found in Pāzand (cf. ANTIÁ, 1909:90; MOEIN, 2012:147).

nevertheless underline that several points seem to be problematic. First and foremost, the entire theory rests on the *h* of hiatus in Middle Persian, for the existence of which several arguments are adduced (KLINGENSCHMITT, 2000:212-213). These arguments are on the whole unconvincing, as they are all based on verbal and post-verbal forms, where analogy could have taken place.<sup>13</sup> Had there been an *\*ēhák* proto-form at any stage of late Old Iranian, or early Middle Persian, I do not see why a metathesis would have occurred, thereby yielding *\*ēáhk-*, when the form should have remained MP †*ēhák*, NP †*ēha*, instead of becoming the phonetically difficult [hk] cluster. Secondly, the reading of all quoted Pahlavi forms is subject to caution. It is equally problematic that no descendant of *\*nyahk* remains, to my knowledge, in modern Iranian. Thirdly, accepting the sound changes of KLINGENSCHMITT, if *\*éh(a)či* yields *hēč*, then why would *\*ēh(á)k* not yield †*hyak*? More importantly, why is *\*ájuəh* yielding Middle Persian (-)ēw (KLINGENSCHMITT, 2000:212<sup>78</sup>) and not †*hē* (through *\*éha* < *\*é'a*, if we want to follow KLINGENSCHMITT's own logic, cf. *hēč*)? Or, in other words, why do we have a trace of the *h* in some words, but not in others? Finally, all these changes need to have happened after the #y- > #j- sound change of Middle Persian, which is “older than the voicing of intervocalic stops, i.e. [it] took place already in a period preceding any Middle Persian texts” (KORN, 2016:407). This supposes that *\*ēáhk-* changed into *yak* at a later period, possibly after the 3<sup>rd</sup> century. This seems chronologically doubtful, considering MMP <yk>: we would expect at least a few occurrences of †<yhk>. What is nevertheless possible is that *\*ēwak* became *\*ēak* in Proto-MP and, in turn, became /jak/ after the described #y- to #j- sound change. This hypothesis is nevertheless contradicted by forms such as MP, NP *nēk* ‘good’.<sup>14</sup> In view of all these difficulties, I believe it is better to propose a new derivation for (M)MP, NP *yak*, which might also account for a related Balochi form (see 9.3).

#### 4. The prehistory of *yak*

To explain the proposed forms that could have given *yak*, a series of morphological and semantic developments will be proposed. I will present two different scenarios, the first is based on an adjectival derivation, and analogy from the word *\*duaja-(ka-)* ‘double’, while the second draws on Ossetic developments. In both scenarios I will propose an identical phonetic

<sup>13</sup> The only lexical examples cited by KLINGENSCHMITT (2000:213<sup>77</sup>), namely *xāk* ‘earth, soil’ and *nazdīk* ‘close, nearby’ could be explained differently: *xāk* finds a more convincing etymology, in my opinion, in KÜMMEL (2018:166), *xāk* < *\*hāsa* (< *\*h<sub>2</sub>ahs-*), and *nazdīk* could be explained as coming from *\*nazdī-ka-ka*.

<sup>14</sup> While the Pāzand forms mentioned in fn 10 might indeed be explained through *\*/ne:ak/*, *nēk* cannot.

reconstruction of the proto-form of *yak* ‘one’: *\*aiyajakā*. The later developments leading to *yak* will be discussed in Section 8.1.

### 5. First scenario

In MP “*ēw* occurs as an enclitic suffixed to the noun it modifies to indicate an indefinite article. *ēw/ē* can also precede the noun, in which case it alternates with *ēk* to express indefiniteness or the numeral ‘one’”<sup>15</sup> (JOSEPHSON, 2011:25). In this scenario, I suggest that, in Proto-MP, *\*ēw/ē* did not alternate with the ancestor of *yak*, *\*īiak* which meant ‘unique, single’, cf. Parthian *ēwag* ‘solely, only’, and Ved. *ekakā-* ‘einzig, alleinig’ (EWAia:263). In this hypothesis, the analogy was made with *\*duaja(h)* ‘double’, yielding the meaning ‘unique, single’. The meaning ‘one’ for *\*īiak* was acquired progressively, starting from negative sentences: “not a single [one] has come”, “there is not a single one”, and then, in affirmative sentences “there is a single one” > “there is one”. There would be a parallel in Vulgar Lat. *ūnus* ‘one’ which originally meant ‘single’.<sup>16</sup> Once the meaning of ‘one’ was acquired and generalized for *\*īiak* > *yak*, the role of *ēw* was progressively confined to that of an enclitic indicating indefiniteness. This perhaps explains why the *-ēw* clitic is not found in inscriptional Pahlavi, notably archaic (JOSEPHSON, 2011:37). It is equally possible that, by taking on this role, *ēw* “encouraged” the new function of *\*īiak*. This scenario is summed up in the following table.

<b>Proto-Persian (1)</b>	<i>*aiya</i> ‘one, an’	<i>*aiyaja(h)</i> ‘unique, single’ ←	<i>*duaja(h)</i> ‘double’
<b>Proto-Persian (2)</b>	id.	<i>*īia-ka-</i> (< <i>*aiyajakā</i> ) ‘unique, single’ ←	<i>*duajaka-</i> ‘double’
<b>Proto-MP</b>	<i>*ēw</i> ‘one, an’	<i>*īiak</i> ‘single (one)’, used in negative sentences	id.
<b>MP</b>	(-) <i>ē(w)</i> ‘one, an’ + indefinite marker	( <i>ī</i> ) <i>yak</i> ‘one’, used in affirmative sentences	n.a.
<b>Late MP, NP</b>	<i>-ē</i> indefinite marker	<i>yak</i> ‘one, a’	n.a.

<sup>15</sup> It should be noted for clarity that *ēk* is a transcription of Pahlavi <’ywk’>, see FIG. 1. It is indeed the expected outcome of *\*aiya-ka-*, but we do not have any further proof that such a form existed in any variety of Middle Persian.

<sup>16</sup> I thank Romain GARNIER for this suggestion and for the Latin parallel.

The first step is not necessary: it is possible that the form *\*īja-ka-* (< *\*aiuajaka*) was directly based on *\*duajaka-*. As is often the case in Iranian, there is little difference in meaning between *-ka-* derivatives and their underived forms.

## 6. Second scenario

### 6.1. Investigation of Proto-Iranian numerals

It is a well-known fact that analogy plays a major role in the formation of number terms: a quick look at the Persian numbers from 11 to 19, for instance, suffices to illustrate this phenomenon.<sup>17</sup> Also, the *h-* of NP *hašt* ‘eight’ is notably due to analogy with *haft* ‘seven’, and the *-h* of *nuh* ‘nine’ is from that of *dah* ‘ten’. It is thus also possible that the word for ‘one’ was influenced by the following numbers. Let us first examine the Proto-Iranian ordinal and cardinal numerals one to five, in the nominative masculine, to try to understand which numbers could have influenced *yak* and its predecessor.

NUMBER	CARDINAL	ORDINAL
‘one’	<i>*aīyah</i>	<i>*fra-tama, *prHu(i)ja-</i>
‘two’	<i>*duyaH</i>	<i>*dūtījah</i>
‘three’	<i>*θrajah</i>	<i>*θritījah</i>
‘four’	<i>*čaθuār-(ah-)</i>	<i>*ktūrījah</i>
‘five’	<i>*panča</i>	<i>*puxθa-</i>

The numbers ‘four’ and ‘five’, being too far from ‘one’, they are not likely to have exerted any influence. This leaves us with *\*duya* and *\*θrajah*. Both *\*aīyah* (< *\*Hoi-uo-*) and *\*θrajah* (< *\*tréjes*, Ved. *tráyas*) end in *-ah*, unlike *\*duya* (Ved. *d<sub>u</sub>vá*). It is thus possible that, under the influence of *\*θrajah*, Proto-Persian underwent a change from *\*aīyah* to *\*aīuajah*. Nevertheless, it does not seem probable — in this context — that ‘one’ underwent analogy from ‘three’, while ‘two’ did not, unless there was a major analogical leveling such as for the numbers 11-19 in Persian (see fn 17).

<sup>17</sup> For example, NP *dawāzdah* ‘12’ and *nawāzdah* ‘19’ have their *-z-* from *sēzdah* ‘13’ and *pānzdah* ‘15’, and possibly from *yāzdah* ‘11’. *šānzdah* ‘16’ and various dialectal variants such as *sīnzdah* ‘13’ have their *-n-* from (var.) *yānzdah* ‘11’, *pānzdah* ‘15’; Iranian Persian *hivdah* ‘17’ and *hiždah* ‘18’ have their vowel from (Ir. Pers.) *sizdah* ‘13’, etc. Cf. SCHMITT (1994). For further analogies in Bal. numbers (borrowed from NP), see KORN (2006:204).

### 6.2. Analogy from three: Ossetic and Wakhi parallels

In Ossetic, ‘two’ is *duw(w)æ* (Digor) and *dywwæ* (Iron). These forms go back to the Proto-Iranian neut. nom. and acc. du. *\*duyai* (CHEUNG, 2002:63). Iron and Digor *ærtæ* ‘three’ goes back to the nom. pl. masc. *\*θrajah* (CHEUNG, 2002:64). According to CHEUNG “[t]he final *-æ* of *ærtæ* is to be derived from the Npl. m. *\*-ajah*, which became *\*-ai* after apocope. The new diphthong *\*-ai* then merged with the reflex of older P[roto-]Ir. *\*-ai* [...]” (*op. cit.*, p. 64). Nevertheless, it seems strange that one numeral derives from the neuter, when all others derive from the masculine. SIMS-WILLIAMS (*forthcoming*:147) proposes that *\*-ai* only becomes *-æ* in “monosyllables, perhaps specifically [...] monosyllabic clitics”. He further suggests that: “The most straightforward solution to this problem [i.e. that of the *-æ* in *duwæ*], [...] is to attribute the final *-æ* of *duwæ* to the influence of *ærtæ* “three”, which seems to be the regular outcome of *\*θrayah*. This idea is not as outlandish as it might at first appear, since it is necessary in any case to assume an analogical spread of final *-æ* to account for *jewæ*, the independent form of the numeral “one” in Digoron (CHEUNG 2002: 195). The same explanation can perhaps also be applied to the final *-æ* of *fæjnæ* “each” and *innæ* “the other” in view of their close semantic relationship to the numerals “one” and “two.”” (SIMS-WILLIAMS, *id.*). If SIMS-WILLIAMS’s proposal is correct, there are two possible developments: **1.** *\*duya* was replaced by *\*duyaja* ← *\*θrajah* before the apocope took place, in Proto-Ossetic; **2.** later, after the apocope, *\*duwwa* was replaced by *duwwæ* ← *ærtæ*.

Wakhi, another Iranian language, shows a similar kind of analogy: *\*dua* becomes *bu* but the form *buy* is found, whose *-y* is taken analogically from *truy* ‘three’ (< *\*θrajah*), cf. STEBLINE-KAMENSKY, 1999:105.<sup>18</sup> This could naturally have happened at different stages of the history of the language.

I have argued in this section that, in Proto-Persian,<sup>19</sup> a similar leveling to that in Ossetic took place. Let us suppose, for instance, that *\*duya* was replaced by a phonetic form *\*/duja(h)/*, by analogy with *\*θrajah*, we thus have a system where *\*aiyah* is “the odd one out”, and may be influenced by ‘two’ and ‘three’, as in the Ossetic example. It may thus undergo analogy and become *\*aiyaja(h)*.

### 6.3. Replacement of *\*duya* ?

A replacement of *\*duya* by *\*/duja(h)/* akin to the situation in Ossetic is possible. There are two different reflexes of the word for ‘two’ in Middle and New Persian, as well as in closely related languages:

<sup>18</sup> I thank Federico DRAGONI for making me aware of this fact.

<sup>19</sup> Proto-Persian is taken here to mean the ancestor of New Persian spoken in the Old Iranian period, which, for a number of reasons, cannot be the Old Persian of the preserved Old Persian texts, but a dialect or a sister language thereof.



- Parthian, MMP, Pahlavi <dw> (DMMPP:143; CPD:26), Early JP <dw> /dō/ (Paul, 2013:91), Bal. *dō*, a loan from NP (Korn, 2005:121), like the vast majority of Balochi numerals (Korn, 2006:207).

- Kāboli *dū* and *du* (GLASSMAN 1972:22), Standard Iranian New Persian *do* < *dū* < *dū*.<sup>20</sup>

Both forms exclude a reconstruction *\*dya*, which would yield Middle Persian †*du*, and they also exclude *\*d(y)āu* as a proto-form (pace NYBERG, 1974:64), which would give †*dā* (see HÜBSCHMANN, 1895:168). Agnes KORN (p. c.) suggests, as a possibility, that *\*duya-* would become *dū*, while *\*duyá-*, or *\*duya-* with stress on another syllable, would become *dō*. If this hypothesis is correct, it would be easy to explain the *dō* forms as reflexes of *\*duyája* (or *\*duyaja*), in which case *\*duyaja* survived alongside *\*duya*. This would account for the existence of those two different forms in different languages and dialects of Persian. Considering Middle Persian <dw'zdh> /dwāzdah/ 'twelve', it could equally come from *\*duādaša* (Ved. *dvādaśa*) or *\*duājadasha*. The existence of a *\*duyaja* proto-form is thus uncertain, but possible. To test the first hypothesis of the present article concerning the word for 'one', I will consider that this replacement did indeed take place, similar to what has been proposed for Ossetic (see Section 6.2.1).

#### 6.4. *\*aiyaja-* as an analogically adapted form after *\*duyaja-* and *\*šraja**h*

Since the form *\*aiyah* is preserved in both Middle and New Persian (<'yw> in (M)MP, and the indefinite clitic *-ēw* > late MP, NP *-ē*), it seems reasonable that, at some point in the history of Proto-Persian, *\*aiya* meant 'a(n)' while the proto-form of *yak* meant 'one'. Such a semantic specialization explains the difference in the later developments of both forms.

The possible developments that led to this reconstructed form recall the foundations of Iranian *Wortbildung*: 1. *\*aiyah* means 'one', as in Avestan and Old Persian; 2. *\*aiyah* takes on the additional meaning 'a(n)', and is often postposed;<sup>21</sup> 3. possibly after the loss of dual grammatical number, the analogy of *\*duyaja* with *\*šraja(h)* occurs; 4. the analogy of *\*aiya* with *\*duyaja* from *\*šraja(h)* happens: *\*aiyaja* replaces *\*aiya* in its function as the numeral 'one' (vs. *\*aiya* which still means 'a(n)'). This situation is identical to that in Digoron, where: "(j)ew [...] is used with nouns, and (j)ewæ [...] is used otherwise" (MILLER, 1881-1887:159 *apud* CHEUNG, 2002:195),<sup>22</sup> 4. *\*aiyaja* is suffixed with *-ka-*.

<sup>20</sup> Perhaps Gav. *dūta* [dyta] (Malati dialect) 'two units (of something)' belongs here as well. It is hard to determine whether Gav. *do* 'two' is inherited or borrowed from another Iranian language.

<sup>21</sup> *aiya-* already means 'a(n)' in Old Persian in some contexts (CHEUNG, in prep.).

<sup>22</sup> The initial *yod* in *jew*, *jewæ* is phonetically triggered by the *#e*, and is completely secondary.

All these steps are outlined in the following table:

<b>First step</b>	<i>*ai̯uāh</i> ‘one’		<i>*du̯uā</i> ‘two’	<i>*ǧrajah</i> ‘three’
<b>Second step</b>	<i>*ai̯uāh</i> ‘one, a(n)’		id.	id.
<b>Third step</b>	id.		<i>*du̯uāia(h)</i> ←	<i>*ǧraia(h)</i>
<b>Fourth step</b>	<i>*ai̯uā(h)</i> ‘a(n)’	<i>*ai̯uāia(h)</i> ‘one’ ←	<i>*du̯uāia(h)</i>	<i>*ǧraia(h)</i>
<b>Fifth step</b>	<i>*(-)ēw</i>	<i>*ai̯uāia-ka-</i>	id.	id.

As *yak* is the only Proto-Persian numeral which receives the otherwise extremely productive Iranian *-ka-* suffix,<sup>23</sup> the suffixation must be explained in this scenario. While *\*du̯uāia* ‘two’ could not be confused with *\*du̯uāia* ‘double’ (Ved. *dvayá*), there was no way to differentiate (at least in the nominative) *\*ai̯uāia* meaning ‘one’ from a derivative in *\*-ia-* meaning ‘simple, unique’, similarly to *\*pr̥Hu(i)ia-* ‘first’, synchronically analyzed as a *\*-ia-* derivation of *\*pr̥Hu-* (Av. *pao<sup>u</sup>ru*). The suffixation in *-ka-* nominalized the numeral, whereas *\*du̯uāia* could still be used for ‘two’. It is possible that there was no such confusion for *\*ǧraia(h)* because there was no *\*ǧra-* from which *\*ǧraia(h)* could be derived by reanalysis.<sup>24</sup>

### 7. *Xujasta Siāmak yakē pūr dāšt*:<sup>25</sup>

#### the use and distribution of *yak* and *yakē* in New Persian<sup>26</sup>

The form *yakē* with the meaning ‘one’ occurs as often – or more often – than *yak* in many early NP texts. It occurs “très souvent” in early NP (LAZARD, 1963:215), and even translates Hebrew *ehād* ‘one’ in Early JP (PAUL, 2013:70).<sup>27</sup> Here, a distinction must be made between *yakē*, synonymous with *yak* ‘one’, and *yak-ē* ‘someone’, ‘once’ (viz. contemporary Iranian Persian *yeki* ‘someone’). The latter derives from a nominalized *yak*, as in the English ‘I have seen that **one**’ with the clitic of indefinite *-ē* (< *-ēw*). However, *yakē* ‘one’ can hardly be analyzed as *yak* + indef. *-ē*: how would a word meaning

<sup>23</sup> Except *hazār* ‘a thousand’ → *hazārag* ‘millenium’ (CPD:43), *sad* ‘a hundred’ → *sadag* ‘century’.

<sup>24</sup> Nevertheless, I admit that *du̯i-* ~ *du̯uāia-* compared to *ǧri-* ~ *ǧraia* is problematic.

<sup>25</sup> “Blessed (*xujasta*) Siāmak had (*dāšt*) one son (*pūr*)”, FERDOWSI, *Šāhnāma*, *Kiūmart* chap. 2.

<sup>26</sup> Most of the points discussed in this section I take from CHEUNG (in prep.). I thank him for giving me access to his manuscript.

<sup>27</sup> Unfortunately, WOLFF (1935:878-79) lists both *yak* and *yakē* ‘one’ together.

‘one’ benefit from being qualified as an indefinite? And even if that were the case, how would it differ from *yak*? Nouns following *yakē* can take the indefinite ending, as much as nouns following *yak*, for example: *yakē ātaš-ē* ‘one fire’, *yakē laškar-ē* ‘one soldier’ (both from the *Šāhnāma*). *Yakē* is, in fact, a remnant of the oblique case (CHEUNG, *in prep.*), stemming from the genitive *\*ai̯uajakahja*. It is most commonly attested in the Pahlavi <’ywky> ‘one’ form found on coins. This explains why the form used in compounds (see Section 8.2) is always *yak* and never *yakē*. It is probable that *yakē* was generalized because it was simply reinterpreted as *yak* + *-ē*. It is also possible that the oblique function of *yakē* was retained in early NP. *Yakē* ‘one’ mostly fell out of use in the Classical and Modern NP periods, but not entirely: it is still found in a number of dialects and languages, which most often have loaned it from Middle or New Persian.

## 8. Phonetic problems of *yak*

### 8.1. Preservation of the *yod*

Two strategies could explain the preservation of the word initial *yod* in this word. The first consists of a morphological dissimilation from right to left in *\*ēyak-ē* > *yak-ē* ‘someone, a single one’. This must have happened after the #y- > #j̄- sound law, but before the earliest Middle Persian texts.

The second solution, suggested by CHEUNG (p.c.), is that *\*ai̯uajaj(h)*, containing twice the diphthong *ai̯* underwent a dissimilation in the (late) Old Iranian period, and became *\*i̯uajaj̄*.<sup>28</sup> This would then give *\*i̯uajaj̄(-ka-)* > *\*i̯uaj̄(-ka-)*.

The *\*ui̯* cluster would then simplify to /j/ (see SALEMANN, 1895-1901:279 for other examples of this simplification), yielding the form *\*i̯jak*. The stress would then shift towards the last syllable, since an unaccented *-yak#* would have yielded †*ī(k)* (SALEMANN, 1895-1901:277). This could have happened because of the (Late Proto-Persian) alternation *\*i̯a-ka- ~ \*i̯á-kēh-* (see Section 7), with the paradigmatic accent change described by KLINGENSCHMITT (2000) – (see Section 3).

This reconstruction either implies that *ī + yod* simplified to *yod* after #y > #j̄ but before any other vowel, or that MMP <yk> could also represent [i(:)jak].<sup>29</sup>

In support of the dropping of *ī* before *yod*: it appears that the NP sound change of #aCV > #CV (HÜBSCHMANN, 1895:120-22) actually consists of two different changes: 1. #ayV > #yV; 2. #aCV > #CV. The first step probably happened in late Sassanian times, and the second in early Islamic

<sup>28</sup> For a similar phonetic development, cf. Old Persian *parīyana-* ‘behavior’ < *\*pariyayana-*, SCHMITT, 2014:227.

<sup>29</sup> Since <yy> notes /yē/, /yī/ or /yi/, it could not be used for [j(:)].

times, which would explain why we find alternations between archaic forms and new forms in Early NP, as well as in poetry, for instance *abar* ~ *bar* ‘on’, *abāz* ~ *bāz* ‘again’, *anār* ~ *nār* ‘pomegranate’ (*anār* is now the Standard Persian form) etc.<sup>30</sup> This also explains the fact that Arabic loanwords are also concerned, such as *mīr* ~ *amīr* ‘commander’, but forms such as *ayār* ‘companion’, *ayād* ‘memory’, etc. are not found in New Persian literature, where *yār*, *yād* are found instead.<sup>31</sup> It is thus possible to add yet another sound change, and suppose that *\*#īy* simplified to *#y* immediately after *#y* became *#j̄*. Unfortunately, there are no other examples of word-initial *īy* or *īy* in Middle and New Persian.

## 8.2. Preservation of the *-k*

*\*(ī)yakē* yielded early New Persian *yakē*, and *\*(ī)yak* yielded NP *ya*. Both *ya* and *yak* are already found in Early JP texts (PAUL, 2013:91).<sup>32</sup> NP *yak* itself is a back-formation, based on compounds and on *yakē*.

The final consonant remained unvoiced (instead of the expected *†-g*) because of compound forms such as *yaksarī* ‘gänzlich’, *yakšambad*, *-ba* ‘Sunday’, *yak-tan* ‘someone’, MP *yakyak* ‘double’ (DMMPP:373), and so on: *yak* was simply generalized from those forms.

The systematic devoicing that occurred in these compounded forms would have yielded *yag*, *yagē* as full forms and *yak-* as the first member of compounds, at least before unvoiced consonants. Some Iranian varieties and languages kept the expected *-g* consonant, usually as a variant of *yak*. Dari, for instance, has a variant *yag* (e.g. *yag zara* ‘a bit’ = contemporary Iranian Persian *ye zare* ‘idem’, *yagān* ‘some’, a pluralized form of *yag* ‘one’ GLASSMAN (1971:116;289), Sistani has *yag* along with *yak* (XOMAK, 2000:418), etc.<sup>33</sup> Finally, <’yw̄k> is an example of the ‘pseudohistorische Orthographie’ of Pahlavi (KLINGENSCHMITT, 2000:194): it is based on <’yw̄> /ēw/, and Av. *aēva* ‘one’.

## 9. Other contemporary forms

### 9.1. Contemporary Iranian Persian *ye* and *yek*

Contemporary Persian *ye* ‘one, a’, goes back to a more archaic form than what is usually assumed (see 8.2.). An interesting point to note is that, in the standard variety of contemporary Persian, /jak/ is almost never heard: /jek/ is

<sup>30</sup> For more examples, see HÜBSCHMANN, 1895:121.

<sup>31</sup> To my knowledge, of all the New Persian literature, only the very archaic JP commentary on Ezechiel has <’y’d> ‘memory’ (cf. LAZARD, 1963:132<sup>37</sup>).

<sup>32</sup> Early JP is constituted of texts written between the 8<sup>th</sup> and the 13<sup>th</sup> centuries CE (PAUL, 2013:9).

<sup>33</sup> Both are likely loanwords from NP, as per the Sistani sound law *#ya- > #a-* (see fn 7).

much more common. This could of course be ascribed to the *a ~ e* phonetic alternation of NP (PISOWICZ, 1985:15), but since [jek] is fairly recent (almost all Iranian languages that loaned their word for ‘one’ from Persian, such as Balochi and Gavruni, have *yak*), it is possible that *yak* took its vowel from *e#*, after the *a# > e#* sound change, which probably happened in the 17<sup>th</sup> century, that is *ye* → *yak* > *yek*.

### 9.2. *yāzdah* ‘eleven’

The expected reflex of OIr. *\*aiyandasa* ‘eleven’ should have been NP †*ēndah*, cf. Parthian /*ēwandas*/ (DMMPP:101). The form *yāzdah* (discussed in fn 17) MP <y’zdh>, is from *\*aiyājazdasa*. For a discussion of the length of the vowel (against expected †*yazdah*) see SCHMITT (1994:19-20), who cites previous literature. An analogy with ‘12’ (< Proto-Iranian *\*duā-datsa*) is often mentioned to explain the vowel length.

### 9.3. Balochi *ēyōk* and *ēwak* ‘alone’

*\*aiya-ka* became *ēwak* ‘alone’ in most Bal. varieties (KORN, 2005:103, 105, 151; KORN 2006:204), cf. Parthian *ēwag* ‘solely, only’ (DMMPP:101). Nevertheless, there is a variant (perhaps coastal, see KORN, 2005:243<sup>325</sup>) *ēyōk* ‘alone’ (but HORN, 1893:252 ‘einzeln’) about which KORN (2005:243) writes: “[...] *ēyōk* ‘alone’ [...] vs. usual *ēwak* may be explained as showing loss of the second vowel, thence /*ēōk*/.” This would be the only one of two examples of the loss of vowel happening between Common Balochi and contemporary varieties of Balochi.

Another etymology could perhaps account for this form: Common Balochi *\*ēwyōk* from *\*aiy(a)ia-* with the *-ōk* suffix (see KORN, 2005:163), which must be present in this word, with subsequent simplification of the cluster *\*wy* as *y*. It is thus possible that the processes of analogy leading to a proto-form *\*aiyāia-ka-*, happened in multiple Iranian languages. *\*ēy* would thus be ‘simple, unique’, to which an agentive suffix *-ōk* is added. This, in my opinion, accounts very well for the meaning ‘alone’. The simplification of /wy/ in word internal position could have had different results in different Balochi dialects, perhaps leading to the form *ēwak*.

This etymology also raises the question of whether *ēwak* was borrowed into Balochi from Parthian, with regular devoicing of the word final consonant in loanwords (KORN, 2005:268-71) and that *ēyōk* is the inherited Balochi word. There is perhaps an argument in favor of this in the fact that post vocalic *\*u* is lost in *pīg* ‘fat’ (cf. Av. *pīuuah-*), yet is retained in *ēwak* ‘alone’ (KORN, 2005:103). If *ēyōk* is indeed the inherited form, then it undergoes, as does *pīg*, the loss of postvocalic *\*u*.

Since numerals of many Iranian languages were strongly influenced or taken from Persian, it is thus difficult to know exactly what happened in those languages, from a language-internal perspective.

## 10. Conclusion

It seems that the etymological solutions provided by HÜBSCHMANN (1895:151), HORN (1898-1901:44) and HASANDUST (2014:2947) on the one hand, and by KLINGENSCHMITT (2000:212) on the other, to explain the (M)MP and NP form *yak* ‘one’ are unsatisfactory. In the present article, two different solutions are presented to explain the origin of *yak*.

The first scenario implies an analogy from *\*duaja(h)-* ‘double’, or possibly *\*duaja-ka-*, yielding *\*aijuaja-(ka-)*, which dissimilated to *\*iūaja-ka-* ‘unique, single’. In this scenario, the *-ka-* derivation is not semantically relevant. From the meanings of ‘unique’ and ‘single’ the word acquires the meaning of ‘one’, possibly through negated phrases such as ‘not a single (one)’. The older word for ‘one’, *\*aiua-*, now realized as *\*ēw*, is pushed towards the function of an indefinite marker, first independent and then enclitic, due to the new function of *\*iūaja-ka-* > *\*iūak*. The second scenario is based on a possible analogy from ‘two’ to ‘one’, yielding *\*aijuaja-*. This kind of analogy finds a parallel in Ossetic. The new form was later suffixed with *-ka-* to disambiguate it from a potential derivative in *\*-ja-*, meaning ‘unique, simple’. The second scenario is more complex and requires more hypotheses than the first one, but I believe it is still possible.

The claims made in this article are not irrefutable, but in light of the difficulties with the derivation of Middle and New Persian *yak*, I hope to have provided two plausible etymologies of *yak*, thereby shedding new light on the (Proto-)Persian numeral system.

## BIBLIOGRAPHY

- ANTIÂ, (Ervad) Edalji, Kersâspji (1909), *Pâzend Texts*. Bombay: Duftur Ashkara Press.
- BARTHOLOMAE, Christian (1895-1901), “Vorgeschichte der Iranischen Sprachen” in GEIGER, Wilhelm & KUHN, Ernst (eds.), *Grundriss der Iranischen Philologie*. Erster Band. 1. Ableitung. Strassburg: Verlag von Karl J. Trübner, 1-151.
- BLAU, Joyce (1975), *Le Kurde de ‘Amādiya et de Djabal Sindjār*. Paris: Klincksieck.
- CHEUNG, Johnny,  
— (2002), *Studies in the Historical Development of the Ossetic Vocalism*. Wiesbaden: Dr Ludwig Reichert Verlag.

- (*in prep.*) “A few remarks on the origin of the Kurdish indefinite/definite category”.
- CPD = MACKENZIE, David Neil (1986), *A Concise Pahlavi Dictionary*. London: Oxford University Press.
- DMMPP = Durkin-Meisterernst, Desmond (2004), *Dictionary of Manichean Middle Persian and Parthian*. Turnhout: Brepolis.
- EWAia = MAYRHOFFER, Manfred (1992), *Etymologisches Wörterbuch des Altindoarischen*, Band I. Heidelberg: Carl Winter – Universitätsverlag.
- GEIGER, Wilhelm (1893), *Etymologie und Lautlehre des Afghānischen*. München: Verlag der k. Akademie.
- GLASSMAN, Eugene H. (1972), *Conversational Dari*. Edited by Susan SMITH. Peshawar: InterLit Foundation.
- HASANDUST, Mohammad (2014), *Farhang-e Riše-šenāxti-ye Zabān-e Fārsi* (Etymological Dictionary of the Persian Language). Tehran: Ātār.
- HÜBSCHMANN, J. Heinrich (1895), *Persische Studien*. Strassburg: Verlag von Karl J. Trübner.
- HORN, Paul,  
 — (1893), *Grundriss der neupersischen Etymologie*. Strassburg: Verlag von Karl J. Trübner.  
 — (1898-1901), “Neupersische Schriftsprache” in GEIGER, Wilhelm & KUHN, Ernst (eds.), *Grundriss der iranischen Philologie*. Erster Band. 2. Ableitung. Strassburg: Verlag von Karl J. Trübner, 1-200.
- JOSEPHSON, Judith (2011), “Definiteness and Deixis in Middle Persian” in MAGGI, Mauro and ORSATTI, Paola, *The Persian Language in History*, Wiesbaden: Dr. Ludwig Reichert Verlag.
- KLINGENSCHMITT, Gert (2000), “Mittelpersisch” in FORSSMAN, Bernhard & PLATH, Robert (eds.) *Indoarisch, Iranisch und die Indogermanistik, Arbeitstagung der Indogermanischen Gesellschaft vom 2. bis 5. Oktober 1997 in Erlangen*. Wiesbaden: Dr. Ludwig Reichert Verlag, 191-229.
- KORN, Agnes,  
 — (2005), *Towards a Historical Grammar of Balochi*. Wiesbaden: Dr. Ludwig Reichert Verlag.  
 — (2006), “Counting Sheep and Camels in Balochi”, in BOGOLJUBOV, Mixail et al. (eds.) *Indoiranskoe jazykoznanie i tipologija jazykovyx situacij. Sbornik statej k 75-letiju professora A. L. Grjunberga*. St. Petersburg: Nauka, 201-212.  
 — (2016), “A partial tree of Central Iranian”, *Indogermanische Forschungen*, 121, 401-434.
- KÜMMEL, Martin J. (2018), “The Survival of Laryngeals in Iranian” in VAN BEEK, Lucien, KLOEKHORST, Alwin, KROONEN, Guus, PEYROT, Michaël, PRONK, Tijmen (eds.), *Farnah: Indo-Iranian and Indo-European Studies in Honor of Sasha Lubotsky*. Ann Arbor – New York: Beech Stave, 162-172.

- LAZARD, Gilbert (1963), *La Langue des plus anciens monuments de la prose persane*. Paris: Librairie C. Klincksieck.
- MILLER, Vsevolod F. (1881-1887), *Osetinskie Ètjudy*, I-III, Moskva.
- MOEIN, Hamid (2012), *On Pāzand: Philological Comparison with Pahlavi*. Masters thesis. Montréal.
- NYBERG, Henrik S. (1974), *A Manual of Pahlavi, volume II*. Wiesbaden: Otto Harrassowitz.
- PAUL, Ludwig (2013), *A Grammar of Early Judaeo-Persian*. Wiesbaden: Dr. Ludwig Reichert Verlag.
- PISOWICZ, Andrzej (1985), *Origins of the New and Middle Persian Phonological Systems*. Kraków: Nakładem Uniwersytetu Jagiellońskiego.
- SALEMANN, Carl H. (1895-1901), “Mittelpersisch” in GEIGER, Wilhelm & KUHN, Ernst (eds.), *Grundriss der Iranischen Philologie*. Erster Band. 1. Ableitung. Strassburg: Verlag von Karl J. Trübner, 249-332.
- SCHMITT, Rüdiger,  
 — (1994), “Die Zahlreihe zwischen „10“ und „20“, zum Beispiel im Iranischen” in *Historische Sprachforschung*, 107. Bd., 1. H., 12-29.  
 — (2014), *Wörterbuch der altpersischen Königsinschriften*. Wiesbaden: Dr. Ludwig Reichert Verlag.
- SIMS-WILLIAMS, Nicholas (*forthcoming*), “Dual and Numerative in Middle and New Iranian”.
- STEBLINE-KAMENSKY, Ivan M. (1999), *Etymological Dictionary of the Wakhi Language*. Saint-Petersburg: Peterburgskoye Vostokovedenye.
- WOLFF, Fritz (1935), *Glossar zu Firdosis Schahname*. Berlin: Reichsdruckerei.
- XOMAK, Javād-Mohammad (2000), *Vāže-nāme-ye Sagzi* (Dictionary of Sistāni). Tehran: Soruš.

RÉSUMÉ.— Le nombre persan *yak* « un » ne peut, pour des raisons phonétiques, dériver du vieil-iranien *\*aiya(-ka-)*, comme il a été proposé auparavant. Le présent article se propose d’offrir deux nouvelles hypothèses afin d’en expliquer la dérivation : la première est une analogie avec *\*duaja(-ka-)* « double ». La seconde se base sur la forme proto-iranienne du nombre « trois », *\*θrajah*, qui engendrerait une série d’analogies et de développements sémantiques menant à une proto-forme qui pourrait expliquer *yak*.



