

Exaptation from Arabic syntax to Persian lexical Morphology

John R. Perry
The University of Chicago
j-perry@uchicago.edu

It has long been natural for linguists to invoke metaphors taken from the life sciences, in particular evolutionary biology. In recent decades it has been argued that such analogies represent not mere rhetoric or metalinguistic convenience, but actual processes in language development; after all, why should not a biologically-conditioned cultural phenomenon such as language, the collective product of a life form, replicate aspects of a form of life – and not just general processes such as evolution and extinction, but actual details of the process, such as natural selection and exaptation?

“Exaptation” is a biological term coined by Stephen J. Gould and Elisabeth Vrba in a 1982 article¹. It may be defined as “a case where an anatomical structure that originally evolved to serve one function was later commandeered to facilitate a quite different function.” For instance, feathers in the proto-avian dinosaur lineage demonstrably evolved before the capacity for flight; their purpose must have been something else, such as thermo-regulation (to keep them warm), or display (to attract a mate) – functions which additionally continued in many cases. However, this structure was fortuitously available when it was later modified to provide flight-control surfaces. Darwin anticipated this process in 1859, applying the term “pre-adaptation,” and citing the example of a fish’s swim-bladder, as having originally evolved for flotation, and in land animals being converted to a wholly different purpose – that of respiration, in the form of lungs².

In linguistics the term has been adopted by, e.g., Roger Lass in a 1990 article³, and by Laura A. Janda in *Back from the brink: a study of how relic forms in languages serve as source material for analogical extension* (Lincom Europa, 1996). My example of the redeployment of a distinctive feature to a novel purpose (which I investigated in the 1980s, before I had heard of exaptation) involves the successful hijacking, by semantic determinants of Persian lexical morphology, of a syntactically-conditioned phonological alternation in the feminine ending of Arabic nominals. The human motivation was the desire, during the development of the Eastern Islamic Kulturgebiet between the seventh and twelfth centuries (see the map, fig. 4) to incorporate useful or prestigious Arabic vocabulary into Persian, using the Arabic writing system (which had been adopted in Persian) but adhering to Persian phonotactics and lexical morphology. The junk element in Arabic feminine-ending nouns and adjectives was not so much that Persian had no grammatical gender, but that it had no use for the typically Semitic syntactic structure known as the “construct state”: this requires that a feminine ending be pronounced as /-at/ with terminal -t when its nominal is the head of a NP modified by a following noun (“pre-juncture position”), and /-a/ in all other situations (“pre-pausal position”). Thus *dawlat al-Sūdān* ‘the state of Sudan’, but *ra’īs al-dawla* ‘head of state’ and *dawla mustaqilla* ‘an independent state’. In each case, the feminine marker was written with an invariant hybrid graph in Arabic. Now, Persian speakers needed definitively to lexicalize a single form of a word as either -at or -a.

The solution they devised has led to an inventory of at least 1400 Arabic Feminine Ending (AFE) loanwords in the modern Persian lexicon, in a ratio of roughly 600 -at: 800 -eh, including forty doublets – i.e., copies of the same word in each form, written with distinctive graphs (final t, and final h for spoken /-a/) and two lexically distinct meanings (c. fig. 3; I use -eh to represent the vocalic termination as being visually quite distinct from -at, and to

¹ “Exaptation – a missing term in the science of form,” *Paleobiology* 8 (1982): 4-15.

² See *The Origin of Species* Chapter VI, under “Modes of Transition.”

³ “How to do things with junk: exaptation in language evolution,” *Journal of Linguistics* 26 (1990): 79-102.

incorporate the incidental vowel change that marks this form in Standard Persian). A simple example, involving a doublet, is: *āyat* ‘sign, portent’ vs. *āyeh* ‘verse of scripture’.

Looking for a rationale behind the distribution of these allomorphs, it struck me how utterly divorced were the current Persian function of the *-at* vs *-eh* dichotomy (as lexical markers, the significance of which was yet to be described) and the Arabic syntactic alternation *-at* ~ *-a*. The anatomical location, so to speak, was identical, but the process of recycling and the new rationale had remained opaque. The forms are a salient feature of Persian’s large Arabic loanword inventory, but one that had so far been taken for granted. Neither Iranian nor foreign scholars (except for one of each, with very limited results) had even taken note of the puzzle; it was indeed the case (to paraphrase Gould & Vrba) that “Current functions cannot be used to infer past causal pressures.” After collecting the two inventories, I was able first to eliminate any copying of Arabic syntagmata (such as nominals in the construct state, noun adjective NPs, prepositional phrases, etc.) as a source for the loanword patterns. Construct state NPs were quite commonly lifted into Persian *in toto*; but the pre-pausal feminine ending in, e.g., *fawq al-‘āda* ‘extraordinary’, does not correspond to the canonical lexical form *‘ādat* ‘custom, norm’ in Persian; nor does the pre-juncture form in the epithet *sayf al-dawla* ‘Sword of the State’ correspond to the lexicalized Persian *dawlat* ‘state’.

Secondly, I eliminated the cognate loanwords in Malay-Indonesian, various African languages, and Spanish, and determined that this regular pattern of distribution between *-at* and *-eh* was exclusive to the Persianate world, i.e., to those languages which had received their Arabic lexical copies in pre-modern times through the medium of Persian – mainly Iranian, Turkic, and Indic languages covering the shaded area of the map (fig. 4).

Thirdly, I realized that the process had been a dynamic one, both diachronically and geographically. On the evidence of Persian literature of the 10th-11th centuries, the initial ratio of incorporation into Persian showed a preponderance of *-at* over *-eh* affiliates of approx 6:1, as against a modern preponderance of 4:3 *-eh* over *-at*. So, over the intervening 1000 years, there had evidently been a considerable shift in individual items from the *-at* inventory to *-eh*. From an areal perspective, the basic Persian pattern of distribution, and the rate of shift, were not 100 percent the same in inventories of cognate copies in other languages of the area. Turkic dialects especially, showed an idiosyncratic dynamic, with instances of shift and doublets clearly based on the same rationale as in Persian, but targeting different words. Indic languages also demonstrated some independence, but in general were more conservative than Turkish and Persian, retaining more original *-at* affiliates.

Finally, what is the rationale for the dichotomy in these two series and their patterning? We may gain an overview of it by shifting our attention between fig. 1, which summarizes the principal linguistic and sociolinguistic features of the system, and fig. 2, where these are exemplified by the location of typical AFE loanwords in the Persian semantic and lexical spectrum. This “semantic spectrogram” maps words across a range of lexical, syntactic, and sociolinguistic usage, from most abstract and least specialized to most concrete and/or specialized. The illustrations in fig. 2 are singletons, i.e., borrowings which remained true to their respective incorporation in *-at* or *-eh* without formal change.

In the case of loanwords which in Arabic were deverbal or deadjectival derivatives, “specialization” may involve re-verbalization as a compound verb (+V) and/or acquisition (whether contextual or permanent) of a particular extension of the verbal sense, such as a noun of instance (English *sleeping*, *sleep* are action nouns, [*a*] *sleep* or *nap*, pl. *naps* are instance nouns), a product noun (as Eng. [*a*] *collection*) or instrumentive or agentive nouns (as *cooker*, *cook*). Thus the same nominal may occupy progressively more tangible, imageable, and countable semantic slots so as to approach the concrete polarity of an entity noun. The hatched line separating the mean point between clusters of *-at* and *-eh* affiliates in these figures I call the “Semantic Watershed.” It illustrates the comparative density and location of *-at* and *-eh* copies across the semantic spectrum.

Bearing in mind the features sketched in the three sections of fig. 1, what stands out intuitively from the meaning of the forms listed in fig. 2 is the following.

(1) From a semantic and syntactic perspective, *-at* marks words for abstract, intangible, low-imagery referents. Here the *-at* affiliates constitute mainly quality and action nouns, as nos. 5 *šo'ûbat* 'difficulty', 10 *hokumat* 'government'. Some may have expanded their semantic range for use as nouns of instance (fourth column: 'a [particular instance of] difficulty, a problem'); or, more rarely, as product, agentive, or instrumentive nouns ('agency that governs; the persons so constituted'). These are then count nouns: 'difficulties, governments'. Though they are more 'imageable', i.e., readily visualized via a speaker's or hearer's particular experiences, they seldom extend as far as tangibilia, or entity nouns (last three columns).

Conversely, *-eh* tends to mark concrete, tangible, high-imagery deverbals such as product nouns, etc.: 1 *moqaddemeh* 'preface, introduction [to a book, etc.]' (instrumentive), 4 *mervahēh* 'instrument for creating a draft', 5 *raqqāšeh* 'dancing girl' (agentive). In everyday Persian usage, the deverbal function encoded in the Arabic morphological pattern may be opaque, and nouns such as this last will fall intuitively into the class "entity noun," the same as nos. 8 'town, city' (and the name of a particular city), 9 'name of a female', 10 'type of camel'.

(2) Sociolinguistically, *-at* marks unspecialized, and *-eh* specialized terms in various ways. Stylistically, those originally incorporated in *-at* as learned words (*mots savants*) often remain in the higher, literary register; some may be characterized as "Classical," i.e., archaic or imperfectly assimilated in modern Persian (*Fremdwörter*). Those that were incorporated in the *-eh* series are (or were) common in the vernacular register (and tend, of course, to be countable and to correspond to concrete and entity nouns). Those which, we might guess, were originally transmitted orally turn up as *-eh* (such as names: no. 9 *Xadijeh*, and the toponym *Maymaneh*² in fig. 3 no. 9; common entity nouns: no. 10 *jam(m)āzeh* 'dromedary').

(3) Questions of diachronic shift *-at* > *-eh*, the generation of doublets, and differential assimilation in different languages (as limned in the third section of this epitome), are best illustrated in condensed form via fig. 3 (Doublets). To begin with a minor lexical point that is not captured in fig. 1: contrastive affiliation can, at its simplest, disambiguate what were homonyms in Arabic (i.e., a mere coincidence of root and surface forms), such as fig. 3, no. 8 *šārārat* 'evil, wickedness' vs. *šārāreh* 'spark'. Homophones and homographs in Arabic, these non-cognate words find themselves, as Persian loanwords, appropriately differentiated in sound and form, and parked at opposite ends of the semantic spectrum in accordance with their affiliation in *-at* or *-eh*. More subtly, no. 4 *mānaviyat* 'Manichaeism' and *Mānaviyeh* '(the) Manichaeans' (collectively, as a sect or community) demonstrates different uses of the feminine ending in Arabic as a quality noun ('Manichaeism-ness') and as a collective noun formative; the resulting surface homonymy in Arabic is disambiguated in Persian by the assignment to, and marking of, separate semantic slots appropriate to *-at* and *-eh*.

No. 6 *mas'ala*, in Arabic, likewise represented not simply two different denotations of the same lexical pattern, but two semantically distinct lexical patterns. As discrete Persian copies, the Classical Persian (CP) action noun *mas'alat* 'asking, questioning' (a *mot savant*) soon took a back seat to the everyday instance and product noun *mas'aleh* 'question, problem, issue, matter; thingy' (a vernacular euphemism for the male sexual organ). Equally ingenious is the doublet pair no. 2 *ešāriyat/-eh*, both being late Arabicate neologisms in Indo-Persian or even Urdu. The abstract 'symbolism' contrasts with the instrumentive 'index' (a count noun, readily visualized, and even tangible when printed). These neologisms were derived from Perso-Arabic doublet pair no. 1 *ešārat/-eh* 'showing; indication, gesture, sign'. The semantic extension in the direction of specialization, imageability, tangibility and vernacularity that prompted the shift *-at* > *-eh* is evident: Classical Persian *ešārat* is a verbal abstract, 'the showing, demonstration'; as +V it is re-verbalized in conjunction with an auxiliary ('to show, indicate, gesture'), and soon shifts in this function to *ešāreh*; but it is so marked also as a count noun of instance/product, 'pointer, gesture, sign'.

No. 3 *mo'ādelat* was similarly abstract, 'equivalence, balancing' when first copied into Persian. In Ottoman Turkish it shifted to *mo'ādeleh*, in the product noun sense 'equation'. Likewise restricted to Turkish is the pair no.5 *hāarakat* 'movement' (an action

noun) and its shifted doublet *ḥarekeh* ‘vowel sign’ (a grammatical device and the written symbol for this; modern orthography *hareket*, *hareke*; Persian uses *-at*).

No. 7 *ta'ziat* was copied into Classical Persian with the meaning ‘condolence, mourning’. In Turkish this shifted to *-eh*, in the more concrete and vernacular sense of ‘obsequies, funeral’; in the usage of Shi'i Turkmen and Iranians of the fifteenth century, it specialized as ‘Moharram mourning rites for Imam Hosayn’; returning to Persian as a *Rückwanderer*, the word was further restricted to the sense ‘Passion play’ (in Persian), and later copied into Indo-Persian with reference to a ritual object, the model of the martyr's tomb carried in procession.

In no. 9 we encounter a neat tripartite distinction of three meanings of Arabic *maymana*, a homonymous doublet from the trilateral root *YMN* ‘right (side); being of good omen’. The word of the two that is an action noun (from a stative verb, hence in effect a noun of condition or quality, ‘wellbeing, prosperity’) is logically affiliated in Persian in the *-at* series. The other word, a locative ‘right wing (of an army)’, as a specialized count noun derivative of the plain, non-metaphorical sense of *YMN*, is allocated to the *-eh* series as *maymaneh*¹. Then there is also a city in Afghanistan called Maymana (*maymaneh*²); is this in origin ‘located on the right (side, wing)’, or a ‘place of prosperity’? Whichever it be, the form is appropriate.

In the course of the next several centuries, hundreds of the *-at* class shifted to the *-eh* class, some leaving behind traces as doublets in *-at*. (Or, to rephrase this in more sociolinguistic terms, a shortened form with a more specialized, imageable connotation evolved in the vernacular, which soon complemented, or superseded, the literary word). In general, the sound (and orthographic) change recapitulates the original rationale: the resulting *-eh* words are semantically more specialized or concretized (cf. fig. 3 no. 3 *mo'ādeleh* ‘equation’, or no. 5 *ḥarekeh* ‘vowel sign’), and/or more firmly established in the vernacular (cf. no. 6 *mas'aleh* ‘question, matter, problem’, and no. 7 *ta'zieh* ‘Moharram passion play; cenotaph’).

There are of course apparent exceptions to the trends observed here. *Nomina actionis* especially (by definition abstract, less imageable, in their basic meanings) seem to be more arbitrarily apportioned; even of these, however, those ending in *-eh* tend to form common compound verbs in Persian (i.e., are more imageable in context, and frequent or vernacular in usage), and have also evolved count-noun referents (no. 1 *ešāreh*). This shift appears to have peaked by the late thirteenth century, by which time the majority of the Arabic loanwords that are in general use today had been incorporated in Persian, and were being transmitted to Turkish, Hindi, etc. Moreover, as is evident from the examples, not only were individual loanwords incorporated into Persian then passed on to nearly all the other languages of mainland Muslim Asia, but the intuitive rules for this binary sorting were transmitted with them, to be used innovatively in the recipient languages.

To conclude: a fortuitous syntactically-triggered dichotomy in a portion of the inventory of Arabic substantives that were copied into Persian was reanalyzed to furnish a sub-system of semantic sorting and lexical expansion in the languages of a significant cultural area. This exaptive morphological redeployment of the Arabic feminine ending in Persian was still active in Persianate cultures at least until the first decades of the twentieth century: in evidence we may cite neologisms such as Urdu *ešāriyat* and *ešāriyeh*, Persian *e'lāmiyeh* ‘manifesto’, Turkish *eṭfā'iyeh* ‘fire service’ and *melliyat* ‘nationalism’ (also adopted in Persian), and (Soviet) Tajik *partiaviyat* ‘party loyalty’. It has ceased to be productive, except to a limited extent in Urdu of Pakistan, since Arabic ceased to be an active source of vocabulary after the language reform movements in Turkey, Iran, and Soviet Muslim Asia during the second quarter of the twentieth century.

Further references to works by J. R. Perry on this topic

1984a “*-at* and *-a*: Arabic Loanwords with the Feminine Ending in Turkish.” *Turkish Studies Association Bulletin* 8/2: 16-25.

1984b “Differential Assimilation of some Arabic Loanwords in Tajik and Uzbek.” *Folia Slavica* Vol. 7, nos. 1-2: 268-82.

- 1984c “Semantic Sorting and Morphological Shift in a Loanword Category.” *Journal of Historical Linguistics and Philology* (University of Maryland) Vol. I, No. 2: 31-53.
- 1989 “Lexical Doublets and Triplets in the Turkic Languages of the USSR,” in Howard L. Aronson, ed., *The Non-Slavic Languages of the USSR: Linguistic Studies*. CLS, University of Chicago: 176-186.
- 1991 *Form and Meaning in Persian Vocabulary: The Arabic Feminine Ending*. Bibliotheca Persica Persian Studies Series No. 12. Mazda Publishers: Costa Mesa, CA. [254 pp.]
- 2004 “Lexical areas and semantic fields of Arabic loanwords in Persian and beyond,” in Éva Á. Csató, Bo Isaksson and Carina Jahani (eds), *Linguistic convergence and areal diffusion: Case studies from Iranian, Semitic and Turkic*. RoutledgeCurzon: 97-110.

Fig. 1. Epitome of the *t.m.* Series Dichotomy

	-at	-eh
Semantic & Syntactic Features	Abstract; <i>and/or</i> intangible, insubstantial, low-imagery. Mass noun +VN (Nominalization of VP)	Concrete; <i>and/or</i> Tangible, substantial, high imagery. Count noun +N; +V (Reverbalization)
Sociolinguistic Features	Mot savant; <i>Fremdwort</i> High register; <i>and/or</i> Classical, literary; ?higher written frequency	Lexically assimilated Low register; <i>and/or</i> modern, vernacular; ?higher spoken frequency
Diachronic & Dialectal Features	Abstract, etc. Mass noun +VN Mot savant > <i>Fremdwort</i> High register, etc. <i>But also:</i> written > spoken	> Concrete, etc. > Count noun > +N; +V (reverbalization) > Lexically assimilated > Low register, etc. written < spoken

Fig. 2. Semantic Spectrogram: Single-Series Loans in Persian

	Qual. N.	Action Noun		Inst. N.	Product Noun, &c.		Entity Noun			
		+N	+V		Int.	Tan.	Mas	Cnt		
1. <i>oxovvat</i>	t				h	h		h	<i>moqaddemeh</i>	1.
2. <i>zarurat</i>	t				h	h		h	<i>majmu'eh</i>	2.
3. <i>morovvat</i>	t					h		h	<i>'amaleh</i>	3.
4. <i>nobovvat</i>	t					h		h	<i>mervaheh</i>	4.
5. <i>šo'ubat</i>	t			t		h		h	<i>raqqāseh</i>	5.
6. <i>xošunat</i>	t	t		t				h	<i>'olufeh</i>	6.
7. <i>kodurat</i>	t	t		t				h	<i>xārejeh</i>	7.
8. <i>sokunat</i>	t	t	t	t				h	<i>madīneh</i>	8.
9. <i>'oqubat</i>	t	t		t				h	<i>xadijeh</i>	9.
10. <i>hokumat</i>	t	t	t	t	t			h	<i>jam(m)āzeh</i>	10.

GLOSSES: *-at* 1. 'brotherhood'; 2. 'necessity'; 3. 'manliness'; 4. 'prophethood';
5. 'difficulty'; 6. 'harshness'; 7. 'turbidity'; 8. 'residence'; 9. 'punishment';
10. 'government'.

-eh 1. 'preface'; 2. 'compendium'; 3. 'worker'; 4. 'fan'; 5. '(female) dancer';
6. 'fodder'; 7. 'foreign parts'; 8. 'town; Medina'; 9. 'Khadija'; 10. 'dromedary'.

Fig. 3. Semantic Spectrogram: Sample Doublets in Persian, Turkish and Urdu (P, T, U)

	Qual. N.	Action Noun		Inst. N.	Product Noun, &c.		Entity Noun		
		+N	+V		Int.	Tan.	Mas	Cnt	
1. <i>ešārat</i>		t	h	h	h	h			<i>ešāreh</i>
2. <i>ešāriyat</i> (U)	t				h	h			<i>ešāriyeh</i> (U)
3. <i>mo'ādelat</i> (T)	t	t			h				<i>mo'ādeleh</i> (T)
4. <i>mānaviyat</i>	t						h		<i>mānaviyeh</i>
5. <i>ḥarekat</i>		t	t	t	h	h			<i>ḥarekeh</i> (T)
6. <i>mas'alat</i> (CP)		t	t		h			h	<i>mas'aleh</i> (P)
7. <i>ta'ziat</i> (P)		t	t		h			h	<i>ta'zieh</i> (T, P, U)
8. <i>šarārat</i>	t							h	<i>šarāreh</i>
9. <i>maymanat</i>	t	t						h	<i>maymaneh</i> ₁
								h	<i>maymaneh</i> ₂

GLOSSES: 1. 'indication'/'gesture; sign'; 2. 'symbolism'/'index'; 3. 'equivalence'/'equation'; 4. 'Manichæism'/'the Manichæans'; 5. 'motion, movement'/'vowel sign'; 6. 'asking'/'problem; thingy'; 7. 'condolence'/'obsequies' (T), 'religious drama' (P), 'cenotaph' (U); 8. 'wickedness'/'spark'; 9. 'prosperity'/'right wing; name of town'.

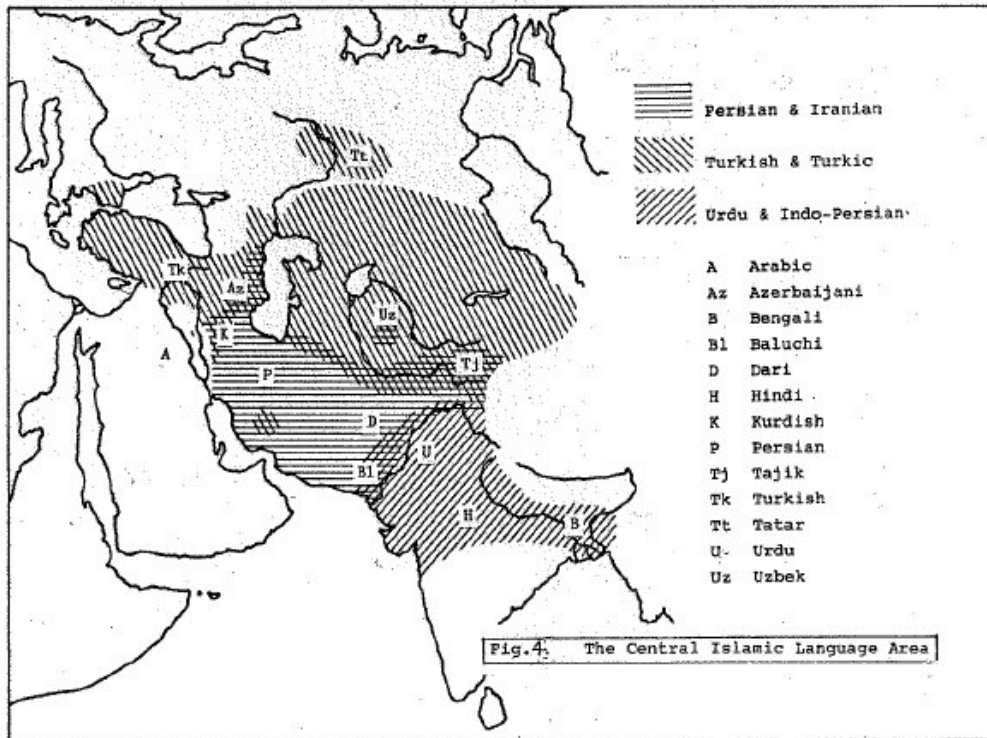


Fig.4. The Central Islamic Language Area