Acta Linguistica Asiatica

Volume 2, Number 1, May 2012

ACTA LINGUISTICA ASIATICA
Volume 2, Number 1, May 2012

Published by: Znanstvena založba Filozofske fakultete Univerze v Ljubljani
(Ljubljana University Press, Faculty of Arts)

Issued by: Department of Asian and African Studies

For the publisher: Andrej Černe, the Dean of the Faculty of Arts

Journal's web page:
http://revije.ff.uni-lj.si/ala/
Journal is published in the scope of Open Journal Systems

ISSN: 2232-3317

Abstracting and Indexing Services:
COBISS, Directory of Open Access Journals, Open J-Gate and Google Scholar.

Publication is free of charge.

Address:
University of Ljubljana, Faculty of Arts
Department of Asian and African Studies
Aškerčeva 2, SI-1000 Ljubljana, Slovenia

E-mail: mateja.petrovcic@ff.uni-lj.si

TABLE OF CONTENTS

FOREWORD

With the present issue of ALA, we are starting the second year in its new incarnation. As the paper is striving to cover the Asian languages in their multiplicity and in the multiplicity of approaches, I am glad to say that this issue offers both, in line with Roman Jakobson's famous paraphrase "*Linguista sum, linguistici nihil a me alienum puto.*"

Among the six papers in this issue, two papers deal with Japanese, two with Iranian langages, and one each with the language of Rigveda and with Arabic. Also, approaches vary from historical phonetics/phonology (paper by John KUPCHIK on the role of *rendaku* in Eastern Old Japanese poetry) and historical syntax (Tamara DITRICH's discussion of coordinative particles in *Ṛgveda*), to typological considerations spannig diachroinc and synchronic views (Yadgar KARIMI's analysis of the evolution of ergative in Iranian languages). Syncronic approaches comprise pragmatics (Biook BEHNAM and Salam KHALILIAQDAM's treatment of hedgind devices in Kurdish and Robert Michael BIANCHI's account of the new hybrid language of 3arabizi), and interdisciplinary interpretation of lexis in the light of cross-cultural psychology (Márton SZEMEREY's paper comparing Japanese and Hungarian linguistic resources for expression of emotions).

Andrej Bekeš

RESEARCH ARTICLES

MORPHEME-BASED *RENDAKU* AS A RHYTHMIC STABILIZER IN EASTERN OLD JAPANESE POETRY

John KUPCHIK
Kyoto University – JSPS
johnkupchik@gmail.com

Abstract

This paper explores the functions of morpheme-based *rendaku*, or "sequential voicing", in Eastern Old Japanese poetry, with a focus on its function of maintaining rhythmic stability in poetic verse. It is argued that this function is implemented to avoid a hypermetrical line when no adjacent vowels exist as candidates for synchronic elision. Furthermore, a comparison with synchronic vowel elision is conducted. Based on the results, it is argued that morpheme-based *rendaku* is preferred to synchronic vowel elision when both are available options for maintaining the rhythmic stability of a line. Linguistic constraints blocking morpheme-based *rendaku* are also discussed to explain hypermetrical examples with potential, yet unrealized, morpheme-based *rendaku*.

Keywords

phonology; phonetics; Old Japanese; *rendaku*; poetry

Izvleček

Raziskava se osredotoča na funkcijo t.i. "sequential voicing" ali zvenečnjenja na nivoju morfemov v stari vzhodno-japonski poeziji, in sicer daje poudarek funkciji, ki ohranja ritmično stabilnost verza. Avtor pokaže, da se v primeru, ko ni okoliških samoglasnikov, ki bi sodelovali pri sinhronem izpadu samoglasnikov, z uporabo te funkcije izognemo hipermetrični vrstici. Vzporedno avtor primerja zvenečnjenje in sinhroni izpad samoglasnikov in ugotovi, da je v primeru, ko sta za ohranjanje ritmične stabilnosti možni obe vrsti posredovanja, zvenečnjenje bolj željena izbira. Nazadnje so analizirane lingvistične omejitve, ki preprečujejo uporabo zvenečnjenja, in s tem tudi hipermetrični primeri s hipotetičnim obstojem zvenečnjenja.

Ključne besede

fonologija; fonetika; stara japonščina; zvenečnjenje; poezija

1. Introduction[1]

The Eastern Old Japanese (EOJ) dialects were spoken in the first half of the 8th century CE in Kamitukeno, Sagamu, Simotukeno, Muzasi, Pitati, Michinöku, Simotupusa, and Kamitupusa provinces[2], all of which were located in the Eastern (Azuma) region of Japan. The language is markedly different in many aspects of grammar and lexicon compared to Western Old Japanese (WOJ), Central Old Japanese, and Töpo-Suruga Old Japanese, all of which were spoken in Japan at the same time.

EOJ is attested exclusively in poetry in books 14 and 20 of the *Man'yōshū* poetry anthology, and has been an important subject of interest for Japanese linguists for over half a century, beginning with the early seminal works by Fukuda (1965), Hōjō (1966), and Mizushima (1972, 1984a, 1984b, 1986) and continuing on in recent studies by Mizushima (1996, 2003, 2005, 2009), Vovin (2005, 2009a, 2009b), and Kupchik (2011). Also notable are the important commentaries by Nakanishi (1981, 1983), Omodaka (1984a, 1984b), Itō (1997, 1998), and Aso (2011). Long heralded as a language that contains crucial pieces of information necessary for the reconstruction of Proto-Japonic and Proto-Japanese, much previous linguistic research has focused on its array of retentions, some of which are shared with Ryukyuan languages (Pellard, 2008). However, an equally interesting, yet seldom discussed, characteristic of the language is that it displays some of the earliest attested examples of *rendaku*, or "sequencial voicing." In Modern Japanese *rendaku* is a morphophonological process that voices the initial consonant of the second element of a compound. For example, when *ao* "blue" and *sora* "sky" combine the output is not *aosora* but *aozora*. There are phonological and lexical restrictions on Modern Japanese *rendaku* and it is not completely regular. In EOJ there are two types of *rendaku*: morpheme-based and process-based. These are described in Section 2 of this paper.

The goal of this paper is to explore the functions of morpheme-based *rendaku* in EOJ through a comprehensive analysis of the corpus. In particular, attention will be paid to its use in speech rhythm, and how it interacts with synchronic vowel elision in hiatus contexts. A comparison between morpheme-based *rendaku* and synchronic vowel elision will also be made to see if their effects in speech rhythm are different, or if one is preferred over the other for any particular function.

[1] Research for this work was supported by Grants-in-Aid for Scientific Research (*Kakenhi*) #23.11784 from the Japan Society for the Promotion of Science (JSPS).

[2] I write all province names in their reconstructed 8th century pronunciations (note that /ö/ represents a schwa). These correspond to the present-day names Kōzuke 上野, Sagami 相模, Shimotsuke 下野, Musashi 武蔵, Hitachi 常陸, Mutsu 陸奥, Shimōsa 下総 and Kazusa 上総, respectively.

1.1 Methodology

Kupchik (2011, pp. 871-1013)'s annotated EOJ corpus was used for this study. The total corpus used consists of 234 poems from *Man'yōshū* books 14 and 20, most of which contain 31 syllables in a syllabic rhythm of 5-7-5-7-7. Any of the poems were marked if they showed morpheme-based *rendaku* or hypermetricality, and then digital searches were conducted to find all examples for each particular type. Examples are written in the IPA, and follow Kupchik (2011, pp. 29-482)'s reconstruction of EOJ phonology and phonetics. Logographic morphemes are written in italics. The original *Man'yōgana*[3] orthography is also given. Textual attestations are given in the following format:

(book:poem.line - province)

For example, if the example is from *Man'yōshū* book 14, poem 3384, line 4, from Simotupusa province, it will be shown as (14:3384.4 - Simotupusa). Sequential lines are marked with a hyphen, while non-sequential lines are separated by a backslash. Thus 14:3506.1-2/5 indicates the poem is from book 14, poem 3506, lines 1, 2 and 5 (but not 3 or 4).

Some poems have no overt province attribution. These are called "unknown dialect" poems, abbreviated as UD.

1.2 The rhythm of EOJ poetry

EOJ poetry is almost exclusively in the *tanka* style, which most often consists of five lines of verse in a sequence of 5-7-5-7-7 syllables. Hypermetrical lines are permitted, at least superficially, but almost never in excess of one syllable[4]. Depending on the scribe or dialect hypermetrical lines may have either been firmly restricted or conditionally permitted.

In order to maintain a stable metrical rhythm, two strategies are used. First, we find quite frequently the synchronic[5] elision of one of two adjacent vowels in hiatus across word boundaries, with V_2 elision being the most prominent (Kupchik, 2011, p. 483), which is typologically rare (Casali, 1997, 1998). I will refer to this as synchronic vowel elision (SVE) in this paper. The second strategy is the use of morpheme-based *rendaku* to contract the line.

---

[3] *Man'yōgana* is a system of Chinese characters used for either their meaning or sound values. Most of the EOJ poems are written in purely phonographic *Man'yōgana*, though a small set of logographic characters appear in some poems as well.

[4] One example with nine syllables in a line is found in 14:3423.5, from Kamitukeno province.

[5] It is important to distinguish between synchronic and diachronic elisions. Synchronic elisions are those that occur optionally at the boundary between two separate word-forms and they are not lexicalized. Diachronic elisions occur in compounds, verbal morphophonology, and lexicalized collocations in the language, and they are completely lexicalized. Many diachronic elisions are shared with Western Old Japanese and certainly go back at least to Proto-Japanese times.

2. Types of *rendaku*

In this paper I distinguish between two types of *rendaku*: process-based and morpheme-based. Process-based *rendaku* is described in section 2.1, while morpheme-based *rendaku* is described in section 2.2.

2.1 Process-based *rendaku*

Process-based *rendaku* involves reduplicating the root and prenasalizing and voicing the onset of the reduplicant. It has a function of pluralization, and there is no derivable morpheme from the resulting prenasalized onset of the reduplicant. This is only attested a few times. An example is shown below.

(i) 20:4391.1-3 – Simotupusa
久尓具尓乃 / 夜之里乃加美尓 / 奴作麻都理
kuni-**ᵑguni**-nə / yasiri-nə kami-ni / nusa matur-i
province-**REDUP**-GEN / shrine-GEN deity-DAT / paper_offering offer-INF
"I make paper offerings to the deities in the shrines of [many] provinces."

The reduplication itself is not adequate to indicate plurality – indeed, without the *rendaku* prenasalization and voicing, reduplication usually has an iterative function in EOJ (Kupchik, 2011, p. 566). Due to the fact process-based *rendaku* is not used to stabilize the rhythm of a line of poetry, it will not be discussed further in this paper.

2.2 Morpheme-based *rendaku* (MBR)

Morpheme-based *rendaku* (henceforth abbreviated as MBR) refers to the word-boundary contraction of the vowel of a nasal-initial syllable of a grammatical morpheme or word form (in the case of the copula *n-*), which in turn fuses with the following voiceless onset, prenasalizing and voicing it in the output. Prenasalized-voiced onsets are restricted from participating, as they are already prenasalized and voiced[6]. This is extremely useful when confronted with a fairly rigid meter and no other allowable contractive processes other than SVE. MBR is well attested in all dialects of Old Japanese and it is the historical source of many of the *rendaku* forms found in Modern Japanese.

The grammatical morphemes and word-forms that may participate in MBR are shown in Table 1 below. While they are not numerous, they are some of the most frequently attested morphemes in the corpus.

[6] EOJ is a language that has the uncommon contrast between voiceless and prenasalized-voiced obstruents, but no plain voiced obstruents.

Table 1: Grammatical morphemes that trigger morpheme-based *rendaku*

Dative/Locative *-ni*
Genitive *-nə*
Copula-infinitive *n-i*
Copula-adnominal *n-ə*
Demonstrative pronoun "this" *kənə*

All contracted *rendaku* forms of the above morphemes are well attested except *kənə*'s form *kə*^{n}-, which is only attested once. An example of each is given below in examples (1) – (5).

(1) 14:3506.1-2/5 – UD
尓比牟路能 / 許騰伎尓伊多礼婆/見延奴己能許呂
nipi muro-nə / **kə^{n}**-dəki-ni itar-e-^{m}ba[7] / *mi*-ye-n-u kənə kərə
new house-GEN / **this**-time-LOC arrive-EV-CONJ / see-PASS-NEG-ATTR this time
"Since the time of the new house has arrived, this is a time when I am unable to see you."

The underlying form of line 2 in example (1) is /kənə təki-ni itar-e-^{m}ba/.

(2) 20:4368.1-2 – Pitati
久自我波々 / 佐氣久阿利麻弖
ku^{n}zi-**^{ŋ}g**apa pa / sake-ku ari-mat-e
PN-**GEN**-river TPT / be.safe-AVINF ITER-wait-IMP
"Be waiting for me safely, [at] Ku^{n}zi river!"

The underlying form of line 1 in example (2) is /ku^{n}zi-nə gapa pa/.

(3) 14:3537b.1-2 – UD
宇麻勢胡之 / 牟伎波武古麻能
uma-se-^{ŋ}-gos-i / mu^{ŋ}gi pam-u koma-nə
horse-fence-**LOC**-cross-INF / grain munch-ATTR stallion-COMP
"Like a stallion that crosses the horse-fence to munch on grain."

The underlying form of line 1 in example (3) is /uma-se-ni kos-i/.

[7] There is another interpretation of this line mentioned by Mizushima (1986, p. 307), in which the first three syllables, *kə^{n}dəki*, are a reduction of *kə-nə təki* "silkworm-GEN time". The problem with this is "silkworm; child" is *ko*, not *kə*. We would therefore have to conclude the vowel change was due to a regressive assimilation to the vowel in *təki* "time", but I am skeptical due to the fact *kə^{n}dəki* is attested just once in all of the Old Japanese corpus. Furthermore, while progressive vowel assimilations are well attested in the EOJ corpus, other examples of a regressive vowel assimilation are not attested.

(4) 14:3412.4-5 – Kamitukeno
可奈師家兒良尓 / 伊夜射可里久母
kanasi-ke *ko*-ra-ni / iya ^{n}-zakar-i-k-umo
be.adorable-AVATTR girl-DIM-DAT /
more.and.more **COP.INF**-be.far.from-INF-go-EXCL
“I am going further and further away from my adorable girl!”

The underlying form of line 5 in example (4) is /iya n-i sakar-i-k-umo/.

(5) 14:3489.5 – UD
左祢度波良布母
sa-ne ^{n}-do parap-umo
LOC-sleep.NML **COP.ATTR**-place clear.away-EXCL
“I’ll clear away a place for us to sleep.”

The underlying form of the line in example (5) is /sa-ne n-ə to parap-umo/.

Normally we only find one occurrence of MBR in a line, but there is one extreme case where the underlying line is nine syllables and two instances of MBR are used to contract it into seven syllables, shown below in example (6).

(6) 14:3497.1-2 – UD
可波加美能 / 祢自路多可我夜
kapa kami-nə / ne-^{n}-ziro taka ^{ŋ}-gaya [8]
river upper.part-GEN / root-**GEN**-white high **COP.ATTR**-grass
“Like the tall grass with white roots in the upper part of the river.”

The underlying form of line 2 in example (6) is /ne-nə siro taka n-ə kaya/.

It should also be noted that MBR only occurs when an underlying line is hypermetrical and it never occurs at the boundary between two lines. Due to this, we can conclude its primary function was to stabilize the rhythm of a single line of verse.

2.2.1 Evidence for the lack of lexicalization of MBR

To show that MBR is not implemented unless needed, examples are given below in (7) – (11) where the underlying line is metrical, and consequently no MBR occurs. While, due to the small size of the corpus, it is difficult to find examples containing all of the same morphemes as in the MBR examples, the following examples are phonologically similar as well as morphosyntactically similar.

[8] While I mark the morpheme boundaries between the prenasalization and the following onset in order to clearly indicate the morphemes involved (i.e. /^{n}-ziro/ “GEN-white”), we must remember prenasalized voiced consonants were unit phonemes. Thus the post-*rendaku* forms are completely fusional and consequently should be understood as being portmanteau morphs.

(7) 14:3460.3-5 – UD
尓布奈未尓 / 和我世乎夜里弖 / 伊波布許能戸乎
nipu nami-ni / wa-ᵑga se-wo yar-i-te / ipap-u **kənə** *to*-wo
new taste.NML-LOC / 1.S-POSS lover-ACC send-INF-SUB /
pray-ATTR **this** door-ACC
"I pray to this door after sending my husband away during the new [rice] tasting."

(8) 14:3369.1-4 – Sagamu
阿之我利乃 / 麻万能古須氣乃 / 須我麻久良 / 安是加麻可左武
asiᵑgari-nə / mama-**nə** ko-suke-nə / suᵑga-makura / aⁿze ka mak-as-am-u
PN-GEN / cliff-GEN DIM-sedge-GEN / sedge-pillow /
why QPT use.as.a.pillow-HON-TENT-ATTR
"Why would you use a sedge pillow made of sedge from the cliffs of Asigari as a pillow?"

(9) 14:3514.4-5 – UD
和礼左倍尓 / 伎美尓都吉奈那 / 多可祢等毛比弖
ware sape n-i / kimi-**ni** tuk-i-n-ana / taka ne tə mop-i-te
1.S RPT COP-INF / lord-LOC attach-INF-PERF-DES / high peak COP think-INF-SUB
"I, too, would like to be with you, my lord. I think of you as a high peak."

(10) 14:3350.4-5 – Pitati
伎美我美家思志 / 安夜尓伎保思母
kimi-ᵑga mi-kesi si / aya **n-i** ki posi-mo
lord-POSS HON-clothes FPT / extremely COP-INF wear.NML be.desired-EXCL
"I so desperately want to put on my lord's clothes!"

(11) 20:4420.4-5 – Muzasi
安我弖等都氣呂 / 許礼乃波流母志
a-ᵑga te-tə tuke-rə / kəre **n-ə** paru məs-i
REFL-POSS hand-COM affix-IMP / this COP-ATTR needle hold-INF
"Put it together again with your own hands, holding this needle."

2.2.2 Unrealized MBR in hypermetrical lines

While there are dozens of cases of unrealized SVE in a hypermetrical line, there are only a few cases where MBR seems to be possible, but remains unrealized in a hypermetrical line. An example is shown in (12) below.

(12) 14:3418.5 – Kamitukeno
伊麻波伊可尓世母
ima pa ika **n-i se**-m-o
now TPT how **COP-INF do**-TENT-ATTR
"Whatever shall I do now?"

In example (12) we find a hypermetrical line even though theoretically the copula-infinitive *n-i* could be contracted to [ⁿ], and voice the following onset to [z]. However, I could not find a single example of the onset of *se-* "do" becoming prenasalized to [ⁿz] in EOJ or WOJ, so it is possible that there was a lexical prohibition on this root

becoming prenasalized. We should note that SVE in the sequence [ai] also does not occur in example (12).

In example (13) below we find a hypermetrical line 5, but no MBR occurring between the copula-infinitive *n-i* and the emphatic particle *si* to create an output of [^{n}zi]. SVE in the sequence [ia] also does not occur in this example.

(13) 20:4351 – Kamitupusa
多妣己呂母 / 夜倍伎可佐祢弖 / 伊努礼等母 /
奈保波太佐牟之/ 伊母尓志阿良祢婆
ta^{m}bi kərəmo / ya-pe ki-kasane-te / i-n-ore-^{n}dəmo /
napo pa^{n}da samu-si / imo **n-i si ar**-an-e-^{m}ba
journey garment / eight-CL wear.INF-pile.INF-SUB / sleep-sleep-EV-CONC /
still skin be.cold-FIN / darling COP-INF EPT exist-NEG-EV-CONJ
"Although I do sleep wearing the many layers [of my] travel garment,
still my bare skin is cold, because my darling is not here!"

As there are no attested examples of a particle undergoing MBR, they were likely restricted from participating in the process, much like words with voiced onsets and the verb *se-* "do". Further strengthening this idea is example (14) below, where we see the same string of morphemes (*n-i si ar-* "COP-INF EPT exist-") with the occurrence of SVE, but not MBR.

(14) 20:4393.1-2 – Simotupusa
於保伎美能 / 美許等尓作例波
opo kimi-nə / mikətə **n-i s-ar**-e-^{m}ba
great lord-GEN / command COP-INF EPT-exist-EV-CONJ
"When [my] great lord's command came to be."

2.2.3 Hypermetricality in the wake of MBR

There is one attestation in the corpus where the application of MBR is not enough to make a line metrical. This example is given in (15) below, where both lines are hypermetrical. There are a few interesting things to note about this example. First, there is nothing to contract in line 1 – the conditions necessary for MBR or SVE to occur are non-existent. Second, in line 2 both MBR and SVE are possible, but only MBR occurs. It is unclear why the adjacent vowels in the sequence *kumo ar-* remain in the output. It is possible that some lexical items were not permitted to undergo SVE (recall, also, the point that there appears to be a restriction on *se-* "do" and particles participating in MBR). Further to this point, as far as I can tell there are no attestations of *kumo* "cloud" losing its final vowel or eliding an adjacent vowel in any Old Japanese text. The verb *ar-* "exist" is similarly unattested in OJ with a synchronically elided vowel.

(15) 14:3516.1-2 – UD
對馬能祢波 / 之多具毛安良南敷
tusima-nə ne pa / sita-^{ŋ}-gumo ar-an-ap-u
PN-GEN peak TPT / below-GEN-cloud exist-NEG-ITER-FIN
“There are no clouds below the peaks of Tusima.”

2.2.4 Lexical parallels to MBR

A few lexical morphemes are also attested behaving similar to MBR. These nouns lose the vowel in their nasal-initial syllable when combined with another word. This is a rare phenomenon in EOJ, but one relevant to the present discussion. One example is *yumi* “bow”. An example of its contracted form *yu*^{n}- is provided below in example (16).

(16) 14:3567 – UD
於伎弖伊可婆 / 伊毛婆麻可奈之 /
母知弖由久 / 安都佐能由美乃 / 由都可尓母我毛
ok-i-te ik-a^{m}ba / imo pa ma kanasi /
mot-i-te yuk-u / a^{n}dusa-nə yumi-nə / **yu^{n}**-duka n-i mo^{ŋ}gamo
leave.behind-INF-SUB go-COND / darling TPT so be.sad.FIN /
hold-INF-SUB go-ATTR / catalpa-GEN bow-GEN / **bow**-grip COP-INF DPT
“If [I] go leaving [her] behind, my darling will be so sad. I wish she could be the grip of the catalpa bow, which I will take with me.”

However, this form is probably lexicalized, as the form *yumi tuka* “bow grip” is unattested in OJ. Thus it is not a good analog to MBR, which occurs synchronically.

A more intriguing example is *miti* “road”, which we often find synchronically contracted to *^{n}di* as the second part of many compounds. This is more relevant to MBR as this contraction only occurs when needed to stabilize the rhythm of the line. This is shown, for example, in example (17):

(17) 14:3405a.3 – Kamitukeno
可波治尓毛
kapa-**^{n}di**-ni mo
river-road-LOC FPT
“Even on the river road”

The underlying form of example (17) is [kapa-miti-ni mo], with six syllables instead of the rhythmically ideal five. By contracting *miti* into *^{n}di* the desired five syllable rhythm for the line is achieved. This word is not contracted unless metrically required, as demonstrated in example (18), which is already metrical with its underlying five syllables that surface in the output:

(18) 14:3477.1 – UD
安都麻道乃
a^{n}duma **miti**-nə
PN road-GEN
“Of Aduma road”

An alternate analysis of *miti*’s contracted form ^{n}*di* is presented by Vovin (2009c, p. 49; 2010, p. 200), who segments it as -^{n}*-di* “-GEN-road” (a contraction of *-nə ti* “id.”), which would make it an example of MBR. I disagree with this analysis for the following reasons: *miti* “road” is often not preceded by the genitive *-nə* when it is uncontracted and preceded by a nominal (such as a place name), as shown in example (18) above; *ti* “road” is unattested preceded by an uncontracted genitive *-nə*; *ti* “road” does not appear at all outside of the compound *ti-mata* “road-fork”, which is only attested logographically in WOJ, the place name 當芸麻知 *ta^{n}gima-ti* “Tagima Road” attested once in *Kojiki Kayō* 77, and possibly also *naga-ti* “long road”, though that also seems to be only attested logographically. Also note the parallelism with *yumi* “bow” ~ *yu^{n}-duka* “bow-grip”, which shows an identical phonological contraction of [mit] > [^{n}d] in compounds. While historically the evidence for “road” being *ti in Proto-Japonic is strong, synchronically only *miti* and its contracted allomorph ^{n}*di* exist in EOJ.

3. MBR or SVE: is one preferred over the other?

The data suggests MBR is preferred over SVE. One piece of evidence can be seen in line 2 of example (1) above, as the adjacent vowels [ii] do not elide, but MBR occurs at the start of the line. Similarly, as mentioned before, MBR occurs in line 2 of example (15) above but no SVE occurs in the adjacent vowels [oa]. In addition, examples (19) and (20) below show MBR, but adjacent vowels remain in the output:

(19) 14:3527.4-5 – UD
伊伎豆久伊毛乎 / 於伎弖伎努可母
iki-**n**-duk-**u i**mo-wo / ok-i-te k-i-n-o kamo
breath-LOC-breathe-ATTR beloved_girl-ACC /
leave_behind-INF-SUB come-INF-PERF-ATTR EPT
“I left my darling weeping and came [here]!”

In example (19) we find the underlying form [iki-ni tuk-u] realized as *iki-^{n}duk-u* due to the implementation of MBR, but no SVE of the following vowel sequence [ui]. Similarly, in example (20) below we find the underlying form [ura-ni tat-i] realized as *ura-^{n}-dat-i* due to the implementation of MBR, with no SVE in the preceding vowel sequence [eu].

(20) 14:3552 – UD
麻都我宇良尓 / 佐和恵宇良太知 / 麻比登其等 /
於毛抱須奈母呂 / 和賀母抱能須毛
matu-ᵑga ura-ni / sawaw**e u**ra-**ⁿ**-dat-i / ma-pitə-ᵑ-gətə /
omop-os-unam-ə rə / wa-ᵑga mop-o-nəsu mo
pine-POSS bay-LOC / noisy tip-LOC-rise-NML / INT-people-word /
think-HON-TENT2-ATTR COP / 1.S-POSS think-ATTR-COMP FPT
"You surely think that people's rumors are like the noisiness rising to the tips
[of the tree branches] in the bay of pines. My thoughts are also like that."

Importantly, metrical lines with SVE along with unrealized MBR are unattested in the corpus. It should also be noted that there are no examples of a hypermetrical line with unrealized MBR (when the aforementioned restrictions on MBR are taken into account), but there are many examples of hypermetrical lines with unrealized SVE.

4. Summary and conclusion

This paper has given evidence that suggests the primary function of MBR was to stabilize the rhythm of a single line of verse. The evidence also suggests that MBR is preferred over SVE when both are available to contract a line into the ideal rhythm of the *tanka* verse. This patterns well with the fact that SVE is not strongly preferred over hypermetricality in the language.

Still, the question remains: why is there a preference for MBR over SVE? One possibility is that SVE and MBR have different stylistic functions in the language. The specific type of SVE found in EOJ quite possibly only occurred in poetic verse (there is no evidence to the contrary), while it seems MBR was something more central to the grammar, due to the fact it continues on in later forms of Japanese while SVE does not. In fact, historically *rendaku* becomes far more prevalent in Japanese as time goes on with morpheme-based and process-based *rendaku* eventually merging into one. Based on this, it is not unrealistic to conclude SVE was mainly a poetic device employed by scribes for verse composition, while MBR was a more basic element of the grammar more widely used in the language. Therefore, this hypothesis maintains MBR had grammatical primacy to stabilize the rhythm of verse, while SVE was a more artificial device used for the same effect mainly when the conditions necessary for MBR to occur were not met. It is unsurprising, then, that when both are equal candidates to stabilize the rhythm of a single line of verse, the primary device is preferred.

Another element that may be at play in these data is the cross-linguistic initial-syllable positional privilege (Beckman, 1998, pp. 50–58) which maintains there is a preference to retain initial syllables of words rather than final syllables. MBR preserves the initial syllable of words, while EOJ's SVE is mainly of the type that elides V_2, which deletes the initial syllable of a word. However, it is difficult to explain the predominantly V_2 elision operating in EOJ if the language really followed this typological trend.

Abbreviations

Linguistic terms

MBR	Morpheme-based *rendaku*
SVE	Synchronic vowel elision

Grammatical terms

1.S	First person singular pronoun
ACC	Accusative case
ATTR	Attributive marker
AVATTR	Adjectival verb attributive
AVINF	Adjectival verb infinitive
CL	Classifier
COM	Comitative case
COMP	Comparative case
CONC	Concessive gerund
CONJ	Conjunctive gerund
COP	Copula
DAT	Dative case
DES	Desiderative
DIM	Diminutive
DPT	Desiderative particle
EPT	Emphatic particle
EV	Evidential
EXCL	Exclamative
FIN	Final predication marker
FPT	Focus particle
GEN	Genitive case
HON	Honorific
IMP	Imperative
INF	Infinitive
INT	Intensifier
ITER	Iterative
LOC	Locative case
NEG	Negative
NML	Nominalizer
PASS	Passive
PERF	Perfective aspect
PN	Place name
POSS	Possessive case

QPT	Question particle
REDUP	Reduplication
REFL	Reflexive
RPT	Restrictive particle
SUB	Subordinative gerund
TENT	Tentative mood
TENT2	Second tentative mood
TPT	Topic particle

Languages

COJ	Central Old Japanese (8th century CE, central Yamato)
EOJ	Eastern Old Japanese (8th century CE, Azuma region)
PJ	Proto-Japanese
TS	Töpo-Suruga Old Japanese (8th century CE, Töpotuapumi (=Tōtōmi) and Suruga provinces)
UD	Azuma poems from an unspecified dialect
WOJ	Western Old Japanese (8th century CE, Nara region)

References

Aso, M. (2011). *Man'yōshū zenka kōgi. Kan daijūsan ~ kan daijūyon* [A commentary on all *Man'yōshū* poems. Books 13 ~ 14]. Tokyo: Kasama Shoin.

Beckman, J. (1998). *Positional faithfulness*. PhD dissertation, University of Massachusetts at Amherst.

Casali, R. (1997). Vowel elision in hiatus contexts: Which vowel goes? *Language* 73(3): 493–533.

Casali, R. (1998). *Resolving hiatus*. New York: Garland.

Fukuda, Y. (1965). *Nara jidai Azuma hōgen no kenkyū* [A study of Nara period Azuma Dialect]. Tōkyō: Kazama shoin.

Hōjō, T. (1966). *Jōdai tōgoku hōgen no kenkyū* [A study of the dialect of the eastern province of the ancient period]. Tōkyō: Maruzen.

Itō, H. (1997). *Man'yōshū shakuchū, 7* [*Man'yōshū* annotated commentary, vol 7]. Tokyo: Shūeisha.

Itō, H. (1998). *Man'yōshū shakuchū, 10* [*Man'yōshū* annotated commentary, vol 10]. Tokyo: Shūeisha.

Kupchik, J. (2011). *A grammar of the Eastern Old Japanese dialects*. PhD dissertation, University of Hawai'i at Manoa.

Mizushima, Y. (1972). *Kōchū Man'yōshū Azuma uta Sakimori uta* [Azuma songs and border guard songs from the *Man'yōshū*]. Tōkyō: Kasama Shoin.

Mizushima, Y. (1984a). Man'yōshū Azuma uta no kokugogaku-teki kenkyū [A linguistic study of the Azuma songs in the *Man'yōshū*]. Tōkyō: Kasama Shoin.

Mizushima, Y. (1984b). *Man'yōshū Azuma uta honbun kenkyū narabi ni sōsakuin* [Indexes to and research on the original text of Azuma songs in the *Man'yōshū*]. Tōkyō: Kasama Shoin.

Mizushima, Y. (1986). *Man'yōshū zenchū. Kan dai 14.* [A complete commentary on the *Man'yōshū*, vol. 14]. Tokyo: Yūhikaku.

Mizushima, Y. (1996). *Kōchū Man'yōshū Azuma uta Sakimori uta: shin zōho kaiteiban* [Azuma songs and border guard songs from the *Man'yōshū*: New reprint with additions and emendations]. Tōkyō: Kasama Shoin.

Mizushima, Y. (2003). *Man'yōshū Sakimori uta zenchūshaku* [A complete commentary on the *Man'yōshū* border guard poems]. Tōkyō: Kasama Shoin.

Mizushima, Y. (2005). *Man'yōshū Sakimori uta no kokugogaku-teki kenkyū* [A linguistic study of the *Man'yōshū* border guard poems]. Tōkyō: Kasama Shoin.

Mizushima, Y. (2009). *Man'yōshū Sakimori uta no kenkyū* [Research on the *Man'yōshū* border guard poems]. Tōkyō: Kasama Shoin.

Nakanishi, S. (1981). *Man'yōshū zen yakuchū genbun tsuki* (*san*) [A complete translation of the *Man'yōshū* with the original text and commentary (vol. 3)]. Tokyo: Kōdansha.

Nakanishi, S. (1983). *Man'yōshū zen yakuchū genbun tsuki* (*yon*) [A complete translation of the *Man'yōshū* with the original text and commentary (vol. 4)]. Tokyo: Kōdansha.

Omodaka, H. (1984a). *Man'yōshū chūshaku kan dai 14* [A commentary on the *Man'yōshū*, Book 14]. Tokyo: Chūōkōronsha.

Omodaka, H. (1984b). *Man'yōshū chūshaku kan dai 20* [A commentary on the *Man'yōshū*, Book 20]. Tokyo: Chūōkōronsha.

Pellard, T. (2008). Proto-Japonic *e and *o in Eastern Old Japanese. *Cahiers de linguistique – Asie orientale* 37(2): 133–158.

Vovin, A. (2005). *A descriptive and comparative grammar of Western Old Japanese – Part 1: Phonology, Script, Lexicon, and Nominals*. Folkestone: Global Oriental.

Vovin, A. (2009a). *A descriptive and comparative grammar of Western Old Japanese – Part 2: Adjectives, verbs, adverbs, conjunctions, particles, postpositions*. Folkestone: Global Oriental.

Vovin, A. (2009b). *Man'yōshū to Fudoki ni mirareru fushigina kotoba to jōdai Nihon rettō ni okeru Ainu go no bunpu* [Strange words in the *Man'yōshū* and *Fudoki* and the distribution of the Ainu language in the Japanese islands in prehistory]. Kyoto: Kokusai Nihon Bunka Kenkyū Sentā.

Vovin, A. (2009c). *A translation of the Man'yōshū Book 15*. Folkestone: Global Oriental.

Vovin, A. (2010). *Koreo-Japonica*. Honolulu: University of Hawaii Press.

THE EVOLUTION OF ERGATIVITY IN IRANIAN LANGUAGES[*]

Yadgar KARIMI
University of Kurdistan
y.karimi@uok.ac.ir

Abstract

This paper presents an attempt to investigate the origins of ergativity in Iranian languages, drawing upon diachronic and synchronic analyses. In so doing, I will trace the development of the ergative structure back to Old and Middle Persian where, it is argued, the roots of ergativity lie. I will specifically show that the ergative pattern as currently obtained in the grammatical structure of some Iranian languages has evolved from a periphrastic past participle construction, the analogue of which is attested in Old Persian. It will further be argued that the predecessor past participle construction imparted a resultative construal in Old Persian and, subsequently, in the transition to Middle Persian, has assumed a simple past reading. The bottom-line of the analysis will be represented as a proposal regarding the nature of the ergative verb, to the effect that an ergative verb, as opposed to a regular (non-ergative) transitive verb, is semantically *transitive*, but syntactically *intransitive*.

Keywords

ergative structure; resultatives; past participle; Old Persian; Iranian languages

Izvleček

Članek poskuša z diahrono in sinhrono analizo prikazati izvor ergativnosti v iranskih jezikih. Avtor sledi razvoju ergativne strukture vse od stare in srednje perzijščine, kamor naj bi segale korenine ergativnosti, in pokaže, da se je ergativni vzorec, ki je prisoten v slovničnih strukturah nekaterih iranskih jezikov, razvil iz perifrastičnega preteklega deležnika, katerega vzporednice so najdene v stari perzijščini. Avtor trdi, da je predhodni pretekli deležnik vodil do nastanka rezultativov v stari perzijščini in je posledično ob prehodu v srednjo perzijščino privedel do enostavnega preteklika. V zadnjem delu analize avtor predstavi naravo ergativnih glagolov in predlaga, da so ergativni glagoli, v nasprotju z običajnimi (ne-ergativnimi) prehodnimi glagoli, semantično prehodni, a skladenjsko neprehodni.

Ključne besede

ergativna struktura; rezultativnost; pretekli deležnik; Stara perzijščina; Iranski jeziki

---

[*] I would like to thank an anonymous reviewer of ALA for insightful comments and helpful suggestions. Thanks are also due to Mojtaba Monshizadeh of Allameh Tabataba'i University, Tehran, for sharing with me his ideas and insights concerning some of the issues raised in an earlier draft of the paper.

1. Introduction

The term "ergative" is widely used to describe a case/agreement pattern in which the subject of an intransitive clause is marked similarly to the direct object of a transitive clause, and differently from the transitive subject (Dixon, 1994). The following examples from Northern Kurdish (Kurmanji) illustrate ergativity:

(1) a. tu be pilîkān čuy-î[1]
you.NOM with stairs went-2.SG
"You went down the stairs."

b. min tu dît-î
I.OBL you.NOM saw-2.SG
"I saw you."

c. te ez dît-im
you.OBL I.NOM saw-1.SG
"You saw me."

In (1a), the subject bears the nominative case and the intransitive verb enters into cross-reference with it. In the transitive clause, however, the verb establishes agreement with the direct object which is marked as nominative (1b-c). The emergent generalization seems to be that the verb in Kurdish shows agreement with the noun phrase that bears the nominative case. The pattern of case-marking and agreement in the Kurdish data is in full conformity with the definition of the ergative structure, given above. It is interesting to note that the subject in the transitive clause is marked as oblique.

The language system, and, concomitantly, its grammatical system, at any given stage of the development of the language is the natural outcome of the interplay of the independent operations and rules of the system. When described synchronically, some of these operations and rules are postulated under the guise of language-particular principles and patterns which are ultimately tantamount to ad hoc stipulations. For examples, the "unaccusative" status assigned to the transitive verb participating in an ergative structure has been derived from the theory-internal considerations in the perspective of which the overall analysis of ergativity is contextualized (Aldridge, 2008; Legate, 2008; Van de Visser, 2006). The result is the designation of the ergative verb as "exceptional", as opposed to the regular (non-Ergative) transitive verb which are designated as "accusative". Accounting for the status of "exceptionality" assigned to an ergative verb simply amounts to the restatement of the status quo in theoretical terms, without ever delving into the fundamental questions concerning the ontological status of the ergative verb.

[1]. The following abbreviations have been used in this paper: 1=first person; 2=second person; 3=third person; ACC=accusative; DAT=dative; GEN=genitive; IMPRF=imperfective; NOM=nominative; OBL=oblique; PL=plural; PTCPL= (past) participle; SG=singular.

However, the diachronic investigations take up where the synchronic studies leave off, when it comes to the analysis of ergativity. Exploring why and how ergativity emerges in the Iranian languages, essentially, boils down to the investigation of the trajectory of the development of this grammatical pattern in the previous stages of these languages. Methodological considerations in conducting a linguistic research of a diachronic orientation impose a descriptive reductionism to the effect that the descriptive discussion of ergativity will be limited to Kurdish, in this paper. The pattern of ergativity and the grammatical properties associated with in Kurdish are taken to represent the ergative structure in the relevant Iranian languages. The diachronic discussion of ergativity will also be restricted to Old and Middle Persian. The only extant historical records the authenticity of which has long been acknowledged by scholars working in the field are from Old and Middle Persian (Kent 1953; Skjærvø 2005). To the extent that I have been able to ascertain, there is no historical document pertaining to a modern Iranian language, other than Persian, that has survived to the present day. Accordingly, given these empirical concerns, I will draw the data from Old and Middle Persian to trace the development of ergativity in Iranian languages back to its origin. However, this is not intended to mean that the modern Iranian languages in which the ergative structure obtains have been derived from Old and/or Middle Persian, rather, I believe, the close genealogical ties of modern Iranian languages and modern Persian makes it possible to extend the results ensuing from this investigation to other Iranian languages. Consequently, in this paper, data drawn from Kurdish will be taken to represent the current status of ergativity in Iranian languages and the data elicited from Old and Middle Persia will constitute the empirical domain based on which the diachronic analysis of ergativity will be conducted.

The linguistic analysis of almost every aspect of the syntax of Iranian languages presents a challenge to the researcher, as these languages are is a state of transition, typologically speaking. This class of languages is in the gradual process of undergoing a typological change from a highly inflectional type to an analytic type and this process of transition confronts the researcher who is trying to provide a purely synchronic study of the syntax of these languages with a stumbling block. The only way to circumvent the obstacle is to integrate the diachronic investigations into the synchronic analyses of the syntax of this set of languages. And, it is only the conflation of the diachronic and the synchronic studies that yields a possibly naturalistic picture of the syntax of Iranian languages.

To derive the syntax of ergativity in Iranian languages, I will draw on data from the earlier stages of Persian to argue that the pattern of ergativity historically descends from a periphrastic past participle construction in Old Persian which initially imparted a resultative meaning. The past participle construction further developed to convey a simple past construal in Middle Persian. It will, explicitly, be proposed that a structural extension of the past participle construction in Middle Persian serves as the immediate predecessor of ergativity in modern Iranian languages. The pursuit of the trajectory of the development of a past participle construction to an ergative construction will

account for the “exceptional” status of the ergative verb as being “unaccusative”. The ergative verb will be argued to be syntactically “intransitive” but semantically “transitive”. This duality follows from the fact that the ergative verb in modern Iranian languages has retained the intransitive (nominal/adjectival) properties of its (past) participle predecessor, while it has assumed a fully-fledged transitive meaning characteristic of two-place predicates. Accordingly, the mixed properties of an ergative verb as straddling intransitivity and transitivity is the repercussion of the typological transition of the grammatical system of such languages from an inflectional type to an analytic one. The concomitant result of the discussion to be presented is that the regularity observed in the behavior of the non-ergative present tense transitive verbs as opposed to the ergative past transitive verbs stems from the relative typological stability reached in the present tense paradigm of these languages.

The structure of this paper is as follows: in section 2, the characteristic features of ergativity, with data chiefly drawn from Kurdish, will be discussed. Section 3 is devoted to the discussion of the past participle construction in Old Persian as the first stage in the development of ergativity. Section 4 presents the second stage of the development of ergativity in Middle Persian. In section 5, the consequences of the diachronic investigation for the synchronic analysis of ergativity will be evaluated. Section 6 constitutes the concluding remarks.

2. The ergative construction: a description

Ergativity in Iranian languages, and in Kurdish for that matter, obtains in the past tense clauses; that is to say, the Iranian languages display a tense-based split ergativity in their grammatical system. The non-past tense clauses in this class of languages tend to follow a nominative-accusative case/agreement system. The following data from Kurdish present the nominative-accusative alignment in the present tense clauses:

(2) a. ez di-kev-im
I.NOM IMPRF-fall-1.SG
“I am falling.”

b. ez te di-bîn-im
I.NOM you.OBL IMPRF-see-1.SG
“I see you.”

c. tu min di-bîn-î
you.NOM I.OBL IMPRF-see-2.SG
“You see me.”

So far, we have seen that the present tense clauses in Kurdish abide by a nominative-accusative case/agreement pattern in that the subjects in an intransitive and a transitive clause are treated alike in terms of case/agreement whereas the direct object

is marked differently (oblique). However, in the past tense clauses a different pattern emerges, as discussed before:

(3) a. tu be pelikān čuy-î
you.NOM with stairs went-2.SG
"You went down stairs."

b. min tu dît-î
I.OBL you.NOM saw-2.SG
"I saw you."

c. te ez dît-im
you.OBL I.NOM saw-1.SG
"You saw me."

In (3a) a past intransitive clause is presented where the subject bears nominative case and the verb cross-references this nominative argument. However, in the past transitive clauses (3b-c), it is the direct object argument that takes the nominative case and triggers agreement on the verb. The subject argument in the past transitive clause is marked oblique. Generally speaking, in the past tense clauses in Kurdish, and in the ergative Iranian languages, the subject of an intransitive clause and the direct object of a transitive clause behave similarly in terms of case/agreement, and differently from the subject of a transitive clause.

That a selective range of Iranian languages conform to an ergative pattern in their case/agreement system is by no means a new finding (Bynon, 1979, 1989, 2005; Trask, 1979); in addition, some theoretically insightful accounts of ergativity in such languages have recently been put forward (Holmberg & Odden, 2004; Karimi, 2010a,b; Van de Visser, 2006). The theoretical treatments of ergativity in general and of Iranian languages in particular share the theory-internal assumption that what makes an ergative transitive clause seem eccentric in terms of case/agreement, as opposed to the regular present tense (non-ergative) transitive clause, has to do with the nature of the ergative verb. These accounts, essentially, maintain the idea that an ergative transitive verb is defective in the sense that it is "unaccusative". In a theoretically neutral sense, an ergative transitive verb fails to license the accusative case on the direct object. While theoretically adequate and empirically correct, the "unaccusative" nature of the ergative transitive verb has not been the subject of a more fundamental investigation as to how and why an ergative verb is different from its non-ergative transitive counterpart. In other terms, the analyses proposed to derive the structure of ergative constructions have always begged the question of why an ergative transitive verb is "unaccusative" whereas its present-tense counterpart is "accusative". On an ontological level, working out an answer to this question will motivate a more realistic view of what an ergative construction essentially reduces to and will lead to a more naturalistic view of the ergative pattern in the paradigmatic system of these languages, as opposed to the non-ergative construction.

Descriptively speaking, the two properties that characterize, and accordingly distinguish, the ergative pattern are as follows:

(4) a. the subject bears an oblique case.
b. the direct object bears a nominative case and controls agreement on the verb.

Coming up with a natural analysis that accounts for the properties (4a) and (4b) in the ergative structure of Iranian languages calls for a reassessment of this structure in terms of its diachronic evolution. This paper presents an attempt to delve into the historical development of the ergative structure as currently observed in Iranian languages. In the next section, I will broach the discussion by exploring the first stage of the development of ergativity in Old Persian.

3. Old Persian: The first stage

The linguists who have addressed themselves to the issue of the origin of ergativity in Iranian languages share the idea that the ergative construction is the natural extension and development of a "past participle construction" attested in the earlier stages of these languages (Benveniste, 1966; Cardona, 1970; Haig, 2008; Skjærvø, 1983; Statha-Halikas, 1979; Trask, 1979). As discussed before, due to the paucity of data of the earlier stages of the Iranian languages exhibiting ergativity, I will focus on data from Old and Middle Persian to investigate the trajectory of the development of ergativity.

3.1 Case system in Old Persian

The Old Persian and its sister Old Avestan constitute the oldest Iranian branch of the Indo-European family of languages. The extant written documents from Old Persian date back to the sixth to fourth centuries B.C. It is important to note that during the time that these documents were created, the Old Persian was already in the state of transition to Middle Persian; therefore, it is difficult to confidently decide that these texts precisely represent the Old Persian. However those working on the earlier stages of the Iranian languages consider the remnant texts as representing Old Persian and I follow the common practice in this respect. Bearing this in mind, let's turn now to the discussion of the case system in Old Persian.

Old Persian makes an extensive use of case-markers to distinguish the different grammatical functions a noun phrase assumes in the clausal syntax. Shown in the following is the list of the case-markers and their corresponding functions in Old Persian (see, Skjærvø 2005 for a fuller discussion on this and related issues):

– Nominative: to mark the subject in the passive, intransitive and transitive clauses.

– Accusative: to mark the object of a transitive verb or of a preposition.
– Dative/genitive: the dative case-marker, though present in the earlier stages of the Old Persian, became extinct and its corresponding functions were taken up by the genitive marker. Accordingly, a genitive case-marker assumes the functions associated with both itself and the dative case-marker. The main functions denoted by the genitive case-marker are possessor marking and indirect object marking. However, it was occasionally used to mark the experiencer subject in a psychological state or the agent/ benefactive in a past participle construction.

The case-markers mentioned above are the only structural cases attested in Old Persian. In other words, for a given noun phrase to take any of these case-markers, that noun phrase must be in a certain grammatical (structural) position in the clause. Besides this set of case-markers, there are case-markers such as instrumental/ablative, vocative and locative that are called inherent (semantic) cases, the distribution of which is sensitive to the idiosyncratic (lexical) properties of the noun phrase they anchor to.

Having laid out this foundational background on the case system in Old Persian, I will turn now to the discussion of the past participle construction in Old Persian, which is widely believed to be the predecessor of the ergative construction.

3.2 The past participle construction in Old Persian

The category of verb in Old Persian was derived from four distinct stems: the present stem, the aorist stem, the participle stem and the perfect stem. The perfect stem was only attested in the earlier stages of the Old Persian and became extinct in the later stages of the development of the languages (Drinka 2003; Skjærvø 2005). The defunct perfect stem in Old Persian had been derived from its counterpart in Old Iranian. The perfect stem in Old Persian was synthetically derived from the verbal root and the reduplication of the initial consonant of the root coupled with a vowel (commonly "*a*") (Skjærvø 2005:104). The conflation of the reduplicated consonant and the vowel is attached to the verbal root as a prefix:

(5) √kar "to do"
[k+a]+√kar → kakār "have done"

However, as pointed out before, the synthetic perfect stem was no longer available in Old Persian and this left the paradigmatic tense/aspect system of Old Persian with a lacuna. To compensate for the deficiency, Old Persian resorts to an analytic construction consisting of a past participle stem and a copular verb. A clause with a perfect construal in Old Persian was thus constructed:

(6) agent/affected (genitive) + patient (nominative) + past participle + copular verb

The following example exhibits a clause in Old Persian constructed according to the formula given in (6)

(7) avaθā=šām hamaranam kartam Ø
then=3.PL.GEN battle.NOM do.PTCPL.NOM be.3.SG
"Then they have fought the battle." (Kent, 1953, DB III, p. 19)

As concerns its distribution, a (past) participle falls into the category of adjectives. Descriptively speaking, a participle is a deverbal adjective. A clausal structure in which the predicate includes a past participle imparts a "resultative" meaning; that is to say, such a construction denotes a past event viewed from the present perspective. The interpretation conveyed by the periphrastic past participle construction is the same as the interpretation imparted by the already deceased synthetic perfect construction. Both yielded a "resultative" meaning.

The past participle construction represented schematically as (6) is widely believed to have triggered the development of ergativity in Iranian languages (Benveniste, 1966; Haig, 2008); however, what the previous work on the origin of ergativity in Iranian languages has failed to account for is the explicit trajectory of the development of the ergativity from the past participle construction. The scholars working on this topic have primarily engaged themselves with the issue of locating a starting-point which laid the foundation for the development of ergativity in Iranian languages. Nonetheless, the issue seems to have escaped the notice of previous scholars as to how the past participle construction precisely evolved to a construction which has come to be called ergative in modern Iranian languages. Although spotting the initial stages of the development of ergativity is a significant undertaking, it is only through the detection of the surmised trajectory of the development of ergativity from a past participle construction that we can be hopeful to provide a naturalistic account of how ergativity is derive in the present Iranian languages. For the results of the diachronic investigation of ergativity to be contextualized in a theoretical perspective, tracing the explicit stages that the ergative construction has undergone is indispensable.

The determination of the interpretation imparted by a past participle construction of the sort represented as (6) has been the subject of debate among the scholars focusing on this construction. The two dominant contentions in this regard are the possessive interpretation and the passive interpretation. Benveniste (1966) states that the past participle construction in Old Persian conveyed a possessive construal much in the same way *manā pita* "my father" did. At the same time, he highlights the parallelism between the possessive noun in a possessive construction and the agent noun in the past participle in (6); both bear the genitive case. On the other hand, he points out that in Old Persian, concurrently with the past participle construction, there has been a separate synthetic passive construction with its demoted agent realizing as a prepositional phrase. Therefore, according to Benveniste (1966), a passive construction is distinguished by the preposition *hača* "by" that licenses the demoted agent noun.

Accordingly, based on these two observations, he maintains that the past participle construction in Old Persian was a possessive one and not passive.

Cardona (1970) argues against Benveniste's (1966) analysis and states that there are passive structures in Old Persian in which the agent noun bears a genitive case and, therefore, the mere presence of the genitive case on a noun does not imply that a structure is possessive. Cardona (1970) suggests that the past participle construction in Old Persian was passive with a genitive agent. Statha-Halikas (1979) states that Cardona's arguments concerning the structure of the past participle construction in Old Persian were not strong enough. He argues that the genitive case had become a versatile case in Old Persian assuming the functions already performed by other oblique cases. Hence, a possessive noun and a passive agent are equally marked by a genitive case, which is a cover oblique case. According to Statha-Halikas, taking the genitive case as the key to determining the semantic/structure of the past participle construction in Old Persian is not a plausible move.

Compiling the full range of past participle constructions in Old Persian, Statha-Halikas (1979) and Skjærvø (1985) draw the descriptive observation that the majority of the attested past participle constructions in Old Persian texts are essentially agent-less. Building on their observations, two general configurations can be envisaged in which a past participle predicate has participated:

(8) a. patient (nominative) + past participle + copular
b. agent/affected (genitive) + patient (nominative) + past participle + copular

The frequency of occurrence of the pattern represented as (8a) is much more than that of (8b) in Old Persian. These latter scholars, independently, believe that the agent-less construction represented as (8a) imparted a perfect passive meaning while the construction in (8b) with the agent noun expressed a perfect active meaning.

The frequency of occurrence reported in Statha-Halikas (1979) and Skjærvø (1985) as a solution to settle the debate of interest is only suggestive and it is by no means conclusive. Although I do concur with the latter two scholars on the conclusions they draw, in the next section, I will present further formal evidences from Old Persian that corroborate the idea that the past participle construction as represented in (8b) did not implicate a passive construal.

3.3 The active predecessor

As discussed in the previous section, the past participle construction which is considered to be the forerunner of ergativity in the Iranian languages comes in two flavors: (a) with an agent participant and (b) without an agent participant. The following examples illustrate the past participles without an agent participant:

(9) a. xšačam tya hacā amāxam taumāyā parābartam āha.
kingdom.NOM which from our family taken-away.PTCPL be.PAST.3.SG
"The kingdom which has been taken away from our family." (Kent 1953, DB I, 61-62)

b. vasiy aniyašciy naibam kartam anā Pārsā
much other good do.PTCPL in Persepolis
"Much other good (construction) has been built in Persepolis."
(Kent, 1953, XPa, pp. 13-14)

c. tya Bardya avajata
that Smerdis slay.PTCPL
"That Smerdis has been slain." (Kent, 1953, DB I, p. 32)

And shown in the following examples are instances of the past participle construction with an overt agent participant.

(10) a. tya **manā** kartam
that.NOM 1.SG.GEN do.PTCPL
"That is what I have done." (Kent, 1953, DB I, p. 27)

b. avaθā=**šām** hamaranam kartam
then=3.PL.GEN battle.NOM fight.PTCPL.NOM
"Then they have fought the battle." (Kent, 1953, DB III, p. 19)

c. tya **manā** kartam utā tya**maiy** **piça** kartam
that.NOM 1.SG.GEN do.PTCPL and that.NOM=1.SG.GEN father.GEN do.PTCPL
"That is what I have done and that is what my father has done."
(Kent, 1953, XPa, pp. 19-20)

The constructions of interest in this paper are the past participle constructions with an agent participant as represented in (10a-c). In (10a), the agent noun bears the genitive/dative case and the patient noun is implied by the nominative relative pronoun which takes the patient participant as its antecedent. The genitive/dative agent noun has been cliticized to the second-position in (10b), and the patient participant *hamaranam* is marked by the nominative case. The data cited in (10c) provides evidence that, unlike the agent nouns in (10a-b) which are pronouns, a common noun phrase may be an agent participant, too.

It is recalled from section 3.2 that the interpretation of the past participle constructions with an agent participant ignited a controversy among the scholars analyzing such constructions. It is interesting to note that Kent (1953), from whose seminal work the data have been drawn, treats all the examples in (10a-c) as agented passive constructions[2]. However, based on the greater frequency of occurrence of the agentless past participle constructions, Statha-Halikas (1979) and Skjærvø (1985)

[2] The glosses given by Kent (1953) for the examples (10a-c) are, respectively, as follows:

i) It was done by me.
ii) Then the battle has been fought by them.
iii) It was done by me and that was done by my father.

argued that the agentless past participle constructions were indeed passive structures whereas the agented past participle constructions imparted an active perfective meaning. It was also discussed that the frequency of occurrence may not constitute a strong argument for the conclusions read by these two scholars because it may be only a coincidence that the extant Old Persian texts contain more agentless past participle constructions than the agented counterparts, and hence they may not have truly reflected the syntax and the frequency of occurrence of such constructions.

In what follows, I will provide more direct evidence drawn from the structural properties of the agented past participle construction to argue that this construction conveyed an active perfective construal and not a passive one, along the lines of Statha-Halikas (1979) and Skjærvø (1985). It will specifically be illustrated that the dative/genitive agent clitic in the past participle construction merges in a high structural position which is typical of subjects. The properties/tests to be discussed include conjunction reduction, clitic placement and the synthetic vs. analytic passive coexistence which all point to the fact the agent clitic in the past participle construction enjoys a structurally and semantically high status which in turn renders the construction active (and not passive).

The first piece of evidence comes from the conjunction reduction test. A coordinate structure in which both conjuncts share a similar constituent can be rearranged so that the shared constituent in the second conjunct is omitted under identity with the antecedent constituent in the first conjunct. One of the frequent cases of conjunction reduction is the deletion of the shared subject(s) which is coreferential with the antecedent subject. It is a cross-linguistically valid generalization that it is only a canonical subject that can control the deletion of the coreferential subjects in the subsequent conjuncts. Now, let's turn to the data from Old Persian:

(11) avaθā=**šam** hamaranam kartam utā avam Vahyazdātam agarbāya
then=3.PL.GEN battle fight.PTCPL and that Vahyazdata took-prisoner.3.PL

tā martiya tyašaiy fratamā anušiyā āhata agarbāya
and men who=3.SG.GEN foremost followers were arrested.3.PL

"They have fought a battle, and have arrested Vahyazdata, and have arrested his foremost followers." (Kent, 1953, DB III, pp. 47-49)

The above data lend credence to an analysis in which the dative/genitive agent in the past participle constructions in Old Persian occupies the syntactic subject position. In (11), the subject in the second and third conjuncts has been elided under the identity with the subject in the first conjunct. Therefore, it suggests that in Old Persian the genitive/dative agent in the past participle construction is treated as high in the clause structure as a canonical subject. The degree of topicality the dative/genitive agent noun enjoys in a past participle construction is in sharp contrast with the measure of topicality assumed by a demoted agent phrase in a passive construction (e.g., *by-agent* in English). The natural reading of the chain of coordinate constructions in (11) imposes a topical construal of the dative/genitive agent NP which enables it to license

all subsequent occurrences of the elided subjects; whereas in a passive construction the topicality of the agent is further demoted due to its low position in the clause structure, and hence the failure to license conjunction reduction.

The discrepancy observed in the behavior of nominative and oblique clitics constitutes the second piece of evidence pointing to an active interpretation of an agented past participle construction. In Old Persian, nominative clitics remain in their base-generated position and do not undergo movement. However the oblique clitics undergo movement and this movement is to the second position in a clause. Among the oblique clitics, however, only *accusative* and *genitive/dative* clitics bear the ability to undergo movement (Haig, 2004, pp. 155-158). It, then, follows that the dative/genitive case, constituting a natural class with the accusative case, is licensed structurally much in the same way the accusative case is (Chomsky, 1981). A structural case is assigned to a noun phrase that is part of the argument structure of the verb and accordingly, it becomes evident that the genitive/dative agent is licensed as part of the argument structure of the verb. This is while a demoted agent noun phrase in a passive clause is not part of the canonical argument structure and assumes a peripheral position:

(12) a. aita=maiy Auramazdā dadātuv
this=1.SG.GEN Ahuramazda may.give
"May Ahuramazda give this to me." (Kent, 1953, DNa, pp. 53-55)

b. Auramazdā=maiy upastām abara
Ahuramazda=1.SG.GEN aid bore.3.SG
"Ahuramazda bore me aid (helped me)." (Kent, 1953, DB I, pp. 87-88)

The data in (12a-b) show that, along with the dative/genitive clitic, an accusative clitic may undergo movement to the second-position in a clause in Old Persian.

The last piece of evidence is drawn from the empirical observation that in Old Persian a synthetic passive construction co-existed with the agented past participle construction. Given the economy-driven considerations, it is logical to argue that the concurrent existence of a synthetic passive construction in the Old Persian has obviated the need to develop a parallel passive construction serving the same purposes. The agent noun phrase in the synthetic passive in Old Persian came in three different guises: (a) as a preposition phrase governed by the preposition *hača* "from", (b) as a preposition phrase followed by the postposition *rādiy* "by", and (c) as a genitive/dative marked noun. Although the attested data representing the synthetic passive in Old Persian texts are scarce, it is patently obvious that each incarnation of the agent phrase in the synthetic passive has been determined by the idiosyncratic properties of the verb:

(13) utā=šam Auramazdā naiy ayadiya
and=3.PL.GEN Ahuramazda NEG worshiped.PASS
"And Ahuramazda was not worshiped by them." (Kent, 1953, DBV, pp. 15-16)

A further interesting point to note concerning the synthetic passive exemplified in (13) is the passive morphology on the verb. While in an agented past participle

construction, the predicate is realized as a past participle form of the verb, in the synthetic passive construction the passive verb takes the passive morphology. This in turn counts as a supporting evidence for the fact that the agented past participle construction in Old Persian has not been passive.

In the next section, I will address the syntactic derivation of the agented past participle construction which is argued to have constituted the initial impetus in the development of ergativity in the Iranian languages.

3.4 Deriving the syntax of the past participle construction

It was discussed in section 2.2 that the periphrastic past participle construction was a system-internal compensatory strategy to replace the defunct synthetic perfect construction. This paradigmatic choice was motivated by the identity of the interpretation read off from a past participle construction. The semantic yield of a past participle construction is a "resultative" construal and this is precisely the interpretation which ensued from the extinct synthetic perfect construction. As far as I have been able to verify, the data attested in the Old Persian texts do not include a past participle construction in which both the agent and the patient participants are realized as common nouns in one clause. In the attested data, either the patient is realized as a relative pronoun and the agent as a common noun, or the patient is realized as a common noun and the agent as a clitic. To illustrate how a past participle is syntactically derived, I will reconstruct a past participle construction where both the agent and the patient are common nouns. In fact, the reconstructed example is a conflation of the constituents in (10b) and (10c):

(14) maiy pica hamaranam kartam
1.SG.GEN father.GEN battle.NOM do.PTCPL
"My father has fought the battle."

The following is the syntactic structure I propose to derive a past participle construction:

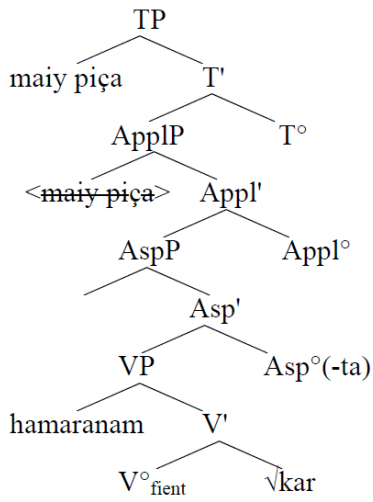

Figure 1: The past participle construction

Building on work couched within the Distributed Morphology framework (Halle & Marantz, 1993) and proposals put forward in Embick (2004), I assume that the root merges with an abstract *fientive* verb which serves to convert the root into a fully-fledged verb [3]. The traditional parsing of a past participle distinguishes two components, which largely corresponds to the result ensuing from the Distributed Morphological parsing of the past participle:

(15) a. √*kar* "root"
b. √*kar* + verbalizer
c. √*kar* + verbalizer + *ta* (participle stem marker)
d. *Karta* "the past participle stem"

In the traditional grammars of Old Persian, "*-ta*" has been considered the past participle marker which turns a root into a past participle stem (Skjærvø, 2005, p. 97). The verbalizer component which corresponds to the fientive verb is phonologically null in Old Persian. Following Embick's (2004) proposals to derive the syntactic structure of a resultative construction, I will assume that the object (patient) is merged as the specifier of the fientive verb, extending it into a verb phrase (VP). Along the lines of Embick (2004), I also assume that what has traditionally been termed "participle marker" in Old Persian is, in fact, the instantiation of the aspect head in syntax. It is this aspectual head merged with the eventive VP that renders the whole structure resultative in interpretation. A noteworthy feature of the past participle construction in Old Persian is that the VP is not merged with a causative v° to project

[3] In Distributed Morphology, roots are syntactically the minimal category-neutral elements that enter the derivation of sentences with an abstract core of meaning (Marantz, 1997). For a root to assume a grammatically independent status, it must combine with a syntactic category-assigner element. In the present case, a fientive verbalizer turns the root into a verb.

into vP. This is because the past participle construction, being resultative semantically, denotes a final state in which the direct object is found in relation to whole event.

The subject/affected in the past participle construction in Old Persian bears dative case. Incorporating ideas developed originally in Marantz (1993) and later elaborated in Cuervo (2003) and Pylkkänen (2002, 2008), I assume that the subject/affected, being dative, is applied to the whole AspP as the specifier of a high applicative head[4]. Semantically, the subject in the past participle construction in Old Persian is assigned a *benefactive* interpretation. For example the past participle construction *hya manā kartam* is ambiguous as between "this is what I have done" with the agentive reading of the subject, or "this is what has been done to my interest' with a benefactive reading of the subject. The last step in the derivation is the introduction of the tense phrase (TP). The head T° enters into agreement with the patient NP which is already caseless in the derivation. This is because, as discussed before, the predicate in the VP is a participle and not a fully-fledged transitive verb, and since participles share adjectival/nominal properties, they fail to assign an accusative case to the patient. As a result of the T°-patient agreement, the patient noun phrase receives the nominative case[5].

In the next section, the route of the development of the ergative construction will be further traced in Middle Persian.

4. Middle Persian: The second stage

The past participle construction in Middle Persian represents the second stage of the development of ergative construction in Iranian languages. As extensively discussed in section 3, the paradigmatic tense/aspect system of Old Persian included four verbal stems: the present stem, the aorist stem, the perfect stem and the participle stem. Of these four stems, the perfect stem which contributed a synthetic perfect tense/aspect became defunct and was no longer available as a paradigmatic choice in the tense/aspect system in the later stages of the language. To make up for the gap, Old Persian resorted to the past participle stem to construct periphrastic structures imparting the resultative meaning typical of a perfect construction.

[4] In the current approaches to event/argument structure, an argument which is not directly engaged in the event denoted by the verb phrase is licensed by an applicative head. The interpretation assigned to an applied argument is thus: the action/state denoted by the verb phrase is in the interests of the applied argument. In languages in which an experiencer argument is marked dative in a psychological construction, it is argued that the experiencer is licensed as an applied argument. For example the Spanish "John (dative) likes the cake" is analyzed as "the cake is appealing to John (dative). The state denoted by the verb phrase "the cake is appealing" is oriented toward the applied argument "John".

[5] A point worthy of mention, in passing, concerns the declension of the past participle itself. It is evident from the examples that the past participle, constituting part of the predicative structure, bears nominative case. This empirical fact gives further credence to the stance that the past participle was still retaining nominal properties in Old Persian. Case in Old Persian was restricted to showing up on the categories endowed with the [+N] feature.

However, the reduction in the tense/aspect oppositional system of Old Persian did not reach a terminus at that stage, rather it extended to Middle Persian with much greater effects. The tense/aspect system further reduced in Middle Persian to the effect that only the participle stem and the present stem were retained. In other words, the additional reduction in Middle Persian resulted in the extinction of the aorist aspect. The logical candidate to substitute the now defunct aorist stem was the past participle stem. Consequently, the following form-function correspondences emerged:

FORM	FUNCTION
Present stem	Present tense
Aorist stem	Past tense
Participle stem	Predicative adjective
Perfect stem (defunct)	Perfect tense/aspect

FORM	FUNCTION
Present stem	Present tense
Aorist stem	Past tense
Participle stem	Predicate adjective
	Perfect aspect/tense

Figure 2: Old Persian

FORM	FUNCTION
Present stem	Present tense
Aorist stem (defunct)	Past tense
Participle stem	Predicate adjective
	Perfect aspect/tense

FORM	FUNCTION
Present stem	Present tense
Participle stem	Predicate adjective
	Perfect aspect/tense
	Past tense

Figure 3: Middle Persian

As of result of this further reduction, the past participle stem assumed the past-tense-marking function which was already born by the aorist stem. Along with the typological change of the Middle Persian from a highly inflectional type to an analytic type, most of the inflectional morphology associated with the nominal/adjectival categories became detached. The common nouns, adjectives and pronouns no longer displayed morphological distinctions for such categories as case and gender. The case system reduced to two cases of nominative (direct) and oblique (indirect). The case marking on the past participle stem did not survive either.

Of particular importance to the present discussion is the fate of the past participle stem. The past participle stem has now a dual function of perfect-marking and past-tense-marking (its primary function as a predicative adjective aside). In order to increase the morphological markedness to establish a form-function parallelism and to unravel the intricacies brought about in the system for the mapping of the form onto the function, the past participle stem in Middle Persian underwent a process of “apocope” to the following effect.

It is recalled from section 3 that the past participle stem in Old Persian was thus derived:

(16) √root + verbalizer + *ta*

It was also mentioned that the final suffix *–ta* was a participle-marker according to the traditional grammarians. In Middle Persian, the past participle underwent apocope and as result the final vowel *–a*, which corresponds to the participle-marker's vowel, was dropped. The remnant portion was treated as a past stem of the verb by the oppositional system of the language:

(17) √root + verbalizer + *t*

An important point to note, however, is that Middle Persian retained the original past participle stem (16) to express the perfect tense/aspect, as in Old Persian. The following example from Middle Persian, with a simple past construal, shows how the past participle form has been reduced in form:

(18) man **abgust** ahénd
I.OBL reveal-PTCPL copular.3.PL
"I revealed them." (Sundermann, 1989, p. 29)

Since the past participle marker in Old Persian and later in Middle Persian had the invariant phonological form of *–ta*, all the simple past tense stems deriving from the process of apocope ended up with the final consonant *–t*. And this phonological form has continued to survive to the present time. The following lists illustrate a selective number of the past tense stems in modern Persian and Kurdish, respectively:

(19) a.

košt	"killed"	šost	"washed"	andāxt	"threw"
foruxt	"sold"	nevešt	"wrote"	šenāxt	"recognize"
kāšt	"implanted"	dašt	"had"	poxt	"cooked"
bāxt	"lost"	xord[6]	"ate"	bord	"took"

b.

kušt	"killed"	xiwārd	"ate"	nāsānd	"introduced"
firošt	"sold"	bird	"took"	xusānd	"soaked"
čānd	"implanted"	xist	"threw"	kird	"did"
šušt	"washed"	kulānd	"cooked"	girt	"held"

A shift in the interpretation of the eventuality encoded in the event structure including a past tense stem was the semantic yield of the reduction of the final vowel in the process of apocope and the concomitant change of the past participle stem into a simple past stem. The eventuality previously expressed by an event structure with a past participle construction was that of a stative/resultative predicate; however, the split in the form of the past participle stem and the associated split in function resulted in a

[6]. The final *–t* in the past tense stem undergoes voicing assimilation into *–d* in the context of a preceding vowel or voiced consonant.

shift from a "stative/resultative" interpretation to a "purely eventive/dynamic" interpretation.

5. Modern Iranian languages: The final stage

The hypothesis defended in this paper is that the ergative structure as currently obtains in a wide range of Iranian languages is the natural extension and development of the past participle and the subsequent past tense construction in the previous stages of these languages, much similar to what happened in the case of Old Persian. The diachronic investigation of the development of ergativity in Old Persian and Middle Persian has immediate implications for the synchronic study of ergative structure in modern Iranian languages.

To begin with the ergative (past transitive) verb, the "exceptional" status ascribed to this class of verbs receives a natural explanation. As is recalled from section 2, the past transitive (ergative) verb has long been recognized as an exceptional verb, as opposed to the regular present transitive verb, in that it fails to assign the accusative case to its object. In theoretical terms, the ergative verb is defective in the sense that it is "unaccusative". According to the diachronic analysis, the current past transitive (ergative) verb has been derived from a past participle stem attested in the earlier stages of these languages. This trajectory of development implies that the current past transitive (ergative) verbs have still retained the syntactic properties of their predecessor, i.e., the past participle. They share with their predecessor the syntactic nominal/adjectival properties and, as a result, fail to assign the (structural) accusative case to their direct object. In the generative tradition of syntactic analysis, the syntactic categories specified for the feature [+N] are not case-assigners (Chomsky, 1970, 1981). The [+N] feature bring nouns and adjective together in a class and sets them in contrast with the syntactic categories endowed with [-N] feature that are case-assigners (verbs and prepositions). Accordingly, the diachronic investigation of ergativity in Iranian languages reveals that the current past transitive verbs are semantically "transitive", but syntactically "intransitive".

As for the oblique case showing up on the subject in the past transitive (ergative) construction, it is recalled that the ergative structure had a predecessor with a resultative construal and that the subject in such constructions was licensed in an applicative phrase, with the applicative head assigning the dative/genitive case to it. As with the past transitive verb, I assume that the subject in the past transitive clause enjoys an in-between status, being semantically an agent but syntactically an applied argument.

The last point to note concerns the nominative case on the direct object in the ergative (past transitive) constructions in Iranian languages. I will assume that the head Tense in the syntactic derivation enters into agreement with the potentially available

direct object and renders it nominative. The following tree diagram represents the syntactic structure I propose for the ergative construction in Iranian languages:

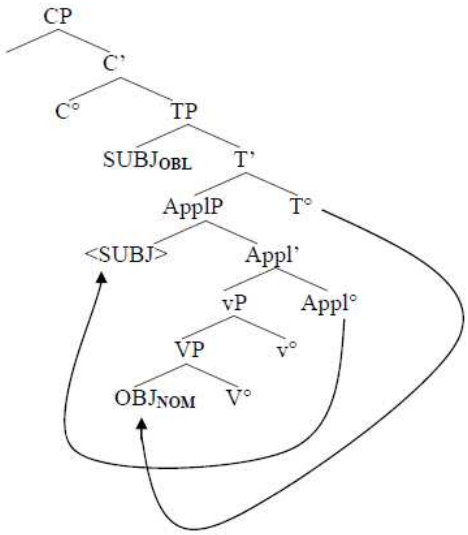

Figure 4: The structure of ergativity in Iranian languages

6. Conclusion

In this paper, I tried to provide a natural analysis of the syntactic structure of ergativity in Iranian languages, with data chiefly drawn from Kurdish, Old Persian and Middle Persian. In so doing, a diachronic investigation was carried out of how ergativity evolved in Iranian languages, and, consequently, the trajectory of the development of ergativity was drawn up from Old Persian to modern Iranian languages. Specifically, it was established that the ergativity in Iranian languages has as its origin a past participle construction with a resultative interpretation, attested in the earlier stages of these languages. The past participle construction underwent a functional extension in the more recent stages of these languages to impart a simple past meaning and this is the basis for the emergence of ergativity in the modern Iranian languages. Given the results ensuing from the diachronic investigation of ergativity, the questions concerning the nature of the ergative verb, the oblique subject and the nominative object in an ergative structure received a natural account.

The present work has as its tangential implication the proposition that the synchronic analyses should be accompanied by diachronic investigations when dealing with the syntactic structure of languages which are in the state of typological change, and that it is only a combination of the two types of study that give rise to natural accounts of the structure of these languages.

References

Aldridge, E. (2008). Minimalist analysis of ergativity. *Sophia Linguistica*, *55*, 123–142.

Benveniste, E. (1966). La construction passive du parfait transitif. In E. Benveniste (Ed.) *Probl`emes de linguistique generale* (pp. 176–186). Paris: Gallimard.

Bynon, T. (1979). The ergative construction in Kurdish. *Bulletin of the School of Oriental and African Studies*, *42*, 211–224.

Bynon, T. (1980). From passive to active in Kurdish via the ergative construction. In E. C. Traugott, R. Labrum & S. Shepherd (Eds.) *Papers from the 4th International Conference on Historical Linguistics* (pp. 151–163). Amsterdam: Benjamîns.

Bynon, T. (2005). Evidential, raised possessor, and the historical source of the ergative construction in Indo-Iranian. *Transactions of the Philological Society*, *103*, 1–72.

Cardona, G. (1970). The Indo-Iranian construction *mana krtam*. *Language*, *46*, 1–12.

Chomsky, N. (1970). Remarks on nominalizations. In R. Jacobs & P. Rosenbaum (Eds.) *Readings in English Transformational Grammar* (pp. 184–221). Waltham, MA: Ginn and Co..

Chomsky, N. (1981). *Lectures on Government and Binding.* Foris: Dordrecht.

Cuervo, M. C. (2003). *Datives at large*. Cambridge, MA: MIT dissertation.

Dixon, R. M. W. (1994). *Ergativity*. Cambridge: Cambridge University Press.

Drinka, B. (2003). The Perfect in Indo-European. Stratigraphic evidence of prehistoric areal influence. In H. Andersen (Ed.) *Language contacts in prehistory. Studies in stratigraphy* (pp. 77–105). Amsterdam: Benjamins.

Embick, D. (2004). On the structure of resultative participles in English. *Linguistic Inquiry*, *35*, 355-392.

Haig, G. (1998). On the interaction of morphological and syntactic ergativity: Lessons from Kurdish. *Lingua*, *105*, 149–173.

Haig, G. (2004). *Alignment in Kurdish: A diachronic perspective*. Unpublished *Habilitationsschrift*: Philosophische Fakultät der Christian-Albrechts-Universität zu Kiel

Haig, G. (2008). *Alignment change in Iranian languages: a construction grammar approach*. New York: Mouton de Gruyter.

Hale, K. & Keyser, S. J. (1993). On argument structure and the lexical expression of semantic relations. In K. Hale & S. J. Keyser (Eds.) *The View from Building 20* (pp. 53–109). MIT Press, Cambridge Mass.

Hale, K. & Keyser, S. J. (2002). *Prolegomenon to a Theory of Argument Structure*. Cambridge, MA: MIT Press.

Halle, M. & Marantz, A. (1993). Distributed morphology and the pieces of inflection. In K. Hale & S. J. Keyser (Eds.) *The View from Building 20* (pp. 111–176). MIT Press, Cambridge Mass.

Holmberg, A. & D. Odden. (2004). Ergativity and role-marking in Hawrami. Paper presented at *Syntax of the World's Languages (SWL 1)*, Leipzig, Germany.

Karimi, Y. (2010a). Unaccusative transitives and the Person-Case Constraint effects in Kurdish. *Lingua*, *120*, 693–716.

Karimi, Y. (2010b). *Ergativity: Its origin and nature*. PhD dissertation, Allameh Tabataba'i University, Teheran.

Kent, R. (1953). *Old Persian*. New Haven, Connecticut: American Oriental Society.

Legate, J. (2008). Morphological and abstract case. *Linguistic Inquiry*, *39*, 55–101.

Marantz, A. (1993). Implications of asymmetries in double object constructions. In S. Mchombo (Ed.) *Theoretical aspects of Bantu grammar* (PP. 113–150). Stanford, CA: CSLI.

Marantz, A. (1997). No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. In A. Dimitriadis, L. Siegel, C. Surek-Clark, & A. Williams (Eds.) *Proceedings of the 21st Annual Penn Linguistics Colloquium* (pp. 201–225). Pennsylvania Working Papers in Linguistics 4.2. Philadelphia: University of Pennsylvania, Penn Linguistics Club.

Pylkkänen, L. (2002). *Introducing Arguments*. Ph.D. dissertation. MIT, Cambridge, MA.

Pylkkänen, L. (2008). *Introducing Arguments*. MIT Press, Cambridge MA.

Skjærvø, P. O. (1983). Case in Inscriptional Middle Persian, Inscriptional Parthian, and the Pahlavi Psalter. *Studia Iranica*, *12*, 47–62.

Skjærvø, P. O. (1985). Remarks on the Old Persian verbal system. *Münchner Studien zur Sprachwissenschaft*, *45*, 211–227.

Skjærvø, P. O. (2005). *An introduction to Old Persian*. Cambridge, Harvard.

Statha-Halikas, H. (1979). How not to tell a passive: the case of Old Persian *mana krtam* reconsidered. *Berkeley Linguistic Society*, *5*, 350–361.

Sundermann, W. (1989). Mittelpersisch. In S. Rudiger (Ed.) *Compendium Linguarum Iranicarum* (pp. 138–164). Wiesbaden: Reichert.

Trask, L. (1979). On the Origins of Ergativity. In F. Plank (Ed.) *Ergativity: Towards a Theory of Grammatical Relations* (pp. 385-404). New York, New York: Academic Press.

Van de Visser, M. (2006). *The marked status of ergativity*. LOT Dissertation Series 141.

SYNTAGMS CONSTRUCTED WITH COORDINATIVE PARTICLES IN *ṚGVEDA* 1.1 –1.50

Tamara DITRICH
University of Ljubljana, University of Sydney
t.ditrich@gmail.com

Abstract

In the *Ṛgveda*, several types of syntactic construction expressing a coordinative or copulative relationship occur: *dvandva* compounds, copulative asyndeta, elliptic duals, and syntagms constructed with coordinative particles. This article investigates the role of coordinative particles in the first fifty hymns of the *Ṛgveda*, focusing in particular on the most frequently used particle *ca* and comparing its use to other copulative conjunctions attested in the text, especially where *ca* is used twice (i.e. *ca … ca*) and the particle *utá*. The article investigates how coordination is expressed between two words and aims to identify differences in the usage of coordinative particles if the words coordinated are theonyms or non-theonyms. By examining how two words coordinated with particles occur elsewhere in other coordinative constructions, the article demonstrates that research into coordinative constructions in Vedic ought to pay special attention to the specific grammatical and linguistic features of theonyms.

Keywords

Vedic Linguistics; Ṛgvedic Exegesis; Coordinative Particles in the *Ṛgveda*; Particle *ca* in Vedic

Izvleček

V *Ṛgvedi* je zabeleženih več tipov koordinativnih nominalnih zvez kot so *dvandva* zloženke, asindetične povezave, eliptične dvojine in sintagme, v katerih so besede povezane z vezalnimi členicami. Prispevek raziskuje sintagme s členicami v ṛgvedskih himnah 1.1–1.50 in se pri tem osredotoča zlasti na najbolj pogosto rabljeno vezalno členico *ca* (»in«) ter primerja njeno rabo z drugimi členicami, ki izražajo koordinacijo, zlasti z rabo dveh členic *ca* (i.e. *ca … ca*) in členico *utá*. Pregled različnih načinov izražanja koordinativnosti med dvema besedama kaže na razlike v rabi členic, ki povezujejo dve (ali več) imen bogov in besedami, ki se ne nanašajo na bogove. Prispevek analizira vse besede iz *Ṛgvede* 1.1–1.50, ki so povezane z vezalnimi členicami in jih primerja z alternativno izraženimi koordinativnimi zvezami (na primer *dvandva*) in ugotovlja, da mora raziskovanje koordinativnih povezav v vedskem jeziku nameniti pozornost specifičnim lingvističnim značilnostim imen bogov v *Ṛgvedi*, ki kažejo drugačno rabo členic kot besede, ki se ne nanašajo na bogove.

Ključne besede

vedska lingvistika; eksegeza *Ṛgvede*; členice v *Ṛgvedi*; vezalna členica *ca* v vedskem jeziku

1. Introduction

This article investigates the role of coordinative particles in the *Ṛgveda*, focusing especially on the particle *ca*, and compares its usage with other copulative conjunctions such as the particle *utá* and the usage of two of the particles *ca* (i.e. *ca … ca*).[1] The research is based on the first fifty hymns of the *Ṛgveda* from which all the syntagms constructed with copulative conjunctions were identified, analysed and compared with other coordinative nominal constructions attested in the entire text of the *Ṛgveda*. The first fifty hymns of the *Ṛgveda*, addressing a variety of deities, provide a substantial amount of materials for investigation of coordinative conjunctions. The main focus of this article is the investigation of different functions of the particle *ca* when situated as a conjunction coordinating two theonyms and two non-theonyms.

In the analysis of the first fifty hymns of the *Ṛgveda*, several types of syntactic constructions expressing a coordinative or copulative relationship occur such as *dvandva* compounds, asyndeta, elliptic duals, and syntagms constructed with copulative conjunctions. It has been shown elsewhere (Ditrich, 2010, pp. 35–44) that examination of such coordinative constructions ought to incorporate a stylistic analysis of the text, paying special attention to the specific grammatical and linguistic features of theonyms. The preferred coordinative construction for two theonyms is *dvandva* compound, followed by asyndeton, whereas syntagms constructed with copulative conjunctions are less common expression and elliptic duals are the most marginal option. In contrast, coordinative relationship between two non-theonyms is most commonly expressed by asyndeton, followed by coordinative conjunctions, whereas *dvandva* compounds are extremely rare (Ditrich, 2007).

2. Coordinative Particles in the *Ṛgveda*

Among coordinative particles used in the *Ṛgveda*, the most common is the particle *ca* (1094 attestations), followed by the particle *utá* (752 attestations), both usually used in the copulative sense “and” (Lubotsky, 1997, p. 344, p. 510). Coordinative particles are rather seldom used to express coordination between nouns in the *Ṛgveda*—they are outnumbered by the usage of asyndeta. In this article only particles used in a coordinative sense are examined, i.e. particles *ca*, *utá* and the rarely attested particles *u*, *na* and *ā*. It has been shown that coordinative nominal constructions with the particle *ca* (and less frequently *utá*), comprised of theonyms, seem to be one of the stylistic variants in hymns that address dual deities, alternating with *dvandva* compounds, asyndeta and, rarely, with elliptic duals (Ditrich, 2007).

Delbrück (1900, pp. 190–192) was the first to comment on the variety of coordinative constructions for two nouns, i.e. asyndeta, *dvandva* compounds and

[1] The research leading to the results in this paper has received funding from the Seventh Framework Programme [FP7/2007-2013] under PIRG02-GA-2007-224432.

syntagms constructed with copulative conjunctions. He thinks that these constructions express the same meaning; asyndeta are, in his opinion, the oldest Indo-European construction whereas the usage of conjunctions is a later development. His hypothesis is based on the development of Sanskrit: in older stages asyndeta are far more frequently used than syntagms with conjunctions whereas in the later language the usage of conjunctions increases. Renou (1955, p. 65) believes that particles may have a special emphatic function —to emphasize the link between the coordinated nouns; however, he provides no evidence to support his claim and, as Gonda (1971, p. 142) later comments, all attempts at reaching a clear understanding of the functions of *ca* and other particles have so far been unsuccessful.

There is a close relationship between syntactic constructions with the particle *ca*, and *dvandva* compounds. As already pointed out by Delbrück (1893, pp. 396–397), Ṛgvedic *ca* is a subclausal coordinative conjunction, very often coordinating pairs in the same case except for two vocatives. Most scholars explain that two vocatives cannot be connected with the subclausal conjunction *ca* because the constituents in vocatives are absolutives, bearing no integral constituent relationship to the rest of the sentence (Klein, 1981, pp. 73–91; Whitney, 1964, p. 90; Delbrück, 1900, p. 396). The syntactic construction of two theonyms in vocatives conjoined by *ca* does not occur in the *Ṛgveda*; on the other hand syntagms comprising a vocative and a nominative conjoined by *ca* are very common (V.+N.+*ca*), e.g. *vā́yav* [V.] *índraś* [N.] *ca* or inverted *índraś* [N.] *ca vā́yav* [V.]. These constructions, identified also in Old Greek and Avestan and seemingly of Indo-European origin, have been studied by many scholars (Humbach, 1982, pp. 95–102; Klein, 1981, pp. 73–91).

Klein (1981, pp. 85–87) comments that dual theonyms in syntagms V.+N.+*ca* usually occur elsewhere in the *Ṛgveda* alternatively as *dvandvas*. Jamison (1988, pp. 16–20) further develops this observation and demonstrates that syntagms V.+N.+*ca* can only be used in the *Ṛgveda* when the two nouns that are involved elsewhere, usually in the same hymn, also form a *dvandva* compound. The two constructions are always linked in discourse; there is a predictable relationship between the order of constituents in the *dvandva* compound and the case role that each element plays in the syntagm.

As pointed out by Jamison (1988, pp. 16–20), the pair Indra and Vāyu is disproportionately represented in syntagms V.+N.+*ca* and it is also the only pair that appears in both regular and inverted order (*vā́yav índraś ca* or *índraś ca vā́yav*). The reason for this, she argues, lies in grammar: the *dvandva* compound *indravāyú* is the only example among *devatādvandvas* of the *Ṛgveda* that has only one accent and the first constituent in stem form. With *indravāyú* operating more like a true morphological compound rather than like the almost chance asyndetic association seen in conventional *dvandvas*, the transformation into the *vā́yav índraś ca* construction may have seemed especially appropriate (Jamison, 1988, p. 20). She claims that all syntagms V.+N.+*ca* are based on related *dvandvas*. When one of the two theonyms, coordinated with *ca*, is in the plural no variant expression in a *dvandva* compound

occurs; e.g. *agniśca … marutaḥ* (RV 5.60.7), Agni and Marutas are never attested in a *dvandva* compound in the *Ṛgveda*.[2] However, Jamison (1988, pp. 22–30) argues that although this is true on the surface there is an underlying *dvandva* involving the plural member; she gives for evidence the post-Ṛgvedic *vṛddhi* derivations *āgnimārutá-* and *aindrārbhava-* that seem to be built to the underlying *dvandva* compounds. She thinks that there is no difference between coordinative constructions addressing theonyms and non-theonyms: gods are addressed in the same fashion as mortals in the *Ṛgveda* as well as in the proto-language (Jamison, 1988, p. 17). This statement seems to be too general, based only on examinations of syntagms V.+N.+*ca*.

The particle *utá* is the second most frequently used coordinative conjunction in the *Ṛgveda*. Klein (1978, pp. 1–23) gives an excellent survey of coordinative conjunctions in the *Ṛgveda* and identifies the main differences between the functions of *utá* and *ca*. He demonstrates that *ca* is mainly used as a subclausal conjunction whereas *utá* has much wider functions: it conjoins subclausal constituents as well as clauses and stanzas. This broader and more general function of the particle *utá* is the reason, as Klein (1978, pp. 1–23; 1981, pp. 77–78) argues, that *utá* can—unlike *ca*—coordinate two vocatives in the *Ṛgveda*.

Other particles have a wider spectrum of functions and are less frequently used in coordinative sense. It is often difficult to determine their function; it depends on the specific textual circumstances. The particle *u* is frequently used in the *Ṛgveda* (608 attestations) and has several functions, including a coordinative one. Klein (1978, pp. 9–23) convincingly demonstrates that the particle *u*, when not employed deictically or anaphorically, occupies virtually the same sphere of usage as *utá* and that the relationship between *u* and *ca* parallels that between *utá* and *ca*. Both *u* and *utá* have numerous attestations in the *Ṛgveda* but because of their wider functions they are far less used in coordinative nominal constructions than the particle *ca*.

3. Syntagms Constructed with Coordinative Particles in *Ṛgveda* 1.1 –1.50

Several coordinative particles are used in the first fifty hymns of the *Ṛgveda*: the most common is the particle *ca*, followed by the particle *utá* and the rarely attested particles *u*, *na* and *ā*. Here the attestations of all particles used in a coordinative sense that occur in *Ṛgveda* 1.1–1.50 are examined in two groups, those coordinating theonyms and those with non-theonyms.

[2] The only possible exception to this is *indrāmarutas* (RV 2.29.4.3) which is, as argued by Klein, "an isolated solecism" (Jamison, 1988, p. 22).

3.1 Theonyms

In the first fifty hymns of the *Ṛgveda* examined, six pairs of deities occur: Indra and Vāyu, Mitra and Varuṇa, Uṣas and Nakta, Indra and Varuṇa, Indra and Agni, and Dyaus and Pṛthivī. They are attested in a variety of coordinative constructions, the most common being *dvandva* compounds and asyndeta. The six pairs of deities that occur in *Ṛgveda* 1.1 –1.50 have their attestations in syntagms with coordinative particles distributed among the ten *maṇḍalas* as follows:

3.1.1 Indra and Vāyu

Table 1: Coordinative particles used in syntagms with Indra and Vāyu

maṇḍala	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**total**
ca	3	0	0	2	0	0	1	0	0	1	7
utá	0	0	0	0	0	0	0	0	1	0	1

Five constructions with the particle *ca* are in V. Sg. and N. Sg.: *vā́yavíndraśca* (1.2.5, 1.2.6, 4.47.3), *vāyo … índraśca* (1.135.4), *índraśca vāyo* (4.47.2). All these constructions also occur in the same hymn in *dvandva* compound, confirming the observations made by Jamison (1988, 16–20) that the two constructions (i.e. *dvandva* and syntagms with *ca*) are always linked in discourse. In two constructions the pair occurs in N. Sg. together with other deities (7.40.2 and 10.90.13) and once in I. Sg. in a syntagm constructed with *utá* (9.61.8 *índreṇa utá vāyúnā*). Indra and Vāyu are disproportionately represented in syntagms V.+N.+*ca*; they are also the only pair of deities that appears in both regular and inverted order (*vā́yav índraś ca* or *índraś ca vā́yo*).

3.1.2 Mitra and Varuṇa

Table 2: Coordinative particles used in syntagms with Mitra and Varuṇa

maṇḍala	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**total**
ca	2	0	0	0	3	0	2	1	2	1	11
ca … ca	0	0	0	0	1	0	0	0	0	0	1
utá	0	0	0	0	0	0	1	0	0	0	1

One construction with the particle *ca* occurs in V. Sg. and N. Sg.: *mitra … váruṇaśca* (5.64.5). Five syntagms have both constituents in N. Sg. (5.40.7, 5.68.2, 6.24.5, 7.66.18), one in A. Sg. (1.2.7), two in D. Sg. (9.100.5, 10.85.17), one in G. Sg. (1.136.2) and one in L. Sg. (9.61.9). Three syntagms (5.64.5, 7.66.18, 1.2.7) occur in

the same hymn also in *dvandva* compound. In one attestation the syntagm constructed with *ca* has one constituent in the dual and the other in the singular (RV 8.25.2 *mitrā́ … váruṇa … ca*), which is considered by some scholars the most archaic stage of development of *dvandvas*, i.e. an intermediate stage between elliptic duals and *dvandvas*. The order of constituents varies: the order Mitra and Varuṇa is more common (10 attestations) than Varuṇa and Mitra (3 attestations).

3.1.3 Uṣas and Nakta

Table 3: Coordinative particles used in syntagms with Uṣas and Nakta

maṇḍala	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**total**
ca	1	0	0	0	0	0	0	0	0	0	1

Only one construction with the particle *ca* occurs in N. A. Du. (1.73.7 *náktā ca … uṣásā*). It is attested in a hymn devoted to Agni which addresses several deities; the two theonyms do not occur in the same hymn in *dvandva* compound.

3.1.4 Indra and Varuṇa

Table 4: Coordinative particles used in syntagms with Indra and Varuṇa

maṇḍala	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**total**
ca … ca	0	0	0	0	0	0	1	0	0	0	1

Only one syntagm constructed with two particles *ca* is attested in the *Ṛgveda*, in A. Sg. (7.83.6 *índraṃ ca … váruṇaṃ ca*), in a hymn which addresses the pair; the theonyms Indra and Varuṇa occur in the same hymn also in *dvandva* compound.

3.1.5 Indra and Agni

Table 5: Coordinative particles used in syntagms with Indra and Agni

maṇḍala	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**total**
ca	0	0	1	0	0	1	0	0	0	0	2
ca … ca	0	0	0	0	1	0	0	0	0	2	3

One construction with the particle *ca* is attested in V. Sg. and N. Sg. (3.25.4 *ágna índraśca*) in a hymn devoted to Agni; the theonyms Indra and Agni do not occur in the same hymn in *dvandva* compound. One construction with the particle *ca* occurs in A.

Sg. (6.60.12 *índram agníṃ ca*) in a hymn addressing the pair; in this hymn Indra and Agni also occur alternatively in *dvandva* compound. All syntagms constructed with two particles *ca* are in N. Sg. (5.51.4, 10.90.13, 10.173.5 *índraścāgníśca*); they are attested in hymns that do not address the pair and in which the pair does not occur alternatively in *dvandva* compound. The order of constituents is, with one exception (3.25.4), Indra and Agni.

3.1.6 Dyaus and Pṛthivī

Table 6: Coordinative particles used in syntagms with Dyaus and Pṛthivī

maṇḍala	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**total**
ca	3	0	1	1	0	1	1	0	0	1	8
ca … ca	0	0	0	1	0	1	0	0	3	2	7
utá	21	0	3	1	2	2	0	0	2	6	37
ná	1	0	0	0	0	0	0	0	0	1	2
ā́	0	0	0	0	0	1	3	1	0	0	5

The most frequently used particle in coordinative constructions of Dyaus and Pṛthivī is *utá*. It is not evenly distributed among the ten *maṇḍalas*: 21.6% of all attestations of *utá* are in the family books which comprise 41.7% of all hymns in the *Ṛgveda*. It occurs most frequently (56.8% of all attestations) in *maṇḍala* 1, which comprises 18.6% of the total number of hymns. 21 attestations are in N. Sg. (always in the order: *pṛthivī́ utá dyaúḥ*), followed by 14 attestations in A. Sg. (always in the order *pṛthivī́m utá dyā́m* with one exception, i.e. 3.32.8 *pṛthivī́m dyā́m utá*) and 1 attestation in Ab. G. Sg. (*divás pṛthivyā́ utá*). Syntagms constructed with *utá* occur only in hymns addressing another deity / deities: in 9 hymns Dyaus and Pṛthivī occur in the same hymn also in *dvandva* compound, whereas in 28 hymns they do not.

Constructions with the particle *ca* are attested most frequently in *maṇḍala* 1; they occur with the constituents in N. Sg. (5 times), in A. Sg. (once) and Ab. G. Sg. (2 times). The order of the constituents does not vary: Dyaus always precedes Pṛthivī, while the particle *ca* either is placed between the theonyms (4 times) or follows them (4 times). Syntagms constructed with *ca* occur only in hymns addressing another deity / deities; the pair does not occur in the same hymn in *dvandva* compound.

Syntagms constructed with two of the particle *ca* have constituents in N. Sg. (3 times) and in A. Sg. (4 times). The order of constituents is always *dyauśca pṛthivī́ ca*. Most syntagms are in hymns addressing another deity / deities; the pair does not occur in the same hymn in *dvandva* compound. However, Dyaus and Pṛthivī coordinated with two of the particle *ca* are attested once in a hymn addressing another deity and in the same hymn the pair also occurs in *dvandva* compound; and once the syntagm

constructed with two of the particle *ca* occurs in a hymn that addresses Dyaus and Pṛthivī in which the pair is also expressed in *dvandva* compound.

Syntagms constructed with the particles *ná* and *ā́* are rare: both particles coordinate the theonyms Dyaus and Pṛthivī in Ab. G. Sg. only (*divó ná pṛthivyā́ḥ*, *divá ā́ pṛthivyā́ḥ*). Mostly the particle *ā́* occurs in the family books and *ná* in the younger *maṇḍalas* 1 and 10; however, the number of attestations is too small to draw any general conclusions about their distribution. All syntagms constructed with these two particles are attested in hymns addressing another deity / deities; the pair Dyaus and Pṛthivī does not occur in these hymns in *dvandva* compound.

Other theonyms occasionally—though rarely—occur in syntagms connected with coordinative particles. In the first fifty hymns examined there are only four attestations of such syntagms:

1. Ādityas and Indra: once in a syntagm constructed with two of the particle *ca*: in I. Sg. / Pl. (1.20.5 *índreṇa ca … ādithébhiśca*). These deities also have 7 attestations in asyndetic constructions but none in *dvandva*.

2. Ahi and Indra: once in a syntagm constructed with two of the particle *ca*: in A. Sg. (1.32.13 *índraśca ... áhiśca*); one other alternative construction is attested.

3. Agni and Mitra (and Varuṇa): once in a syntagm constructed with the particle *utá*: in N. Sg. / Du. (1.36.17 *agníḥ … mitrā́ utá*). These deities have one attestation in asyndeton but none in *dvandva*.

4. The Vasus, the Rudras and the Adityas: twice in a syntagm constructed with the particle *utá*: in A. Pl. (1.45.1) and in I. Pl. (10.125.1). These deities have also 10 attestations in asyndeton but none in *dvandva*.

From the survey of the syntagms constructed with coordinative particles for the six pairs of deities examined in the *Ṛgveda*, the following observations can be made:

The most frequently used particle for coordination of theonyms is *ca*: it is used exclusively to express coordination between Uṣas and Nakta, Indra and Varuṇa, Indra and Agni, and also, except for one attestation with *utá*, to coordinate the pair Indra and Vāyu, and the pair Mitra and Varuṇa. The distribution of the syntagms with coordinative particles for the six pairs of deities among the ten *maṇḍalas* seems generally very similar to the distribution of *dvandva* compounds formed from these theonyms (Ditrich, 2006); however, the number of attestations is too small to draw any general conclusion.

- The construction with *ca* usually occurs in the same hymn as the alternative construction in *dvandva* compound;[3] it indicates that the two

[3] All constructions with the particle *ca* of Indra and Vāyu, Indra and Varuṇa, Indra and Agni, and three of the total six of Mitra and Varuṇa occur in the same hymns also in *dvandva* compound.

coordinative constructions are stylistic variants, especially since they occur in the hymns that address the pair which follow specific stylistic patterns which include variations of different coordinative constructions for dual theonyms (i.e. several types of *dvandva* compounds, the elliptic dual, syntagms constructed with copulative conjunctions, and asyndeta) (Ditrich, 2007).

- The only pair of theonyms that shows a different pattern in the usage and distribution of coordinative particles is Dyaus and Pṛthivī. This pair occurs in the largest number of syntagms constructed with coordinative particles, most frequently with the particle *utá*.[4] Constructions with *utá* are mostly in N. and A. Sg. with the highest frequency in *maṇḍala* 1; these syntagms usually do not occur in the same hymn as *dvandva* compounds or elliptic duals. Dyaus and Pṛthivī—unlike other dual theonyms—are coordinated by a variety of particles: one *ca*, two *ca*, *utá*, *ná* and *ā́*. Other dual theonyms only occasionally occur in syntagms constructed with *utá* or with two of the particle *ca*.
- Generally, it remains uncertain what the function of the various particles coordinating theonyms is, but it seems that one function of syntagms comprising dual theonyms is stylistic variation, especially since these variants often occur in the same hymn, usually addressing the very same pair of deities.

3.2 Non-theonyms

Coordinative particles are used in the *Ṛgveda* also to express coordination between nouns which are not theonyms, as well as — though more rarely — between adjectives, adverbs, verbs and sentences. All the particles coordinating non-theonyms that are attested in *Ṛgveda* 1.1–1.50 are identified below and all other coordinative constructions between the two examined non-theonyms are drawn from the entire text of the *Ṛgveda*.

3.2.1 Particle *ca*

1. Coordination of nouns (non-theonyms):

1. *sahásrapradhana-* and *vā́ja-*: 1 attestation in I. Pl. (1.7.4 *vā́jeṣu … sahásrapradhaneṣu ca*); no alternative coordinative construction is attested in the entire text of the *Ṛgveda*.

[4] The particle *utá* coordinates Dyaus and Pṛthivī 37 times but other dual deities only twice.

2. *ukthá-* and *stóma-*: 2 attestations: in N. Sg. (1.8.10 *stómaḥ ukthấṃ ca*) and in I. Pl (6.24.7); 4 attestations in asyndeton in the entire *Ṛgveda* (1.5.8, 3.5.2, 3.41.4, 6.23.1).

3. *sutá-* and *sakhyá-*: 1 attestation in I. Pl. (1.10.5 *sutéṣu ... sakhyéṣu ca*); no alternative coordinative construction is attested.

4. *yajñá-* and *havís-*: 1 attestations: in N. Sg. (1.12.10 *yajñáṃ havíṣca*); 4 attestations in asyndeton in the *Ṛgveda* (1.24.14, 2.35.12, 4.50.6, 10.20.6).

5. *bheṣajá-* and *agní-*: 1 attestation in N. Pl. (1.23.20 *bheṣajấ agníṃ ca*); no alternative coordinative construction is attested.

6. *apacyavá-* and *upacyavá-*: 1 attestation in A. Sg. (1.28.3 *apacyavám upacyavấṃ ca*); no alternative coordinative construction is attested.

7. *práyas-* and *máyas-*: 1 attestation in N. A. Sg. (1.31.7 *máyaḥ … práyaḥ ca*); no alternative coordinative construction is attested.

8. *tanūkṛ́t-* and *prámati-*: 1 attestation in N. Sg. (1.31.9 *tanūkṛ́t … prámatiśca*); no alternative coordinative construction is attested.

9. *maghávan-* and *tanū́-*: 1 attestation in A. Pl. (1.31.12 *maghónaḥ … tanváśca*); no alternative coordinative construction is attested.

10. *míh-* and *hrādúni-*: 1 attestation in A. Sg. (1.32.13 *míham … hrādúniṃ ca*); no alternative coordinative construction is attested.

11. *amṛ́ta-* and *mártya-*: 1 attestation in A. Sg. (1.35.2 *amṛ́taṃ mártyaṃ ca*); no alternative coordinative construction is attested.

12. *arká-* and *ukthá-*: 1 attestation: in I. Pl. (1.47.10 *ukthébhiḥ … arkaíśca*); 1 attestation in asyndeton in the entire *Ṛgveda* (6.5.5).

13. *hṛdrogá-* and *harimán-*: 1 attestation in A. Sg. (1.50.11 *hṛdrogám … harimā́ṇaṃ ca*); no alternative coordinative construction is attested.

2. Coordination of adjectives:

1. *śáma-* and *śṛṅgín-*: 1 attestation in G. Sg. (1.32.15 *śámasya ca śṛṅgíṇaḥ*); no alternative coordinative construction is attested.

2. *rudát-* and *jákṣat-*: 1 attestation in A. Pl. (1.33.7 *rudató jákṣataśca*); no alternative coordinative construction is attested.

3. Coordination of adverbs:

1. *adyá* and *nūnám*: 1 attestation (1.13.6 *adyá nūnáṃ ca*); no alternative coordinative construction is attested.

2. *doṣā́* and *uṣás-*: 1 attestation in A. Pl. / G. Sg. (adverbial use: 1.34.3 *doṣā́ḥ … uṣásaśca*); 5 attestations in adverbial function in asyndeton in the entire *Ṛgveda* (1.179.1, 2.8.3, 4.2.98, 7.3.5, 8.22.14).

4. Coordination of verbs:

4 attestations of two verbs (X+Y+*ca*): 1.14.1c, 1.31.17cd, 1.42.9ab, 1.48.3ab;

2 attestations of two verbs (X+*ca*+Y): 1.15.9b, 1.17.6ab.

5. Coordination of sentences:

6 attestations of X+*ca*+Y: 1.8.5ab, 1.23.21abc, 1.25.11c, 1.25.19ab, 1.26.8abc, 1.34.12cd;

2 attestations of X+Y+*ca*: 1.13.1abc, 1.23.20abc.

3.2.2 Two of the particle *ca*

1. Coordination of nouns (non-theonyms):

1. *brahmán-* and *yajñá-*: 1 attestation in A. Sg. (1.10.4 *bráhma ca … yajñáṃ ca*); 1 attestation with one *ca* in the entire *Ṛgveda* (6.38.40).

2. *pitṛ́-* and *mātṛ́-*: 3 attestations in A. Sg. (1.24.1 1.24.2, 10.54.3 *pitáraṃ ca … mātáraṃ ca*); 5 attestations with one *ca*, 1 with *utá*, 15 in asyndeton, 83 in the elliptic dual, 1 in *dvandva* in the entire *Ṛgveda.*

3. *dyú-* and *kṣam-*: 5 attestations in Ab. G. Sg. (1.25.20, 1.37.6, 1.100.15, 5.38.3, 10.49.2, 10.22.6 *diváśca gmáśca*); 1 attestation with one *ca*, 3 in asyndeton in the entire *Ṛgveda.*

4. *náva-* and *navatí-*: 1 attestation in N. Sg. (1. 32.14 *náva ca navatíṃ ca*); 9 attestations with one *ca*, 9 in asyndeton in the entire *Ṛgveda*.

2. Coordination of adverbs:

1. *dūrā́t* and *āsā́t*: 1 attestation (1.27.3 *dūrā́ccāsā́cca*); 1 attestation with ā́ (4.20.1).

3. Coordination of sentences:

1 attestation of X+*ca … ca*+Y: 1.35.11cd.

3.2.3 Particle *utá*

1. Coordination of nouns (non-theonyms)

1. *yā́ma-* and *rātí-*: 1 attestation in N. Sg. (1.34.1 *yā́ma utá rātiḥ*); no alternative coordinative construction is attested.

2. *toká-* and *vásu-*: 1 attestation in A. Sg. (1.41.6 *vásu … tokámutá*); no alternative coordinative construction is attested.

2. Coordination of adjectives

pū́rva- and *nū́tana-*: 1 attestation in I. Pl. (1.1.2 *pū́rvebhiḥ … nū́tanaiḥ utá*); 2 attestations in asyndeton in the entire *Ṛgveda*.

3. Coordination of adverbs:

1. *adyá* and *aparám*: 1 attestation (1.36.6 *adyá … utá aparám*); 2 attestations in asyndeton in the entire *Ṛgveda* (1.184.1, 8.27.14).

4. Coordination of sentences:

6 attestations X+*utá*+Y: 1.10.6cd , 1.11.8cd, 1.23.19ab, 1.32.13cd, 1.34.5b, 1.34.5c; 3 attestations of X+Y+*utá*: 1.17.6ab, 1.32.4ab, 1.39.2ab.

5. Coordinations of stanzas:

at the beginning of the stanza: 1.4.5ab, 1.4.6ab, 1.20.6ab, 1.25.15ab, 1.28.6ab, 1.31.18cd.

3.2.4 Particle *u*

1. Coordination of sentences:

2 attestations of X+*u*+Y: 1.34.2d, 1.34.6ab.

3.2.5 Particle *ā́*

1. Coordination on nouns:

ánta- and *parāká-*: 1 attestation in Ab. Sg. (1.30.21 *ántādā́ parākā́t*); no alternative coordinative construction is attested.

From the survey of the particles coordinating non-theonyms that are attested in *Ṛgveda* 1.1–1.50, the following observations are made:

- The particle *ca* is the most frequently used particle for coordination of non-theonyms in *Ṛgveda* 1.1–1.50: it has 45 attestations and most frequently coordinates nouns (13 attestations). Syntagms comprising two nouns coordinated with the particle *ca* seem to express a casual relation between the nouns; usually they have only one attestation in the whole *Ṛgveda* and do not occur in any other coordinative construction (e.g. asyndeton or *dvandva).*
- Only 3 out of a total of 13 pairs of nouns coordinated with *ca* occur also in asyndeton: these nouns have a larger number of attestations in

asyndeton than in constructions with *ca*.[5] In these syntagms nouns occur most frequently in N. and A. Sg. and I. Pl.

- The particle *ca* is also used to coordinate adjectives (2 times) and adverbs (2 times); these syntagms have only one attestation in the whole *Ṛgveda* and do not occur, with one exception, in any other coordinative construction.
- The particle *ca* also coordinates verbs (6 times) and sentences (8 times); the frequent use of *ca* here does not support the claim developed by Klein (1978, pp. 1–23) that *ca* is mainly used as a subclausal conjunction whereas *utá* has much wider functions, i.e. it conjoins subclausal constituents as well as clauses and stanzas. In *Ṛgveda* 1.1–1.50 the particle *ca* is used for coordination of sentences as frequently as the particle *utá*.
- The particle *utá* is less frequently used for coordination of nouns (2 attestations), adjectives (1 attestation) and adverbs (1 attestation) than for coordination of sentences. It also occurs in its special function as a coordinative particle situated at the beginning of a stanza (6 attestations). Syntagms comprising two nouns coordinated with the particle *utá* seem to express a casual relation between the nouns; usually these nouns occur only once in the coordinative relation and do not have attestations in any other coordinative construction (e.g. asyndeton or *dvandva*). The cases most frequently used in these syntagms—as in those constructed with *ca*—are N. and A. Sg. and I. Pl.[6]
- Two of the particle *ca*, attested in *Ṛgveda* 1.1–1.50, are mainly used for coordination of nouns (4 attestations) and, less frequently, coordination of adverbs (1 attestation) and sentences (1 attestation). These syntagms—unlike those constructed with one *ca*—have several attestations in the *Ṛgveda* and do occur in other coordinative constructions as well: all of them have at least one attestation with one *ca*, often they occur in asyndeton and, in the case of *pitṛ́-* and *mātṛ́-*, also in the elliptic dual and *dvandva*. It seems that syntagms constructed with two *ca* do not coordinate nouns casually, as do those with one *ca* or with *utá*, but are one of the stylistic variants expressing coordination between nouns and are especially often attested in *maṇḍalas* 1 and 10. Examination of all syntagms constructed by using two *ca* from *Ṛgveda*

[5] E.g. *ukthá-* and *stóma-* (2 attestations with *ca*, 4 attestations in asyndeton), *yajñá-* and *havís-* (1 attestation with *ca*, 4 attestations in asyndeton), *arká-* and *ukthá-* (1 attestation with *ca*, 1 attestation in asyndeton).

[6] The usage of cases in syntagms constructed with coordinative particles and in asyndeta has not been, to my knowledge, investigated; this is an area which requires future research that would contribute to a better understanding of the style of the Ṛgvedic poetic language.

1.1–1.50 indicates that these constructions are different from the syntagms constructed with other coordinative particles and have a special function which may be the stylistic variation.

4. Conclusion

Several coordinative particles are used in the hymns examined: the most common is the particle *ca*, followed by the particle *utá*. When dual theonyms occur in constructions with the coordinative particle *ca*, they are usually attested in the same hymns in *dvandva* compound as well, thus indicating that the two coordinative constructions are stylistic variants, especially in the hymns that address the pair in question. Dyaus and Pṛthivī is the only pair of deities that is coordinated by a variety of conjunctive particles (one particle *ca*, two particles *ca*, particles *utá*, *ná* and *ā́*), displaying the widest variety of stylistic expressions. It remains uncertain what the function of the various particles coordinating theonyms is; however, it seems that one function of syntagms comprising dual theonyms is stylistic variation, especially since these syntagms usually occur together with other coordinative constructions (*dvandvas*, elliptic duals, asyndeta) in the same hymn.

Syntagms comprising two non-theonyms coordinated with the particle *ca* or *utá* seem to express a casual relation between the nouns; they have usually only one attestation in the entire *Ṛgveda* and do not occur in any other coordinative construction. Unlike syntagms constructed with one particle *ca*, those constructed with two particles *ca* usually have several attestations in the *Ṛgveda* and also occur in other coordinative constructions (i.e. in syntagms constructed with one particle *ca*, in asyndeton and sometimes also in the elliptic dual and *dvandva* compound). It seems that the syntagms constructed with two particles *ca* do not coordinate nouns casually, but represent one of the stylistic variants expressing coordination between the nouns.

There is a considerable difference in style between coordinative constructions comprising two theonyms, and those consisting of non-theonyms, as demonstrated by the distribution and usage of coordinative particles in the first fifty hymns of the *Ṛgveda*. By examining how two words coordinated with particles alternatively occur in other coordinative constructions, the article demonstrates that research into coordinative constructions in Vedic ought to pay special attention to the specific grammatical and linguistic features of theonyms that distinguish them from non-theonyms.

Abbreviations

A. accusative
Ab. Ablative
D. Dative

Du.	Dual
G.	genitive
I.	instrumental
L.	Locative
N.	Nominative
Pl.	Plural
RV	*Ṛgveda*
Sg.	Singular
V.	vocative

References

Delbrück, B. (1893). *Vergleichende Syntax der indogermanischen Sprachen*, Theil 1. Strassburg: Karl J. Trübner.

Delbrück, B. (1900). *Vergleichende Syntax der indogermanischen Sprachen*, Theil 3. Strassburg: Karl J. Trübner.

Ditrich, T. (2006). *Dvandva* Compounds and the Chronology of the *Ṛgveda*. *Crossroads, 1* (Autumn 2006), 26–35.

Ditrich, T. (2007). Stylistic Analysis of Coordinative Nominal Constructions for Dual Deities in the *Ṛgveda*. In *Proceedings of the 13th World Sanskrit Conference*, Edinburgh, July 2006. Delhi: Motilal Banarsidass. [forthcoming]

Ditrich, T. (2010). The Variety of Expression for Heaven and Earth in the *Ṛgveda*. *Crossroads, 5*(1), 35–44.

Gonda, J. (1971). *Old Indian*, Handbuch der Orientalistik, zweite Abteilung, 1. Band, 1. Abschnitt. Leiden-Köln: E. J. Brill.

Humbach, H. (1982). Vedic índraśca vāyo ~ Old Avestan mazdāscā ahurāŋhō. Münchener Studien zur Sprachwissenschaft, 41, 95–102.

Jamison, S. W. (1988). Vāyav Indraś ca Revisited. *Münchener Studien zur Sprachwissenschaft, 49*, 13–59.

Klein, J. S. (1978). The System of Coordinative Conjunctions in the Rigveda. *Indo-Iranian Journal, 20*, 1–23.

Klein, J. S. (1981). The Origin of the Rigvedic *vā́yav índraś ca* Construction. *Münchener Studien zur Sprachwissenschaft, 40*, 73–91.

Lubotsky, A. (1997). *A Ṛgvedic Word Concordance* (Vol. 1). American Oriental Series 82. New Haven, Connecticut: American Oriental Society.

Renou, L. (1955). *Études védiques et pāṇinéennes* (Vol. 1). Publications de l'Institut de Civilisation Indienne, série in-8, fasc. 1. Paris: Éditions E. de Boccard.

Whitney, W. D. (1964). *Sanskrit Grammar*, 2nd ed. Cambridge, Mass.: Harvard University Press.

LINGUISTIC REPRESENTATION OF EMOTIONS IN JAPANESE AND HUNGARIAN: QUANTITY AND ABSTRACTNESS

Márton SZEMEREY
Károli Gáspár University of the Reformed Church in Hungary
mszemerey@gmail.com

Abstract

In the present paper, two linguistic aspects of emotion expression are studied in the form they are performed in present day Japanese and Hungarian. After a brief summary on the recent emotional researches connected to Japanese culture and language, the concept of Linguistic Category Model is introduced. The quantitative study presented afterwards investigates emotion expression in terms of amount and abstraction. Translations were used for comparison and the results showed that 1) Japanese tend to use less explicit emotion terms compared to Hungarians and 2) emotion language in Japanese is characterized by the choice of less abstract phrases compared to Hungarian. These findings are discussed in the light of their relevance to former researches of cross-cultural psychology and linguistics.

Keywords

Emotion; Emotion regulation; Linguistic Category Model; Linguistic Abstraction

Izvleček

Članek razpravlja o dveh lingvističnih vidikih čustvenih izrazov v japonščini in madžarščini. V uvodu avtor predstavi dosedanje raziskave s tega področja, ki zadevajo japonski jezik, in se osredotoči na t.i. Lingvistični kategorični model. V kvantitativni raziskavi spremlja pojavnost in abstrakcijo čustvenih izrazov – za primerjavo med jezikoma so uporabljeni prevodi – in rezultati so pokazali dvoje. Prvič, japonci v primerjavi z madžari uporabljajo manj abstraktne čustvene izraze, in drugič, čustveni jezik je v japonščini opisan z manj abstraktnimi frazami kot v madžarščini. Ugotovitve so na koncu analizirane v luči dosedanjih raziskav s področij psihologije in lingvistike.

Ključne besede

Čustva; ureditev čustvenih izrazov; Lingvistični kategorični model; lingvistična abstraktnost

1. Introduction

The way emotions appear in the communicational habits of Japan is a largely investigated field in both linguistics and psychology. A number of studies focus on the lexical and grammatical features characterizing emotion terms in Japanese (Maeda, 1993; Ohso, 2001, Kiyomi, 2006), and the semantic structure of emotion expression is also closely connected to this topic (Romney, Kimball, Rusch, 1997). In cross-cultural psychology, several authors describe a specific range of emotions, more accessible than others to Japanese (Doi, 1981; Morschbach & Tyler, 1986; Markus & Kitayama, 1991; Kitayama, Mesquita, Karasawa, 2006). Other studies discuss the Japanese cultural scripts, which guide the order of expressing and understanding emotions (Wierzbicka, 1996), emotion inference (Uchida et al., 2009) and co-occurrence of positive and negative emotions in Japan (Miyamoto, Uchida, Ellsworth, 2010; Uchida & Kitayama, 2009).

Emotion regulation is another research area, which has been given considerable attention recently. Extensive studies have shown, that emotional display rules in Japan result in a reduced expression of negative emotions compared to U.S. and Canada (Safdar et al., 2009). Among members of other cultures that value hierarchy and emphasize long-term orientation, Japanese have also been reported to get relatively high scores on both reappraisal and suppression, i.e. a frequent application of the tools for emotion regulation was found in Japan (Matsumoto et al., 2008). Furthermore, studies using retrospective memory tests show a tendency to dampen positive emotions in the case of participants from Japan along with those from China and Korea (Miyamoto & Ma, 2011). These results are in line with the findings of Ekman (1972) about emotional display rules allowing Japanese a reduced expression of disgust when engaging in social interactions.

The above mentioned psychological researches examine and compare the way Japanese and non-Japanese people customarily express emotions during events of everyday communication. To do so, sometimes participants were asked to judge the appropriateness of showing certain emotions in certain situations, or their facial expressions and bodily gestures were analyzed under artificially created conditions, while the question of emotional language use was often left untouched. Considering the difficulties of coding every feature of speech that can have an emotional implication, such as tempo, intonation, accent, tone, etc., the decision to design studies that are less dependent on verbal aspects of the interactions is not at all implausible. However, unlike the above interactions, that have the merits of both verbal and nonverbal communication, any classical genre of written communication renders emotion expression unpronounced and almost exclusively linguistic.[1] Thus it seems to

[1] The sudden spread of emoticons, the pictorial representations of facial emotion expressions, like =) 'smile' has created a new method of displaying feelings. Nevertheless, emoticons appear mostly in online communication, whereas offline writings of any genre still use primarily language as a tool of emotion expression.

be reasonable to ask, whether the tendencies found in a series of basically nonlinguistic researches are also reflected in the emotion expression of written language.

Before exploring that question it may be beneficial to discuss an instrument the linguistic representation of emotions can be examined with.

2. The Linguistic Category Model

The Linguistic Category Model (Semin & Fiedler, 1988, 1991; henceforth: LCM) is a framework of vocabulary analysis that makes it possible to decipher the linguistic representation of an interaction in terms of psychological concreteness and abstractness. Concreteness here refers to form of describing an interaction in a way that provides information more or exclusively about the action and less or not at all about the character of the agent. A concrete statement does not allow the listener to value the personality of the agent as such a description refrains from exceeding beyond the scope of an explanation of the interaction on the physical level ("A gave a rose to B"). A word-choice of that kind makes it unclear what kind of person A is, and also the speaker's evaluation about the interaction remains obscure (Bigazzi & Nencini, 2008, p. 100). I.e. in different contexts giving a rose can be a sign of courtesy, affection, congratulation, insult, threat, etc., but without additional information about the situation the listener can know neither the exact meaning of the act nor the psychological features of the agent. It is possible though to explain the same situation including the speaker's interpretation about the scene ("A congratulated B"), or describing it not so much as an action, but as a state, and thus offering information in relation to the somewhat long-lasting inner conditions of the agent ("A likes B"). In LCM such a interpretative formulations are understood as more abstract, as the listener has the opportunity to evaluate both the interaction and the character of the agent. The latter wording is least concrete among these in that it tells the listener less about the interaction and more about the agent. The less information about the situation is needed to obtain knowledge about a described person's characteristics and features, as supposed by the speaker, the more abstract the linguistic forms are considered in LCM. The most conceptual and thus abstract forming of the description would be an overt, general evaluation containing only the agent (Semin, 2004) and requiring no further information to draw an inference about his general attitudes ("A is a kind person").

The LCM divides words describing interactions into the following six categories:

a) Descriptive Action Verb (DAV): Verbs describing an action with a clear beginning and end, containing an invariant physical element. E.g., hit, kick, walk

b) Interpretative Action Verb (IAV): Verbs describing a number of actions with a clear beginning and end, without an invariant physical element, reflecting the speaker's interpretation about the interaction. E.g., help, congratulate, hurt

c) State Action Verb (SAV): Verbs describing an action but referring more to the emotional state resulting from it. E.g., surprise, frighten, fascinate

d) State Verb (SV): Verbs describing an enduring cognitive or emotional state, without a clear beginning or ending. E.g., love, hate, admire

e) Adjective (ADJ): Adjectives describing characteristics of a person.[2] E.g., kind, depressed, precise

f) Noun (N): Nouns describing characteristics of a person. E.g., thief, criminal, angel

These categories are further classified into four levels of abstractness. When coding a word in terms of concreteness-abstractness, DAV has score 1, IAV and SAV representing only slight difference in abstraction both have score 2, SV has score 3, whereas ADJ and N have score 4 again both being almost equally abstract in a psychological sense. Adverbs are coded and scored as ADJ.[3] It is notable that despite abstraction here refers to a psychological feature, as it is displayed by linguistic markers, in the LCM terminology it is referred to as linguistic abstraction.

During the last twenty five years, LCM has proven a useful tool to reveal new aspects of several phenomena, such as attribution (Semin & Fiedler, 1989), stereotype formation (Maass, Salvi, Arcuri, Semin, 1989; Maass, 1999; Tanabe & Oka, 2001) and person-representation (Maass, Karasawa, Politi, Suga, 2006). Another field of application was the linguistic representation of emotions (Semin, Görts, Nandram, Semin-Goossens, 2002). In their study, Semin et al. examined cultural perspectives of emotion expression and found systematic differences between cultures that value relationships and interdependence (e.g., Turkish and Hindustani-Surinamese people) and those emphasizing more the value of the individual (e.g., Dutch people). While the formers seem to use emotion expressions in the function of relationship-markers, the members of the latter group are likely to use emotions to mark the self and tend to describe emotionally filled events of interaction with more abstract linguistic categories (ADJ, N).

This section has been devoted to the LCM and a brief introduction to a research that contributes directly to the understanding of cross-cultural patterns in emotion expression on the basis of that model. It can be seen, that LCM helps us decode emotion expressions in terms of abstractness. This is a different dimension, compared to the emotion display rules, discussed in the Introduction. On the one hand, rules of emotion regulation give us a hint on the amount of the emotion terms. On the other

[2] In e) and f) *person* refers to any kind of agent in an interaction, including animals, natural phenomena, etc.

[3] In the case of Japanese, much onomatopoeia belong to explicit emotion terms what makes it necessary to include them in the above categorizing and scoring system. However, as in the chosen text no such term appears, that question is not discussed in the present paper.

hand, LCM shows the degree of concreteness of emotion terms used in a language. Thus we arrived to the two topics to be discussed below: abstraction and quantity.

3. Hypotheses

(1) Emotion regulation characterizing a culture should have a measurable impact on the linguistic representation of emotions in that culture. In case of Japan less emotion terms should appear in a Japanese text compared to its corresponding version produced in a Western language.

(2) Cultures that emphasize interdependence tend to avoid direct evaluation through highly de-contextualized representation of emotions and rather refer to the present situation, or the present relationship. Thus emotion expression in Japanese should be formulated in less abstract terms compared to those in a Western language.

4. Method

Participants. 5 female (Mean age=22 years; Standard Deviation=1,581) and 5 male (Mean age=24,8 years; Standard Deviation=4,086) students from universities in Budapest volunteered to participate in this study. All of them were majoring Japanese language and were in the last year before, or the year after graduation (Mean time of Japanese language learning=31,1 months; Standard Deviation=9,338). All of them are native Hungarians without considerable experience in professional translation.

Procedure and materials. In order to produce Japanese and Hungarian texts corresponding in terms of topic, style and length the study was based on translation. A part (330 words) of the popular novel *Kitchin* (Yoshimoto, 1991) was chosen. As the main interest of the study lies in the expression of emotions, the part was selected with regard to the emotions appearing in it. A monologue of an emotion event occupies more than half of the text and the narrator reports her feelings at the end of the section as well. A literal translation of the text was made to Hungarian, strictly preserving part of speech, number and case of every word.

To identify emotion expressions in this translation we used the interface of the program NooJ (www.nooj4nlp.net), an instrument that allows one to construct large-coverage dictionaries and complex grammars in order to parse corpora. Hungarian version of NooJ makes automatic annotation of emotion terms possible. Using a general Hungarian dictionary part of speech, case, and number of each word was identified. This linguistic analysis was followed by a selection of emotion terms based on the NooJ dictionary of Hungarian emotion expressions. The automatic analysis was

completed by a manual recheck of the emotion terms. At that point the concordance of explicit emotion terms, including 10 items altogether, was completed.[4]

In the next phase participants were asked to create a translation from the original Japanese text to Hungarian. The instruction was to aim for a literary translation that would convey the atmosphere of the text rather than a word-to-word translation. Through computational analysis and manual selection of these translations 10 concordances of emotion terms were produced. These wordlists varied in length from 8 to 14 items.

Concordances provided the possibility of 1) immediate quantitative comparison of explicit emotion terms, 2) analysis of corresponding terms and those not having an equivalent in the original text, and 3) the calculation and comparison of the level of linguistic abstraction in the 11 texts.

To test the results, the same excerpt from the English translation of the novel (Yoshimoto, 2001; translator: Backus, M.) was literally translated to Hungarian and analyzed the above explained way.

5. Results

Comparing the quantity of explicit emotion terms, an average surplus of 10% was found in participants' translations (109 phrases in the 10 texts).

The analysis of corresponding emotion terms has shown that explicit emotion expressions in the original text appeared in most cases also in Hungarian as emotion terms. However, in 12 cases explicit emotion phrases were missing from the different translations of the participants. Note that 7 out of 12 were missing from translations with at least 10 explicit terms. On the other hand, one part of the original text, that did not contain explicit reference to feelings, was translated using emotion terms by 6 participants.

[4] *Explicit* here refers to the distinction between words carrying emotional meaning by definition, such as 'anger', 'happy' and 'cry', and other linguistic markers of emotions, that are not necessarily lexicalized and are highly dependent on the context to function successfully as emotion markers. Markers that belong to the latter group are called *implicit* emotion expressions in the present paper. For a more detailed explanation see Bodor (2004).

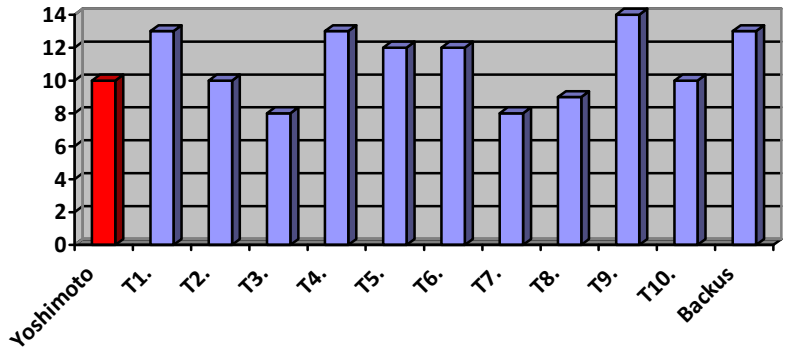

Figure 1: Number of explicit emotion expressions in the translations

Figure 1 shows the number of explicit emotion terms used in the 12 texts. Yoshimoto refers to the literal Hungarian translation of the original text, T1.-10. show the participants' translations of the original Japanese text to Hungarian, whereas Backus stands for the literal translation to Hungarian from the English edition of the novel.

A one-sample t-test using the level of linguistic abstraction in the original text (3,3) and the mean level abstraction of the translations (3,507; Standard Deviation=0,246) yielded a significant difference ($t(9)=2,653$; $p<0,05$) in the linguistic abstraction of the Japanese text and the participants' Hungarian translations.

Results of the text translated from English correlate with those of the participants' translations. A surplus of 30% was identified in the text, with emotion terms appearing here too at the paragraph without explicit emotion terms in the original and lacking equivalent of one term of the original text. The t-test also revealed that the level of abstraction (3,61) does not differ significantly from that of the Hungarian translations ($t(9)=-1,32$; $p>0,05$ n.s.), though it does from the Japanese.

6. Discussion

The data about the quantity of explicit emotion terms in the texts seem to support the hypothesis about the correlation of emotion regulation and linguistic representation of emotions. Nevertheless we have to consider the questions of (1) how can we account for the surplus in the Hungarian and English translations and (2) what could explain the missing emotion terms in the same translations?

For the first question a twofold answer may be given, in accordance with the nature of linguistic emotion representation. Extra emotion terms appear mostly as adjuncts of expressions that are translated as an equivalent of emotion words used in

the original text. These terms seem to have the function of strengthening the effect of the equivalent, as if one word would not suffice to create the same atmosphere in the reader. An example of this can be seen in (1) and (2), the emotion terms underlined:

(1) となりに人がいては淋しさが増すからいけない。 (Yoshimoto, 1991, p. 25)
Tonari ni hito ga ite wa sabishisa ga masu kara ikenai.
"If there is someone next to me, sadness increases, so I cannot [be close to anyone]"

(2) I was too sad to be able to sleep in the same bed with anyone; that would only make the sadness worse. (Backus' translation; Yoshimoto, 2001, p. 16)

The terms "sad" and "sadness" both belong to the category referred to in the present paper as explicit emotion expressions and their relation to an emotion term in the original text can be identified with ease. This is the case by the majority of the surplus. Still, there is a group of additional emotion words in the translation, which seem to lack a straight connection to any of the original expressions and appear in the translation of the same paragraph.

An explanation for this may be the use of implicit emotion expressions, a group of lexical, grammatical and possibly even structural linguistic markers that add an invariable emotional flavor to an utterance. As Nomura (2003, p. 37) proposes, "it is recommended, that emotion terms are not limited to the classical types of emotion- and wish expressions, but one should understand them as one type of expressions showing emotions, i. e., temporary movements and states of a person's mind, verbalized in a text." In other words, he refers to an extended range of linguistic devices as emotion terms, similar to what we call implicit emotion expressions here. Nomura, examining the introductory sentences of the same novel by Yoshimoto states, that not only phrases, like "*tamaranaku suki da*" ("I love it desperately") and "*sabishiku hoshi ga hikaru*" ("stars are shining lonely") belong to emotion expressions, but considering their role in maintaining the emotional atmosphere of the narration, even words such as "*kagayaku*" ("sparkle"), "*ii*" ("good") and "*motarekakaru*" ("lean against") should be counted among emotion expressions. The paragraph in question, describing the night view of the city, contains several words and phrases that can fall into the above explained category, such as "*kasuka na akari ni ukabu shokubutsutachi*" ("plants floating in the dim light"), "*sotto ikizuite ita*" ("breathing softly"), "*tōmei na kūki ga kirakira kagayaite*" ("the transparent air sparkling brilliantly") and "*migoto ni utsutte ita*" ("was reflected beautifully"). These may imply feelings in the readers and many translators would even find words for those impressions.

As for the function of such implicit emotion expressions, Nomura (2003, p. 38) calls the attention to their close relation to the underlying evaluation that frames the narrated episodes. As he states, "to the extent [such expressions] share the connectedness of this evaluation, in the text there is an inherent structuring effect of the emotion expressions." Emotions thus seem to play an important role in creating the unity of story-telling. Narratives consist of episodic scenes, which in turn may be defined by their emotional atmosphere. If so, implicit emotions are expected to appear

in any language at those parts of the narration, where for some reason explicit representation of feelings cannot take place. This may be related to the culture-specific rules of emotion regulation on the one hand, but it is most probably a universal feature of language use at the same time, considering the possibility that any language is likely to be dependent on indirect expression of feelings to some extent.

The idea of implicit and explicit emotion terms constituting an emotional and evaluative frame for segments of narration may be helpful in explaining the missing emotion words as well. It is notable, that emotion words explicitly stated in the original were left out from the translations only when a series of such expressions were concentrated within a sentence or sentences following each other. This can be illustrated by (3), a clause in which the first one of the two expressions was changed to a non-explicit emotional term in 6 translations.

(3) お気に入りの台所に立てたうれしさで目がさえてくると…
Okiniiri no daidokoro ni tateta ureshisa de me ga saete kuru to…
"As my eyes cleared with the happiness of standing in my beloved kitchen…"

An example of changing an emotional term into an evaluative one can be seen in (4), a sentence from the text of T5.

(4) Amint a számomra oly' eszményi konyhában boldogan álltam, hirtelen…
"As I was standing happily in the kitchen so ideal to me, suddenly…"

The fact, that more than half of the translators decided to change one of the two explicit emotion terms in that clause to an evaluative phrase that may implicitly carry emotions suggests, that emotions require a somewhat even expression on the explicit level. The other half of the terms not to be found in the translations have become invisible in a similar, emotionally concentrated context in the first monologue of the text.[5]

Through including a literal Hungarian translation of the English version by Backus, a comparison of the Japanese, Hungarian and English results become possible. 13 explicit emotion terms are used in the English translation, what is on a parallel with many of the participants' works (13 emotions terms by T1 and T4, 14 by T9, 12 by T5 and T6). A considerable surplus in English is also comparable to the results of the studies on emotion regulation cited in the introductory section.

Finally, the question of linguistic abstraction is to be discussed. The significant difference in the level of abstraction between the original text and the Hungarian and English translations appears to provide evidence that a more concrete formulation of

[5] Obviously, there is a possibility of the translators' personality influencing word-choice, but in that case no research on the personality traits of the translators was conducted. Undoubtedly this is a limitation of the study. Still, the homogeneity of the participants in terms of language learning and translation experience suggested that their translations can reveal common features of Hungarian emotion expression tendencies.

emotions characterizes cultures highly valuing interdependence. The minor difference of Hungarian and English might imply some closeness in the habits of emotion expression in Western cultures. This result about Japanese and Hungarian cultures is in line with the findings of Semin et al. (2002) examining Turkish, Hindustani-Surinamese and Dutch participants.

It is important to emphasize that the little sample of the present study does not allow us to draw any extensive conclusion. The fact that our results correlate with those of previous studies on the regulation and linguistic representation of emotions reassures that Japanese culture can and has to be placed in its context in cultural, linguistic and psychological sense as well. It is especially an important task in the light of the diverse results on the phenomena of individualism-collectivism and emotions (Kashima et al., 1995; Matsumoto, 1999). Cultural and linguistic habits may differ from people to people, yet in our view it is mostly a question of the degree a culture represents the universal regularities.

7. Conclusion

The present paper was devoted to two aspects of emotion expression as they appear in the language and culture of Japan compared to Hungary. The quantitative analysis of translations implies that emotional display rules of the Japanese culture are reflected in linguistic emotion expression too, resulting in a reduced use of explicit emotion terms. At the same time, lower level linguistic abstraction was found in Japanese language, suggesting a possibility of a difference in the function of emotion expressions in social interactions.

The way implicit emotion expression operates is one of the topics waiting for further research. Also, the question of emotional associations may be closely related to that. Another field of interest in connection with emotions is their role in constituting the unity during narration.

References

Bigazzi, S. & Nencini, A. (2008). How evaluations construct identities: the psycholinguistic model of evaluation. In O. Vincze & S. Bigazzi (Eds.), *Élmény, tötrénet – a történetek élménye* (pp. 91-105). Budapest: Új Mandátum Könyvkiadó.

Bodor, P. (2004). Az érzelmi fejlődés diszkurzív megközelítése: az érzés (feel). In M. Győri (Ed.), *Az emberi megismerés kibontakozása – Társas kognició, emlékezet, nyelv* (pp. 17-40). Budapest: Gondolat.

Doi, T. (1981). *The anatomy of dependence*. Tokyo: Kodansha International.

Ekman, P. (1972). Universals and cultural differences in facial expresions of emotion. In. J. Cole, (Ed.), *Nebraska symposium on motivation* (pp. 207-283). Lincoln, NE: University of Nebraska Press. (idézi Oatley és Jenkins, 2001)

Kashima, Y., Yamaguchi, S., Kim, U., Choi, S., Gelfand, M., & Yuki, M. (1995). Culture, Gender, and Self: A Perspective From Individualism-Collectivism Research. *Journal of Personality and Social Psychology, 69*, 925-937.

Kitayama, S., & Mesquita, B., & Karasawa, M. (2006). Cultural Affordances and Emotional Experience: Socially Engaging and Disengaging Emotions in Japan and the United States. *Journal of Personality and Social Psychology, 91*, 890-903.

Kiyomi, S. (2006). Kanjōhyōgen no nichiēhikaku. *Surugadaidaigakuronsō, 32*, 91-114.

Maass, A. (1999). Linguistic intergroup bias: Stereotype perpetuation through language. In M. P. Zanna, (Ed.), *Advances in Experimental Social Psychology,* (Vol. 31, pp. 79- 121). New York: Academic Press.

Maass, A., Karasawa, M., Politi, F. & Suga, S. (2006). Do Verbs and Adjectives Play Different Roles in Different Cultures? A Cross-Linguistic Analysis of Person Representation. *Journal of Personality and Social Psychology, 90*, 734-750.

Maass, A., Salvi, D., Arcuri, L., & Semin, G. R. (1989). Language use in intergroup contexts: The linguistic intergroup bias. *Journal of Personality and Social Psychology, 57*, 981-993.

Maeda, T. (1993). Nihongo no kanjō wo arawasu kotoba. *Nihongogaku, 12*, 4-13.

Markus, H. R., & Kitayama, S. (1991). Culture and the Self: Implications for Cognition, *Emotion, and Motivation. Psychological Review, 98*, 224-253.

Matsumoto, D. (1999). Culture and Self: An empirical assessment of Markus and Kitayama's theory of independent and interdependent self-construals. *Asian Journal of Social Psychology, 2*, 289-310.

Matsumoto, D., Yoo, S. H., Nakagawa, S., ... 37 Members of the Multinational Study of Cultural Display Rules. (2008). Culture, Emotion Regulation, and Adjustment. *Journal of Personality and Social Psychology, 94*, 925-937.

Miyamoto, Y., & Ma, X. (2011). Dampening or Savoring Positive Emotions: A Dialectical Cultural Script Guides Emotion Regulation. *Emotion, 11*, 1346-1357.

Morschbach, H., & Tyler, W. J. (1986). A Japanese emotion: Amae. In R. Harré (Ed.), *The social construction of emotions* (pp. 289-307). Oxford: Blackwell.

Nomura, M. (2003). Gendaigo no tekusuto ni okeru kanjōhyōgen. *Nihongogaku, 22,* 36-44.

Ohso, M. (2001). Kanjō wo arawasu dōshi-keiyōshi ni kansuru ichikōsatsu. *Gengobunkaronshū, 22*, 21-30.

Romney, A. K., Moore, C. C., & Rusch, C. (1997). Cultural universals: Measuring the semantic structure of emotion terms in English and Japanese. *Anthropology, 94*, 5489-5494.

Safdar, S., Friedlmeier, W., Matsumoto, D., Yoo, S. H., Kwantes, C. T., Kakai, H., & Shigemasu, E. (2009). Variations of Emotional Display Rules Within and Across Cultures: A Comparison Between Canada, USA, and Japan. *Canadian Journal of Behavioural Science, 41*, 1-10.

Semin, G. R., & Fiedler, K. (1988). The cognitive functions of linguistic categories in describing persons: Social cognition and language. *Journal of Personality and Social Psychology, 54*, 558-568.

Semin, G. R., & Fiedler, K. (1989). Relocating attributional phenomena within a language-cognition interface: The case of actors' and observers' perspectives. *European Journal of Social Psychology, 19*, 491-508.

Semin, G. R., & Fiedler, K. (1991). The linguistic category model: Its bases, applications, and range. In W. Stroebe & M. Hewstone, (Eds.), *European Review of Social Psychology* (Vol. 2, pp. 1-30). Chicester, UK: Wiley.

Semin, G. R., Görts, C. A., Nandram, S., & Semin-Goosens A. (2002). Cultural perspectives on the linguistic representation of emotion and emotion events. *Cognition and Emotion, 16* (1), 11-28.

Semin, G. R. (2004). The Self-in-Talk: Towards an Analysis of Interpersonal Language and its Use. In J. Jost, M. R. Banaji & D. A. Prentice (Eds.), *Perspectivism in Social Psychology: The Yin and Yang of Scientific Progress* (pp. 143-159). Washington, DC: APA Press.

Tanabe, Y., & Oka, T. (2001). Linguistic intergroup bias in Japan. *Japanese Psychological Research, 43,* 104-111.

Uchida, Y., Townsend, S. S. M., Markus, R. H., & Bergsieker, H. B. (2009). Emotions as Within or Between People? Cultural Variation in Lay Theories of Emotion Expression and Inference. *Personality and Social Psychology Bulletin, 35,* 1427-1439.

Wierzbicka, A. (1996). Japanese Cultural Scripts: Cultureal Psychology and "Cultural Grammar". *Ethos, 24,* 527-555.

Yoshimoto, B. (1991). *Kitchin.* Tokyo: Fukutake Bunko.

Yoshimoto, B. (2001). *Kitchen.* New York: Faber and Faber Limited.

A CROSS-CULTURAL STUDY ON HEDGING DEVICES IN KURDISH CONVERSATION

Biook BEHNAM
Department of English Language, Tabriz Branch,
Islamic Azad University, Tabriz, Iran
behnam_biook@yahoo.com

Salam KHALILIAQDAM
Department of English Language, Boukan Branch,
Islamic Azad University, Boukan, Iran
skhaliliaqdam@yahoo.com

Abstract

Hedges are words whose job is to make things fuzzier or less fuzzy. Truth and falsity are a matter of degree, and hedges make natural language sentences more/less true or more/less false. The purpose of the study is to investigate hedging devices in Kurdish spoken language. The aim is to know how hedging devices are used in Kurdish spoken discourse. Also the researchers are willing to know whether Kurdish speakers use hedging devices to indicate a lack of complete commitment to the truth of the proposition, and a desire not to express the commitment categorically, or to lessen the impact of an utterance. The data needed for the study was collected through observation, tape recording, and interviews. The dialogues of 35 people were recorded by the researcher as well as the researcher has interviewed with 21 people from different social classes.15 classes and meetings which Kurdish language was the means of communication were observed. The research showed that hedging as a mitigating device is extensively employed in different conversations. The study shows that hedging devices have the same roles in Kurdish as they have in English. They are used to reduce the certainty and sureness of the utterances. It indicates that some pragmatic devices modify the epistemic strength of the statement in Kurdish language just as they do in English and Arabic.

Keywords

hedging devices; Kurdish; spoken discourse

Izvleček

Omejevalci so besede, katerih naloga je narediti bolj ali manj nerazločno. O resnici oziroma nepravilnosti lahko govorimo le stopenjsko, in omejevalci naredijo naravni jezik bolj ali manj resničen ter bolj ali manj nepravilen. Namen tega članka je raziskati mehanizme omejevalcev v pogovorni kurdijščini in ugotoviti, kako so uporabljeni v kurdijskem pogovornem diskurzu. Poleg tega avtorja ugotavljata, ali uporaba omejevalcev nakazuje na pomanjkanje popolne

predanosti resnici izrečenega in željo po nekategoričnem izražanju obveze ali na zavestno zmanjšanje vtisa izrečenega. Podatki za raziskavo so bili pridobljeni z opazovanjem (15 predavanj v kurdijščini), snemanjem (35 ljudi) in intervjuji (21 ljudi iz različnih socialnih razredov). Splošni rezultati kažejo, da je uporaba omejevalcev kot blažilcev izrečenega v veliki meri uporabljena v vseh vrstah pogovorov in da ni opaznih razlik med uporabo omejevalcev v kurdijščini in angleščini. V obeh jezikih se uporabljajo kot mehanizmi, ki zmanjšujejo gotovost izrečenega, kar nakazuje, da v kurdijščini pragmatični mehanizmi modificirajo epistemično moč stavka, tako kot v angleščini in arabščini.

Ključne besede

mehanizmi omejevalcev; kurdijščina; pogovorni diskurz

1. Introduction

A hedge is a mitigating device used to lessen the impact of an utterance. Typically, they are adjectives or adverbs, but can also consist of clauses. It could be regarded as a form of euphemism. Hyland (1996) illustrates that hedging devices are used to indicate a lack of complete commitment to the truth of the proposition, and a desire not to express the commitment categorically. Hedges may intentionally or unintentionally be employed in both spoken and written language since they are crucially important in communication. Hedges help speakers and writers communicate more precisely the degree of accuracy and truth in assessments. Linguists almost unanimously define hedges as a means to tone down utterances and statements, to reduce the riskiness of what one says, to mitigate what might otherwise seem too forceful, to be polite or show deference to strangers or superiors etc. Lakoff, the pioneer in this field, defined the items like *largely, rather* etc as the words which "make things fuzzier or less fuzzy" (Lakoff, 1972, p. 195). Non native English writers are expected to manifest in their research articles not only grammatical competence in the English language but also sociolinguistic competence in the form of expressing "politeness", showing openness to criticisms, and confidently stating uncertainty regarding findings or claims. Markkanen and Schröder (2006) distinguish two types of hedges, or two reasons for hedging: one type of hedges deals with certain linguistic items that affect the truth-conditions of propositions; the other type reflects the degree of the speaker's commitment to the truth-value of the whole proposition. Vold (2006) also writes about *real hedges,* which serve to give an accurate picture of the level of certainty, and *strategic hedges,* which may fulfill a variety of functions.

In short, hedges are used to express the writer's attitude to both proposition and readers. This study aims at investigating hedging devices in Kurdish language. A large number of conversations of Kurdish speakers have been recorded and transcribed in order to be analyzed what hedging devices Kurdish speakers use in their speech to convey their message.

2. Review of related Literature

Hedging is the expression of tentativeness and possibility. It is therefore central to academic/scientific writing, where statements are rarely made without subjective assessment of their reliability and where claims need to be presented with caution and precision. Science indeed is skepticism doubt, refutation, speculation, formulation of hypothesis, criticism. As a consequence, the expression of doubt and possibility is central to the negotiation of claims, and what counts as effective persuasion is influenced by the fact that evidence, observations, data, and flashes of insight must be shaped with due regard for the nature of reality and their acceptability to an audience.

In medical writing, hedges play a critical role in gaining ratification for claims for a powerful peer group by allowing writers to present statements with appropriate accuracy caution, and humility, expressing possibility rather than certainty and prudence rather than overconfidence. In a context where the accreditation of knowledge depends on the consensus of the research community and the need to evaluate evidence, comment on its reliability, and avoid potentially hostile responses (the "boomerang effect"), expressions such as "may", "might", "could", "possible", and "likely" can contribute to gaining the acceptance of research claims.

While research on hedging and hedges has progressed and expanded enormously over the past four decades, it is still apparent that the semantic category of hedges has not been precisely defined yet. Perhaps the lack of such a category is attributed to the complexity of the meanings of the hedging devices, a fact that has presented a serious challenge for researchers.

Apart from the semantic category of hedges, it seems that researchers have a broad consensus on what hedging is. Lakoff (1972) associates hedges with unclarity or fuzziness: "for me some of the most interesting questions are raised by the study of words whose job is to make things more or less fuzzy." (p. 195). It has been observed that the term hedging which was first used to refer to fuzziness has been widened to cover a number of interrelated concepts, namely indetermination, vagueness, indirectness and approximation (Zuck & Zuck, 1986; Brown & Levinson, 1987; Hyland, 1998; Btoosh, 1999; Btoosh, 2004; Varttala, 2001; Vass, 2004; Chavez; 2004; Ayodobo, 2007; Vazques & Giner, 2008, Donesch-Jezo, 2010).

In a more comprehensive account of the term, Bruce (2010) associates hedging with all means leading lack of full commitment (p. 201).

Hedging is a rhetorical strategy. By including a particular term, choosing a particular structure, or imposing a specific prosodic form on the utterance, the speaker signals a lack of a full commitment either to the full category membership of a term or expression in the utterance (content mitigation), or to the intended illocutionary force of the utterance (force mitigation).

The impact of hedging devices in the discourse is measured by their overall effect on meaning or the message of the text oral/written. Hyland (1996) illustrates that

hedging devices are used to indicate a lack of complete commitment to the truth of the proposition, and a desire not to express the commitment categorically. The same function is found in economics discourse. Pindi & Bloor (1987) argue that "economics forecasters are shown to have three ways of modifying their commitments to a prediction: by hedging, using such as modal verbs as '*may*' or other lexical items such as possibility and by specifying conditions." (p. 55).

Hedging may also stem from the inner conflict between intention and desire: "being indirect is a mechanism for dealing with conflicting intentions and desires. The general form of the conflict is that the speaker wants to convey *X* for some reason and he does not want to convey *X* for other reasons. By being indirect he can convey *X* in one sense but not in another." (Pyle, 1975).

Like English, Arabic does employ lexical, syntactic (conditionals or passive) as well as strategic hedges. However, one of the most common structural hedging devices employed in Arabic discourse is the *conditional sentences*. Safi (1988) argues that: "probability is one of the most difficult issues associated with conditionality. In English the use of the different tenses of verbs and modals usually stand for probability whereas in Arabic it is possible for the conditional particles and different tenses of verb to stand for probability".

A text, of course, is said to have hedging by its having any of the different means that express hedging directly or indirectly. Channel (1994) argues that "one of the most useful and enduring insights to come out of the recent study of language use is that speakers and writers tailor their language to make it suitable to the situation (when, where and why?) and the linguistic context (is it gossipy chat, an interview, a story in a popular newspaper?)". Hedging use, as the literature shows, is affected by gender. Lakoff (1972) asserts that in order to show their femininity, women tend to adopt an unassertive style of communication.

Language scholars have claimed that gender differences in communication mirror and reproduce broader political inequalities between the sexes (Fishman, 1978; Rakow, 1986; Thorne & Henley, 1975; Thorne, Kramerae, & Henley, 1983; Uchida, 1992). The use of hedging devices is one area of inquiry in which this argument has found support. Research on gender and hedging has been strongly influenced by Robin Lakoff's book, *Language and Woman's Place*. Lakoff (1975), an American linguist, argued that women's speech lacks authority because, in order to become "feminine", women must learn to adopt an unassertive style of communication. That is, they must learn to denude their statements of declarative force. Lakoff coined the phrase "women's language" to refer to a group of linguistic devices that serve this function, including hesitations, intensive adverbs, empty adjectives, tag questions, and compound requests. Hedges form part of this group.

The term *hedge* refers to a class of devices that supposedly soften utterances by signaling imprecision and noncommitment. Examples include the pragmatic particles *about, sort of,* and *you know* and the modal terms *possibly* and *perhaps.* Since Lakoff's

pioneering work, hedges have featured prominently in research on gender and communication. Based upon data about their distribution in the speech of men and women, researchers have made bold theoretical claims—most commonly that women's language is indecisive and deficient.

Fraser (1975) introduced the term HEDGED PERFORMATIVE, where certain performative verbs such as apologize, promise, and request when preceded by specific modals such as can, must, and should, as in

a) I should apologize for running over your cat.
b) I can promise that I will never again smoke grass.
c) I must request that you sit down.

result in an attenuated illocutionary force of the speech act designated by the verb. In these examples, the modals were considered as hedges. Example (a) is still an apology, just one less strong than if should were not present. Prince et al. (1982) made a clear distinction between two types of hedging, one type that involves the propositional content and affects the truth condition of the proposition conveyed (propositional hedging), and a second type that involves the relationship between the propositional content and the speaker and serves as an index of the commitment of the speaker to the truth of the propositional content conveyed (speech act hedging).

These authors divided up the hedging world into two classes. The first, APPROXIMATORS, operate on the propositional content proper and contribute to the interpretation by indicating some markedness, that is, non-prototype, with respect to class membership of a particular item.

There are two subclasses: ADAPTORS (acknowledged to be what Lakoff called hedges), relate to class membership; for example, somewhat, sort of, almost describable as, some, a little bit, etc.,

a) He also has a somewhat low interior larynx.
b) She noticed that he was a little bit blue

and ROUNDERS, convey a range, where the term is typical, for example, about, approximately, something around, etc.

a) His weight was approximately 3.2 kilograms.
b) The baby's blood pressure was something between forty and fifty.

Both sub-classes occur when the speaker is attempting to correlate an actual situation with some prototypical, goal-relevant situation, where the hedging indicates that actual situation is close to but not exactly the expression modified. SHIELDS, their second class, change the relationship between propositional content and the speaker by implicating a level of uncertainty with respect to speaker's commitment.

Here, again, there are two subclasses. PLAUSIBILITY SHIELDS are expressions that relate doubt, such as I think, I take it, probably, as far as I can tell, right now, I have to believe, I don't see that, etc., illustrated by the following.

a) I think we can just slow him down to a little over maintenance.
b) As far as I can tell, you don't have anything to lose by taking that path.

The second subclass, ATTRIBUTION SHIELDS are expressions such as according to her estimates, presumably, at least to X's knowledge, etc., which attribute the responsibility of the message to someone other than the speaker, often via plausible reasoning.

a) He was not very ill, according to her estimates.
b) There was no reason to worry, as far as anyone knew.

The authors also point out that one usually does not impose belief on another when the speaker believes that the proposition at issue is false. For example, in the following sentence,

a) According to Dr. Jenkins, we should take out the shunt before we move him.

the speaker is typically committed to the truth of the statement.

Hubler (1983) made a similar two-way distinction of hedging, between what he called UNDERSTATEMENTS and HEDGES, although he uses UNDERSTATEMENT as a cover term for both. Understatement means that "the emotional negatability (of sentences) is restricted through the indetermination of the phrastic," that is, they concern the propositional content of the sentence. It is a bit cold in here, contains an understatement. HEDGING "is restricted through the indetermination of the neustic," that is, it concerns the speaker's attitude to the hearer regarding the proposition, the claim to validity of the proposition the speaker makes. It is cold in Alaska, I suppose, contains a hedge. Hubler's division resembles that of Prince et al. (1982), whose APPROXIMATORS correspond to Hubler's UNDERSTATEMENTS and whose SHIELDS correspond to his HEDGES.

In the 1980s, there was considerable effort to sub-classify the class of hedges, based on certain class membership criteria. The following list reflects the array of proposals involving: attenuating hedges. adaptors, agent avoiders, approximators, attenuators, attribution shields, bushes, committers, compromisers, consultative devices, deintensifiers, diffusers, diminishers, down-toners forewarners, indicator of degrees of reliability, minimizers, mitigators, plausibility shields, play-downs, politeness markers, scope-staters, understaters, validity markers, vocal hesitators, weakeners, etc.

Since the 1980s, there has also been an emphasis on the properties of individual hedges, for example, see Kay (1984), and Aijmer (1984). In addition, there has been considerable interest in exploring the use of hedging within different genres of language use, for example, the research article, mathematics talk, politician talk, negotiation talk, and the speech of language learners, to name but a few of the areas. I do not explore these areas (cf. Schröder & Zimmer, 1997).

There is general agreement today that HEDGING is a rhetorical strategy, by which a speaker, using a linguistic device, can signal a lack of commitment to either the full semantic membership of an expression (PROPOSITIONAL HEDGING),

a) He's really like a geek.
b) The pool has sort of a L-shaped design.
c) Peter's house is almost 100 feet wide.

or the full commitment to the force of the speech act being conveyed (SPEECH ACT HEDGING),

a) Come over here, can you?
b) I guess I should leave now.
c) That type of comment isn't made around here. [Agent less passive]
d) Perhaps you would sit down a minute.

The notion of REINFORCEMENT, initially considered a part of hedging, has pretty much been laid aside. Thus, sentences such as

a) I certainly do insist that you sit down.
b) He is extremely tall.

are not generally viewed today as instances of hedging but rather of reinforcement. Aijmer (1984) believes that the reason for this narrowing of the concept stems from the fact that the sense of hedging on the positive side of a concept (be it to involve a proposition or a speech act) seems counterintuitive:

> Hedging is simply not a symmetrical notion, and it does not connote reinforcement. Linguistic hedges include linguistic devices, both morphological and syntactic forms used in the process of hedging. These include: adverbs, adjectives, impersonal pronouns, concessive conjunctions, indirect speech acts, introductory phrases, modal adverbs, modal adjectives, hedged performatives, modal nouns, modal verbs, epistemic verbs, negation, tag questions, agentless passives, parenthetic constructions, if clauses, progressive forms, tentative inference, hypothetical past, metalinguistic comments, etc.(p.128)

3. Methodology

The purpose of the study is to investigate hedging devices in Kurdish spoken language. The researcher would like to know how hedging devices are used in Kurdish spoken discourse. Also the researchers are willing to know whether Kurdish speakers use hedging devices to indicate a lack of complete commitment to the truth of the proposition, and a desire not to express the commitment categorically, or to lessen the impact of an utterance. As far as we know hedges help speakers and writers communicate more precisely the degree of accuracy and truth in assessments. Linguists almost unanimously define hedges as a means to tone down utterances and statements, to reduce the riskiness of what one says, to mitigate what might otherwise seem too forceful, to be polite or show deference to strangers or superiors etc. In this survey, the

researcher investigates if hedging devices play the same roles as they do in other languages such as English and Arabic in the following examples:

(1) ***rubbama*** naḫnu nantaḍer mustaqbal 'aswa
"***Perhaps***, we are awaiting a worse future."

(2) ***ana a'taqid*** 'anna hadhihi ḫaraban leisat 'adelah
"***I believe*** that this is not a fair war."

(3) ***qad tastamer*** alḫarb limodat osbo'ayein
"The war ***may*** last for two weeks."

The data needed for the study was collected through observation, tape recording, and interviews. The dialogues of 35 people were recorded by the researcher as well as the researcher has interviewed with 21 people from different social classes.15 classes and meetings which Kurdish language was the means of communication were observed.

4. Data analysis and discussion

The following dialogue is between mother and her seven-year-old daughter:

(4) Daughter crying and wanting to go to bazaar with her mother, but her mother doesn't want to.

Mother: "shaa, shaa, **daley** shetoo shta"
"look, look, (addressing her husband) **as if** she is mad."

Father: "**agar qawlt** pe nadabaaya, away nadakrd."
"**If** you didn't **promise** her, she couldn't behave like that."

Her aunt: "**hamisha** aawaaya?"
"Does she **always** do so?"

Her father: "naawallaa **jaar jaar** waadakaa awish **pemvaabe** dayki rqi hastaandbet."
"No, **sometimes** she does so, **perhaps** her mother made her angry."

Her mother: "nawalla **hamisha** waya. **Nazanm** bowa daka wala hichm pe nakotwa."
"No, she **always** gives excuses. **I don't know** why she does. I swear God I have not told her anything."

Daughter to her father: "nawalla lemi daa baba jian. Awal **qawli** pemda, alan dale naatbam."
"No, she hit me, dear papa dear papa. At first, she **promised** me to take me, now she doesn't want to."

Her mother: "kotm **agar** Nasrin khaanm Baran benet amnish ato dabam."
"I told **if** Nasrin Khanm drove Baran with her, I would take you with me."

According to Lakoff (1972), hedges are "words whose job is to make things fuzzier or less fuzzy". He states that truth and falsity are a matter of degree, and hedges make natural language sentences more/less true or more/less false. Therefore, the

examples mentioned above include hedging devices in kurdish language. By investigating the conversation , it is understood that Kurdish speakers use ***agar*** (if: pseudo-conditionals), ***qawl*** (promise: a performative verb), ***pemwaabe*** (I think: a plausibility shield), and ***naazaanm*** (I don't know: a tentativizer) in order to make utterances be neither absolutely true, nor absolutely false, but rather true/false to a certain extent, or true in certain respects and false in other respects.

Considering the following example:

(5) ...**hamisha** aawaaya…
"Does she ***always*** do so?"

As far as we know the term "***hamisha***" (always) is not a hedging device; it plays as boosters in order to intensify what is being said, whereas the term "***jaar jaar***" which means "sometimes" in sentence (6) has been used to reduce the degree of certainty. See the following example :

(6) …naawallaa **jaar_jaar** waadakaa…
"No, ***sometimes*** she does so."

We can say that in Kurdish language, approximators are used to operate on the propositional content proper and contribute to the interpretation by indicating some markedness.

The following conversation occurred among a history teacher (T) and his students (S). They were talking about the future of Iran's economy:

(7) S1: "aga zer **awaa** chowata sare va dolaar **wahshatnak** gran bowa, **bley** khalk raeii bdan?"
"Proffessor!! The price of gold has raised **a lot** and dollars has become **terribly** expensive. Do you **think** people take part in election?"

T: "mn **mo?taqidm** ka khalk waqean narazin wa ba ehtimali qawi kam raei dadan."
"I **believe** that people are really unsatisfied and **most likely a few** will vote."

S2: "amma agaa amn **ehsas dakam** amjaar khalk **zortr** raeii dadan."
"But My Proffessor, I **feel more** people will vote this time."

S3: "wali mn **bawar nakam**. Agarish raei bdam, tedadyan kama."
"But **I don't think** so. **If** they vote, **a small number** will do."

S1: "aga **bley** Amrikaa hamla bka ba Eran?"
"Proffessor, **do you think** US attack Iran?"

T: "**bochuuni** mn awaya ka **zorbay** awaanay ka raeii dadan ya faqirn ya sonnati fkr dakanawa."
"I **guess most** of the people who will vote either are poor or think traditionally."

S2: "ay **agar** khalk **kam** raei bdan **ehtemali haya** Amrika hamla bka?"
"**If a small number** of people vote, **may** US attack Iran?"

T: "naazaanm awa maluum nia ba zor sht bastagii haya."
"**I don't know**. It is **not clea**r. It ups to various factors."

The notion of REINFORCEMENT, initially considered a part of hedging, has pretty much been laid aside. Thus, Taking into account the sentences such as:

(8) "agha!! zer **awaa** chowata sare va dolaar **wahshatnak** gran bowa…."
"Proffessor!! The price of gold has raised **a lot** and dollars has become **terribly** expensive…"

(9) "…..khalk **waqean** narazin wa **ba ehtimali qawi kam** raei dadan…"
"… people are **really** unsatisfied and **most likely a few** will vote…"

These sentences are not generally viewed today as instances of hedging but rather of reinforcement. The words such as ***awaa*** (a lot), ***wahshatnak*** (terribly), ***waqean*** (really), and ***ehtimali qawi*** (most likely) are considered as reinforcement. The way Kurdish speakers reinforce something is quite similar to the way English speakers do. However, the sentence (10) includes an adaptor "***kama***" which means "***a small number***". A Kurdish speaker tries to reduce the degree of the subject by using the adaptors just as an English speaker

(10) **Agarish** raei bdam, tedadyan **kama**
If they vote, **a small number** will do

The next dialogue happened among an English teacher (T) and her students (S). They argued on the word which the teacher doesn't know its meaning:

(11) S1: "agha aw kalimaya laweda yaani chi? Fela?"
"Proffessor, what does that word here mean? Is it a verb?"

T: "bale wa manaakay…. **Hin**…**aw shtaa**…**korra to ble**…..**la sar zmanma** … dwaya pet dalem."
"yes, and its meaning?..... **hin**….**that thing**…….boy? **you say**….**it's on tip of my tongue**…..I **will** tell you later."

S1: "**datwaanin** bleyn ba manaay waday ghazaeiya."
"we **can** say it is a meal."

T: "**dabe** tmashay farhangi logat bkam……. Teyda niya ajiba**….. fkr kam** englisi niya…"
"I **should** look it up…it's not here…. **I think** it's not an English word…"

S2: "na englisi niya **dabe**.. italyayi bet….**kamek** italyayii dazanm.. ahha wabirm hatawa..la jegayekda ditoma.. awa **pemwabe** yani chorti dway nahar."
"No, it's not an English word. It **must** be Italian… I know Italian **alittle**…..ahaa I remembered…..I have seen it **somewhere….. I guess** it means taking a short sleep after lunch."

S3: "**bley** dorost bet?"
"**Do you think** it is right?"

T: "**ba tawajoh ba** wshakani dawrobari **datwane** aw manaaya bdat."
"**According to** surrounding words. That meaning **can** be right."

What is important to me is that vocalizations play an outstanding role in hedging a subject or an idea. Holmes (1999) attributes to hedges vocal hesitations (*um, er*) and

such linguistic forms as *you know, I think, sort of* which she labels as pragmatic particles. The terms such as "***Hin, aw shtaa***", (***hin****….****that thing)*** and "***korra to ble***" (boy? ***you say)*** are considered as hedged vocal hesitations. These vocal hesitations like plausibility and attributed shields impose the commitment on the proposition. Nevertheless, thinking of the following illustration,

(12) "….na englisi niya **…dabe**.. italyayi bet….**kamek** italyayii dazanm…"
"No, it's not an English word. It **must** be Italian… I know Italian **a little**."

we can say that the word ***kamek*** (a little) is a sort of hedging device. Though it isn't a hedging device in this utterance, it is considered as reinforcement. Since the student, by using such reinforcement as ***kamek*** (a little), wants to persuade the teacher that what he says is true. Therefore, the students try to foreground to the extent of Italian he knows by using the word ***kamek*** (a little). The other case which is worth maneuvering is the term ***ba tawajoh ba*** (according to) which is a sort of attributed shield. As far as we know attributed shield in Kurdish has the same function as that in English.

The following dialogue went on among a mother (M) and her 5-year-old son (S) watching a documentary movie:

(13) S: "maman awa chiya awa chiya?"
"Mom! What's this? What's this?" (pointing to the animal he saw on TV)

M: "**naw?e** balandaya"
"**A kind of** a bird"

S: "balanda?"
"A bird? "

M: "shteka ka haldafre dandoki haya ….. bali haya …"
"A thing which fly, has a peak, and has wings."

S: "newi chiyaa?"
"What do we call it?"

M: "pey dalen boqla…"
"It is called turkey."

S: "ay kotr chiya?"
"And what's a pigeon?"

M: "awish balandaya"
"That's a bird too."

S: "farqiyan chiya?"
"What's the difference?"

M: "**are chozanm** rola…… **taqriban wako** yakwan amaa **kamek** boqla gawratra…."
"**I don't know** , son. …they are **almost the same** but a turkey is **a bit** larger."

The hedge *sort of* is a device provided by the linguistic system itself to guide the hearer in the sort of pragmatic process. According to the above example, not "***naw?e***" (sort of) is the matter of semantics, but it is a matter of pragmatics. For instance the sentence:

(14) "**….naw?e** balandaya……"
"**A kind of** a bird"

(15) "….balandaya….."
"It is a bird."

The sentence (14) is different from the following sentence (15) since there is not as much certainty in sentence (14) as there is in sentence (15).

Since there is not as much certainty in sentence (14) as there is in sentence (15). What has reduced the certainty of the sentence (a) is the use of the hedge (***naw?e***) ,meaning sort of. A child is aware of the fact that a pigeon is a bird; nevertheless, he hesitated to accept that turkey is a bird, as well. But hesitancy is reduced when his mother uses the hedge (***naw?e***) along with the word "turkey". The following example by Lakoff (1972) affirms our claim:

(a) A robin is a bird
(b) A penguin is a bird
(c) A penguin is sort of a bird

While (a) is uncontroversial, some people hesitate to affirm (b) since a penguin does not fly and is thus not felt to be a good example of a bird; in Lakoff's terms, it is aperipheral member of the bird category. Hesitancy is greatly reduced when *sort of* is employed, as in (c). It seems that the effect of *sort of* is to loosen or broaden the concept encoded by *bird*, so that it more comfortably encompasses creatures which do not have all the stereotypical properties of birds. In other words, the hedge *sort of* is a device provided by the linguistic system itself to guide the hearer in the sort of pragmatic process he is to carry out in order to arrive at the intended interpretation of a particular word. The pragmatic use of epistemic devices is further complicated by the fact that they not only convey the writer's confidence in the truth of referential information, but also help contribute to a relationship with the reader. This affective dimension involves the need for cooperation and deference. In addition, devices of imprecision such as *about* and *almost* can also modify the epistemic strength of statements (Dubois, 1987) while expressions used to manipulate definiteness, such as *frequently* and *usually*, also contribute to the "scales of probability and usuality to which the term modality strictly belongs" (Halliday, 1985, p. 86). Moreover, while the expression of writer "commitment" is mainly a lexical phenomenon; conditional clauses, questions, contrast markers, and tense can also be used to convey epistemic meanings (Perkins, 1983).

However, consider the following example:

(16) S: "ay kotr chiya?"
"And what's a pigeon?"

M: "awish balandaya"
"That's a bird too."

S: "farqiyan chiya?"
"What's the difference?"

M: "**are chozanm** rola…… **taqriban wako** yakwan amaa **kamek** boqla gawratra…."
"**I don't know** , SON. …they are **almost** the same but a turkey is a bit larger."

"***taqriban***" (almost) is an epistemic device in Kurdish which indicates a sort of imprecision.

Table 1: Types of hedging used in Kurdish along with their English equivalence

Hedging device in Kurdish	**Its equivalence in English**	**Type of hedging device**
daley	As if	Pseudo-conditional
agar	if	Pseudo-conditional
qawl	promise	Performative verb
hamisha	always	Reinforcement
Jaar jaar	Sometimes	Approximant/downgraders
pemvaabe	I guess	Plausibility shield
Nazanm	I don't know	Tentativizaer
wahshatnak	Terribly	Downgrader
bley	Do you think	Plausibility shield
mo?taqidm	I believe	Plausibility shield
waqean	Really	Reinforcement
ehtimali qawi	To a large extent	Reinforcement
maluum nia	It is not clear	Plausibility shield
kam	A few	Approximant/adaptor
ehsas dakam	I feel	Plausibility shield
zortr	More	Reinforcement
bawar nakam	I don't believe	Tentativizer
Bochuuni mn	I guess	Plausibly shield
zorbay	Most	Reinforcement
Hin… …..	…..	Vocalization
aw shtaa…	…..	Vocalization
korra to ble	Boy you say	Vocalization
la sar zmanma	It is on tip of my tongue	Plausibility shield
ba tawajoh ba	According to	Attribution shield
fkr kam	I think	Plausibility shield
datwaanin	We can	Plausibility shield
naw?e	A sort of	Tentativizer
are chozanm	I don't know!!!!	Tentativizer
taqriban	almost	Approximant/adaptor
La jegayek taqriban	Somewhere almost	Approximant/adaptor Rounders

5. Conclusion

The research showed that hedging as a mitigating device is extensively employed in different conversations. An attempt has been made to identify, quantify and analyze different hedging devices employed in spoken discourse in Kurdish. The study has sought to examine how hedging devices are used in Kurdish and also investigate the types of hedging devices which are used in Kurdish conversations. The study shows that hedging devices have the same roles in Kurdish as they have in English. They are used to reduce the certainty and sureness of the utterances. It is typical that some pragmatic devices modify the epistemic strength of the statement. Analyzing the pragmatic use of some hedging devices, we found out that a hedging device which reduced the accuracy, truthfulness, and certainty of the statement in one context could play as reinforcement in other context. Hedging devices used in Kurdish indicate a lack of complete commitment to the truth of the proposition, and a desire not to express the commitment categorically, or to lessen the impact of an utterance. As a result people employ hedges as a means to tone down utterances and statements, to reduce the riskness of what one says, to mitigate what might otherwise seem too forceful, to be polite or show deference to strangers or superiors or someone else. It was understood that attributed shield (***ba tawajoh ba***: according to), plausibility shield (***ehtimalan***: probably), approximates (***kamek***: alittle), tentativizers (***naazanm***: I don't know), pseudo-conditional (***agar***: if), vocalization (*hin*: umm) and amazingly the term "***la sar zmaanma***" meaning "It's on tip of my tongue" were used as hedges in Kurdish.

References

Aijmer, K. (1984). Sort of and kind of in English conversation. *Studia Linguistica, 38(2),*18–128.

Austin, J. (1962). *How to do things with words*. London: Oxford University Press.

Ayodapo, J. O. (2007). Hedging: The pragmatics of politeness in English. *Lagos Papers in English Studies, 1*, 257-270.

Brown, P. & Levinson, S. (1987). *Politeness: Some universals in language usage*. Cambridge: Cambridge UP.

Bruce, F. (2010). Hedging in political discourse: The Bush 2007 press conferences. In Okulska, Urszula & Cap, Piotr (Eds.). *Perspectives on Politics and Discourse, 36*, 201-214. Amsterdam/ Philadelphia: JohnBenjamins Publishing Company

Btoosh, M. (1999). *Hedging in journalistic Arabic*. MA Dissertation. Yarmouk University, Jordan.

Btoosh, M. (2004). *Interlanguage lexicology of Arab students of English: A computer-learner corpus-based approach.* Doctoral dissertation. University of Texas at Arlington, USA.

Channell, J. (1994). *Vague Language*. Oxford: Oxford University Press.

Chavez, D. (2004). *The language of uncertainty in a new illness: Hedging and modality in the biomedical discourse of Sever Acute Respiratory Syndrome (SARS).* Doctoral dissertation. University of Mahidol.

Donesch-Jezo, E. (2010). Teaching academic discourse writing in ESP courses for medical students and professionals. *US-China Foreign Language, 8(1),* 32-39.

Dubois, B. L. (1987). Something in the order of around forty to forty four. Imprecise numerical expressions in biomedical slide talks. *Language and Society, 16,* 527-541.

Fishman, P. M. (I978). Interaction: the Work Women Do. In B. Thome, C. Kramarae & N. Henley (Eds.), *Language, Gender and Sociery* (pp. 89-101). Massachusetrs: Newbury House.

Fraser, B. (1996). Pragmatic markers. *Pragmatics 6(2),* 167–190.

Goffman, E. (1967). Interaction ritual: essays on face to face behavior. New York: Garden City.

Halliday, M.A.K. (1985). *An Introduction to Functional Grammar*. London: Edward Arnold.

Holmes, J. (1992). *An Introduction to Sociolinguistics*. London: Longman.

Holmes, J. (1995). *Women, Men and Politeness*. New York: Longman.

Holmes J. (1999). *Women, Men and Politeness.* Longman Group, UK Limited.

Hubler, A. (1983). *Understatements and hedges in English*. Amsterdam: John Benjamins.

Hyland, K. (2000). Disciplinary discourses: social interactions in academic writing. London: Longman, 93.

Hyland, K. (1998). *Hedging in scientific research articles.* Amsterdam: John Benjamins.

Hyland, K. (1996). Writing without Conviction? Hedging in Science Research Articles. *Applied Linguistics, no. 17.4.*

Kay, P. (1984). *The kind of/sort of construction*. Proceedings of the 10th Annual Meeting of the Berkeley Linguistic Society.

Lakoff, G. (1972). Hedges: a study in meaning criteria and the logic of fuzzy concepts. *Chicago Linguistic Society, 8,* 183–228.

Lakoff, R. (1975). *Language and Women's Place.* New York: Harper & Row.

Markkanen, R. & Schröder, H. (2006). *Hedging: a Challenge for Pragmatics and Discourse Analysis*. Retrived February 6, 2012 from http://www.sw2.euv-frankfurt-o.de/Publikationen/hedging/markkane/markkane.html.

Perkins, M. (1983). *Modal express low in English.* London: Frances Pinter.

Pindi, M., & Bloor, T. (1987). Playing safe with predictions: Hedging, attribution and conditions in economic forecasting. In T. Bloor & J. Norrish (Eds.), *Written Language* (CILT, 55-69). *Politics and Discourse* (36, pp. 201-214). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Prince, E., Frader, J., & Bosk, C. (1982). On hedging in physician-physician discourse. In R. J. Di Pietro (Ed.), *Linguistics and the professions* (pp. 83–97). Proceedings of the second annual Delaware symposium on language studies. Norwood, NJ: Ablex.

Pyle, C. (1975). *The functions of indirectness*. Paper Read at N-WAVE IV Georgetown University. Quirk, R. (1978). Language and Tabu. New Society.

Rakow, L. F. (1986). Rethinking gender research in communication. *Journal of Communication, 36,* 11-26.

Safi, M. (1988). *Conditional sentences in English and Arabic: A contrast- error analysis for translation purposes*. (Unpublished M.A. Thesis).

Sandell, R. (1977). *Linguistics style and persuasion*. London: Academic Press.

Schröder, H., & Zimmer, D. (1997). Hedging research in pragmatics; A bibliographical research guide to hedging. In R. Markkanen & H. Schröder (Eds.), *Hedging and discourse. Approaches to the analysis of a pragmatic phenomenon in academic texts* (pp. 249–271). Berlin: Walter de Gruyter.

Thome, B., & Henley, N. (1975). Difference and dominance: An overview of language, gender and society. In B. Thorne & N. Henley (Eds.), *Language and gender: Difference and dominance*. Rowley, MA: Newbury House.

Thome, B., Kramarae, C., & Henley, N. (Eds.) (1983). *Language, gender and society*. Rowley, MA: Newbury House.

Uchida, A. (1992). When 'difference' is 'dominance': A critique of the 'anti powerbased' cultural approach to sex differences. *Language in Society, 21,* 547-568.

Varttala, T. (2001). Hedging in scientifically oriented Discourse: Exploring variation according to discipline and intended audience. Doctoral dissertation. University of Tampere.

Vass, H. (2004) Socio-cognitive aspects of hedging in two legal discourse genres. *IBÉRICA, 7,* 125-141.

Vazquez, I. & Giner, D. (2008). Beyond mood and modality: Epistemic modality markers as hedges in research articles. *Revista Alicantina de Estudios Ingleses, 21,* 171-190.

Vold E. T., (2006). Epistemic modality markers in research articles: a cross-linguistic and cross-disciplinary study. *International Journal of Applied Linguistics, 16(*1*).*

Zuck, J. & Zuck, L. (1986). Hedging in news writing. In A. M. Cornu, J. Vanparijs, & M. Delahaye (Eds.) *Beads or Bracelects: How Do We approach* LSP?, 172-181. Leuven, Belgium: Oxford University Press.

3ARABIZI - WHEN LOCAL ARABIC MEETS GLOBAL ENGLISH ON THE INTERNET

Robert Michael BIANCHI
Virginia Commonwealth University in Qatar
rmbianchi@qatar.vcu.edu

Abstract

Arabic is the official language of Jordan. Yet, English is a language of prestige among many upwardly mobile Jordanians. Sakarna (2006) dubs a hybrid language comprised of a mixture of these two languages "Englo-Arabic". In online contexts, a similar hybrid language has emerged. Often popularly labeled "3arabizi" or "Arabish", a blended word based on the words "Arabic" and "English", this mixed code is the most commonly encountered form of language for composing forum messages on the popular Jordanian website, Mahjoob.com (http://www.mahjoob.com). The most striking feature of 3arabizi is that it is written in Latin script and uses arithmographemics i.e. numbers as letters to represent Arabic sounds that do not occur in English. This article presents the key orthographical features of 3arabizi and discusses its topical occurrence when compared to both Arabic and English as observable within a purposive sample of web forum messages collected from Mahjoob.com.

Keywords

Arabic; English; script; CMC; globalization

Izvleček

Uradni jezik Jordanije je arabščina, vendar med mnogimi Jordanci, ki se vzpenjajo po družbeni lestvici, velja angleščina za jezik prestiža. Sakarna (2006) poimenuje hibridni jezik, ki je mešanica angleščine in arabščine "angloarabščina" ("Englo-Arabic"). Na internetu se je pojavil podoben hibridni jezik, ki mu pogosto pravimo "3arabizi" ali "Arabish". To je najbolj pogosta oblika jezika na forumih priljubljene internetne strani v Jordaniji, Mahjoob.com (http://www.mahjoob.com). Najbolj vpadljiva lastnost 3arabizija je pisava, saj se uporablja latinica, določene glasove v arabščini, ki nimajo ustreznice v angleški abecedi, pa nadomeščajo številke. Članek predstavi osrednje ortografske lastnosti 3arabizija in razišče njegovo pojavnost v primeravi z arabščino in angleščino na izbranem vzorcu internetnih forumov na internetnih straneh Mahjoob.com.

Ključne besede

arabščina; angleščina; pisava; CMC; globalizacija

1. Introduction

3arabizi is a mixed language comprised of Arabic vernacular written in Latin script and English found in Computer-mediated Communication (CMC) contexts. Despite being a relatively new form of language, on certain websites such as mahjoob.com, 3arabizi is actually more commonly used for composing forum main messages than either Arabic or English especially for certain topics. This article provides a brief description of the unique orthographic features of 3arabizi, namely its use of arithmographemes (Bianchi, 2005). This is followed by a discussion of the topical use of 3arabizi vs. both Arabic and English within the 41 topical forums on the English website of mahjoob.com.

1.1 Background

Auer (1998, 2008) deplores a monolingual bias in code-switching research which makes a priori assumptions about the existence of distinct and discrete linguistic systems which are then mixed in the speech of bilinguals to produce code-switching. Instead, Auer (1998) posits the possible existence of mixed codes or "fused lects" as the normative code of interaction among certain groups. Such a categorization naturally blurs the lines between discrete linguistic varieties. Prior research into 3arabizi seems to point toward the existence of such a hybridized form of language which, while using the Latin script, incorporates lexicogrammatical elements from both Arabic and English (Abdallah, 2008; Al-Tamimi & Gorgis, 2007; Al Share, 2005; Palfreyman & Al Khalil, 2003; Sakarna, 2006; Warschauer, El Said, & Zohry, 2002). Further, most of these studies also point to a unique feature of this hybrid language, namely the use of numerals as graphemes in order to represent Arabic sounds which have no ready or widely agreed upon equivalents in the Latin alphabet (Abdallah, 2008; Al-Tamimi & Gorgis, 2007; Al Share, 2005; Palfreyman, 2001a; Palfreyman & Al Khalil, 2003; Warschauer et al., 2002). The author labels such numerals used as graphemes "arithmographemes" (see Bianchi, 2005).

2. 3arabizi and its arithmographemes

Table 1 adapted from Palfreyman and Al Khalil (2003) illustrates the use of arithmographemes in 3arabizi:

Table 1: The Arithmographemes of 3arabizi

Sound	Arabic letter	Numeral	Example (with translation)
/ħ/ (a heavy /h/-type sound)	<ح>	<7>	<wa7ed> (one)
/ʕ/ (a tightening of the throat resembling a light gargle)	<ع>	<3>	<ba3ad> (after)
/t'/ (the emphatic version of /t/)	<ط>	<6>	<6arrash> (he sent)
/s'/ (the emphatic version of /s/)	<ص>	<9>	<a9lan> (actually)
/ʔ/ (glottal stop)	<ء>	<2>	<so2al> (question)
/x/ (final sound in Scots "loch")	<خ>	<'7>	<'7ebar> (news)
/ɣ/ (voiced version of above)	<غ>	<'3>	<'3ada> (lunch)
/ð'/ (the emphatic version of /ð/)	<ظ>	<'6>	<'6ahry> (my back)
/d'/ (the emphatic version of /d/)	<ض>	<'9>	<man3ara'9> (not shown)

As can be seen above, a certain visual similarity exists between most of the Arabic letters and the numerals selected to represent them when Arabic is written in Latin script. For instance, note that the Arabic letter < ع> becomes <3> in 3arabizi-style Latinization, where the Arabic letter appears to be inverted. The sound being represented here is the voiced pharyngeal affricate [ʕ], which has no commonly agreed upon graphic representation when Arabic is transliterated using Latin script.

Now that the unique orthographic features of 3arabizi have been surveyed, its topic-related occurrence on the mahjoob.com website will be discussed. But first, a brief overview will be given of the mahjoob.com, the website from which the data were collected.

3. The Data: Mahjoob.com and its Web Forums

The Mahjoob.com is website owned by Mr. Emad Hajjaj, a popular political cartoonist from Jordan, currently living in London. The website itself is hosted in Jordan and a significant number of its users appear to be Jordanian as well. However, the site also attracts posters from across the Arabic-speaking world and from the Arab diaspora as reflected by its advertising. Mahjoob.com is organized into two parallel websites, an Arabic one, and an English one. By November 2008, the Arabic site contained 35 forums, 1,330,999 posts, 58,855 threads, and 28,025 members while the English site contained 41 forums and sub-forums, with 982,084 messages (or posts), and 13,724 members. The website is actually composed of several linked web pages, the largest of which are the forum web pages. On the main portal to the English side of mahjoob.com there is a menu that provides various links. For instance, visitors can select the Arabic link and be taken to the Arabic website of mahjoob.com where they

can then enter the Arabic-language forums. Alternatively, the visitor can stay within the English website and visit its blogs, archives, and, of course, its forums.

Given this surface division of the website by language, one might expect the English side of the mahjoob.com website to feature only English-language content and the Arabic side to feature only Arabic-language content given the pre-eminent status of English on the internet noted by researchers such as Crystal (2001). Accordingly, one would expect that the English used on the English website to be written in Latin script and that the Arabic on the Arabic website be written in the Arabic script as they are conventionally written in most offline domains[1]. However, even the most superficial browsing of both the Arabic and English forums of mahjoob.com makes it clear that forum posters do not follow these well-established conventions. Consider, for instance, the following screenshot (Figure 1) taken from the main page of the English website:

Figure 1: Screen shot of the mahjoob.com English website main page

[1] Holes (2004) does note the existence of a written "mixed" variety of Arabic-scripted Arabic composed of standard and vernacular forms in limited use among Egyptians in popular print media such as magazine editorials. He observes that this kind of colloquial written style helps authors get closer to their readership and appear more 'folksy' (2004, pp. 381-382).

The English featured here is ostensibly standard with readily understood lexis and normative spelling conventions. However, on the left-hand side of the screen shot we can see the menu entitled, "Discussions". Underneath this, there are several forums which a website visitor can choose from. Already, from the red, bolded, asymmetrical font for label "Discussions", we anticipate a more playful, less formal kind of discussion. The punctuation accompanying the subsequent discussion options such as "Men's Corner", "Men Only!!" and "Girls Talk", "No MEN!!" marked as flaming type language (see Herring, 1996) indexes informality. Upon deeper probing, it becomes clear in fact that much of the English used in the forums is actually of an informal, non-standard style. Several of the features of Netspeak (Crystal, 2001) are immediately observable: abbreviations such as "plz", neography (cf. Anis, 2007) such as "u" for "you", "coz" for "because" and "r" for "are" are all commonplace in the forum texts. Thus, there is a picture of English which is not altogether uniform: different levels of formality are indexed by distinct lexis and orthography.

Interestingly, in addition to English, other forms of language are immediately discernible on this web page such as Arabic. For instance, the website's logo and name are given as Abu Mahjoob written in Arabic alongside Mahjoob.com in English. Also, on the top menu, the rightmost link says "Arabic" written in Arabic and thus provides an option for website viewers who want to visit the Arabic side of the website. Again, on the right side menu, there are links to articles with Arabic names. The political cartoon of the website's mascot, Abu Mahjoob, is also written in Arabic-scripted Arabic albeit in a Jordanian vernacular variety.

Similar to the English observable on the English website, the Arabic found on mahjoob.com cannot be said to be of one single variety or style. For instance, discussions are often carried out in written vernacular Arabic as opposed to Modern Standard Arabic, which is the normative code for virtually all traditional written discourse in offline contexts (Holes, 2004). In this connection, it is telling that in Figure 1 above the cartoon character Abu Mahjoob "speaks" in Jordanian vernacular in the cartoon snippet just as one would expect in a "real" offline spoken context[2]. And yet, in order to create this "real life" feeling in the cartoon, the cartoonist was forced to break the conventions of Modern Standard Arabic writing by employing vernacular lexis and structure and non-standard orthography such as doubled vowels which signal to the reader the oral nature of the interaction (cf. Holes, 2004, pp. 381-382).

3arabizi is also apparent on this webpage, in the phrase "e7ke wa fadfed" (*trans.* "get it off your chest") found under the "discussions" menu of forums. But while

[2] The cartoon approximately translated says, "So, basically you're (the government) telling us that the global economic crisis has actually increased the size (lit. shoe-size) of the global economy! Well, there's someone here who would like to direct a question to you (lit. towards your head)!" This is a play on words in Arabic as "your head" means "you" emphatically (cf. AAVE 'your ass' for 'yourself'). This cartoon, which was published on March 7th, 2009, invokes the shoe-throwing incident which took place on December 14th, 2008 when then US President George W. Bush was the target of an irate Iraqi reporter during an unannounced visit to Iraq.

scarcely noticeable on the main page, the presence of such 3arabizi is in fact much greater within the web forum main messages[3] as will be seen below. Of peripheral interest, Arabic messages are also posted within the English website and, to a far lesser extent, a smattering of English messages within the Arabic website. There are also several discussion threads on the website featuring code-mixed and script-mixed messages where posters switch between Arabic, English, and 3arabizi.

Having provided an overview of the English website of mahjoob.com, it will now be useful to describe how the data were collected and analyzed.

4. Method

In order to determine how widespread the use of 3arabizi was within the forums of the English-language website of Mahjoob.com, a purposive sample of all messages posted between March 2007 and May 2008 was downloaded and compiled into a corpus. This resulted in a corpus of 460,220 messages, found within 21,626 discussion threads. The discussion threads, in turn, were found within 41 topical forums.

Using language-specific wordlists based on the Arabic Gigaword and the British National Corpus (BNC) wordlists, each message was annotated to indicate whether it was composed in Arabic, English, or a mixture of the two. A third wordlist was developed to indicate messages written in 3arabizi. In the process of annotating Arabic, BNC English, and 3arabizi messages, three more types of messages were discovered: those that contained a mixture of Latin script and Arabic script (Mixed Script messages), those that contained typically single Islam-related Arabic items transliterated without arithmographemics (Salafi English messages), and those that contained items that were English mixed with items that could not be categorized (Non-BNC English messages)[4]. This resulted in the following six linguistic labels for messages in the corpus:

1. Arabic
2. BNC English
3. 3arabizi
4. Mixed Script

[3] Main messages refer to all messages except the initial or seed message in a discussion thread. Wodak and Wright's (2007) earlier research into multilingual threaded discussions discovered linguistic differences between seed messages and following messages. Following their lead, the author has decided to differentiate between seed messages and main messages, arguing that the potential effect of topic on code choice would be more apparent the further removed a message was from the initial message Consequently, seed messages, which accounted for less than 5% of all messages in the corpus, were excluded from the chi-squared analyses of topic and code choice since they were found to be linguistically distinct from main messages in the present data set as well..

[4] Such items were either neologisms such as “wiki” or “Obama” not contained in the original BNC wordlist, or items from languages other than vernacular Arabic e.g. Turkish, Circassian, or transliterated Hebrew items as found in the World Talk forum, dedicated to learning other languages besides Arabic and English.

5. Salafi English
6. Non-BNC English

Using these six labels, all messages were annotated for linguistic content. Next, in order to determine the statistical frequencies of messages composed in each of these six groupings, the data was converted into an SPSS[5] format in which each corpus message became a unique case defined by several variables such as language of message, thread of message, topical forum of message, author of message, etc. These variables could then be used to carry out different statistical procedures such as chi-squared tests of significance. Another advantage was that frequencies of messages in each language could be displayed in a table format and analyzed for statistical significance. Each message was also annotated to indicate which forum it belonged to. Since there was a relatively large number of forums (41 in total), it was desirable to code these into broader topical categories. Thus, a second step involved recoding all 41 forums into eight overarching topics adapted from Bentahila (1983) in his study of topic-related code choice involving Arabic and French. The adapted topics in the present study were 1) Humour and jokes, 2) Poetry, 3) Work and study, 4) Friends and family, 5) Local/Regional culture, nationality, and politics, 6) Hobbies and pastimes, 7) Gender and age-related forums, and 8) General discussion/opinion (see Bentahila, 1983). Initially, forums were coded into these topics based on their forum titles. However, in order to verify that these forums were in fact connected to the topics into which they had been placed, threads from all 41 forums were randomly sampled and read for topical content. This process did in fact confirm the original topical coding of most forums[6]. This codification resulted in the creation of eight overarching topics as seen below in Table 2:

Table 2: The Eight Overarching Forum Topics based on the Original 41 Topical Forums

	Topic Name	**Topical Content**	**Forum(s) within Topic**
1	Humour and Jokes	Mainly imported jokes from other Arabic-language websites along with comments on these	*Joke Zone*
2	Poetry	Imported and original poetic compositions in both English and Arabic along with comments on these	*A7la Alkalam, Copied Material, San7 w Le3b W Jadd w 7obb, and Soul Retreat*

[5] SPSS refers to "Statistical Package for the Social Sciences". It is a popular suite of software applications that can be used to carry out several different types of statistical tests on data sets such as chi-squared, ANOVA, and multiple variable regression tests.

[6] A notable exception was the *Soul Retreat* forum which was initially coded as a General Discussion/opinion forum. Upon random sampling of its threads it was found to belong in fact to the topic of Poetry.

	Topic Name	Topical Content	Forum(s) within Topic
3	Fields of work and study	Mainly imported texts within various professional and scholastic areas as well as job-seeking and study-related discussions	*Engineering, Graphic Design and Architecture, Health and Science, and Politics, Business and Economy*
4	Friends and Family	Poster-composed messages of well-wishing and congratulations for important life events such as births, weddings, birthdays, and holidays	*Family Matters, Pink and Blue, and Wishes and Greetings*
5	Local/Regional culture, nationality, and politics	Mixture of imported news pieces and poster-composed discussions on political, religious, and cultural issues of Jordan, Palestine, Iraq, and the Middle East	*Arab Arts Symposium. Iraq under Occupation, Kuluna Al Ordun, Palestine, Patriotic and Spiritual Art, and Religion Forum*
6	Hobbies and pastimes	Mixture of imported texts and poster-composed texts sharing interests and links to websites on common areas of interest such as cars, movies, mobile phones, language learning, art, music, cooking, etc.	*Art Gallery, Automotives, Chef's corner, Entertainment, Home Decoration, Humanities, Male Chef Corner, Mobiles, SIG (Special Interest Groups), Sports, Tech Talk, Tv and Movies, and World Talk*
7	Gender- and age-related	Mainly poster-composed messages regarding gender-issues with same sex and same age group peers, including some humour as well	*Girls Talk, Men's Corner, and Teens Thoughts*
8	General Discussion/opinion	Mainly poster-composed messages discussing a broad range of serious and light topics such as future aspirations, general complaints and pet peeves, as well as want-ads.	*Announcements, Dababees and Hala 3ammi, e7ke w fadfed, General Forum, On Focus, Sentiments, and Suggestions and Complaints*

The messages were then recoded to reflect which of the eight overarching topics they occurred with. Once this was done, frequencies of language of main message across overarching forum topic were compiled. These frequencies were then tested for statistical significance using the cross-tab function of SPSS, applying Chi-squared with the value p-0.05.

5. Findings

Tabulation of language type across overarching forum topic resulted in Table 3 below:

Table 3: The Six Conflated Codes Across Overarching Forum Topics

Topic	Arabic	BNC English	3arabizi	Mixed script	Salafi English	Non-BNC English	Notable exceptions found for specific forums within each topic
General Topics	22.3%	22.1%	39.6%	2.2%	4.3%	9.3%	Forums featuring mainly complaints have more BNC English than 3arabizi e.g. *e7ke w fadfed* with 31% vs. 29.6% and *Suggestions and Complaints* with 35.4% vs. 35% respectively.
Humour	66.7%	4.3%	19.5%	5.3%	2.1%	2.0%	*Joke Zone* accounts for 20% of Arabic main messages in the whole corpus
Gender/ Age-related	12.4%	23.5%	48.8%	1.4%	5.3%	8.4%	
Hobbies	8.7%	25.7%	44.7%	1.5%	5.9%	13.3%	*Tv and Movies* and *Tech Talk* had 43.4% BNC English vs. 30.2% 3arabizi and 37.5% vs. 35% respectively.
Local Culture	39.5%	16.5%	25.6%	5.8%	4.8%	7.9%	Jordan-related forumshad more 3arabizi than Arabic e.g. *Kuluna Ordun* (35% vs. 28.8%) and *Patriotic and Spiritual Art* (44.1% vs. 8.3%) respectively. Iraq Under Occupation featured more BNC English (30.6%) than either Arabic (21.1%) or 3arabizi (27.7%). This was because much content came from English-language news websites.
Poetry	70.6%	7.3%	10.3%	7.1%	1.7%	3.0%	Soul Retreat featured 51.6% BNC English as it was a forum dedicated to English-language poetry
Family/ Friends	20.4%	15.9%	49.4%	1.9%	5.5%	6.8%	
Work/ Study-related	8.9%	36.4%	28.4%	2.3%	4.9%	19.1%	
Total	**32.3%**	**17.5%**	**35.5%**	**3.2%**	**4.1%**	**7.5%**	

Examining this table, a number of observations can be made. First, in terms of overall code occurrence, 3arabizi is the most common form of language representing 35.5% of the total main messages, followed closely by Arabic at 32.3%. This is so despite the fact that these forums occur within the English website of mahjoob.com. Both BNC English (17.5%) and Non-BNC English (7.5%) combined account for only 25% of the main messages. Salafi English comes in fifth place with only 4.1% of all

main messages. In last place, Mixed script main messages account for a mere 3.2% of main messages in the corpus.

In terms of language and topic, Arabic was found to predominate in forums related to the topics of Poetry, Humour and to a much lesser extent, Local Culture. It was anticipated that Humour-related main messages might actually be composed in 3arabizi as it is often linked stylistically with Vernacular humour. However, it was discovered that the presence of Arabic with the topic of Humour was accounted for primarily by the Joke Zone forum where Arabic was, in fact, the dominant code. Indeed, Joke Zone main messages account for well over half of all Arabic main messages in the entire corpus. A frequency wordlist of Joke Zone forum's Arabic lexis, revealed that its most frequent items were actually Vernacular Arabic items written in Arabic script as opposed Modern Standard Arabic. The Local Culture topic which included nationalistic and religious forums was also largely written in Arabic, however 3arabizi was relatively common here as well which seemed understandable given the cultural link of such topics to Vernacular Arabic. In contrast, Hobbies and Work/Study-related topics were least commonly expressed in Arabic. Instead, these same two topics were most commonly discussed using BNC English and 3arabizi, followed by Non-BNC English. Regarding Work/Study-related forums, these had an especially high percentage of Non-BNC English main messages, suggesting that technical neologisms may have been prevalent in these forums. In contrast to Arabic, BNC English was least common with both the Humour and Poetry topics. These observations reveal a pattern of Arabic and BNC English being diametrically opposed in terms of the topics with which they most frequently occur. 3arabizi was found to dominate Family and Friends, Gender/Age-related forums, Hobbies, and General Discussion forums. This finding was interesting in that it highlighted the observation that these forums were predominantly composed of message texts that were not imported from other websites, but appeared to be written instead by forum posters themselves. Thus, a pattern emerged strongly suggesting the use of 3arabizi to write personal, intimate, and general texts while Arabic and English were used more often because they were part of imported texts and perhaps also because Arabic was tied thematically to cultural and local topics while English was linked to professional and academic topics.

Interestingly, Mixed script main messages were prevalent in the Joke Zone forum as well. This suggests that material may have been imported from other Arabic websites where a certain amount of Latin script was present alongside Arabic script due to hyperlink-related strings. Meanwhile, Salafi English reveals a relatively low frequency across virtually all topics with 4.1% overall. Still, this is more than double the amount of Mixed script main messages (3.2%), highlighting the paucity of biscriptal main messages in the corpus.

5.1 Summary of Findings

To sum up, there are several clear patterns of code choice related to topic. Topical forums that are both local and formal in content feature relatively high amounts of Arabic (Arabic-scripted Arabic). In this regard, Arabic language poetry, Arabic politics and nationalism, and Islamic religion all favour the use of Arabic. The major exception to this formal Arabic usage trend is the preponderance of Vernacular Arabic language jokes in the Joke Zone forum. Regardless, thematically, all of these forums could be described as connected to local Arabic culture. In contrast, topical forums dealing with more specialized technical content especially related to fields of work and study are dominated by BNC English main messages, perhaps because sources for such content are to be found primarily in English language elsewhere on the web. 3arabizi, the most frequently encountered code in the corpus, was found to dominate topical forums which were less formal and more intimate in content as well as forums that encouraged the sharing of general discussion and opinions. In this connection, Hobby-related forums were also dominated by 3arabizi main messages, though both BNC English and Non-BNC English were also frequently featured in such forums. Again, given the non-culturally-localized nature of many of the hobbies, this was not a surprising finding.

6. Conclusions

This report has described the key orthographic features of 3arabizi, noting its novel features such as the use of arithmographemes in place of letters to represent Arabic sounds not found in English. One of the principal domains for the use of 3arabizi is the internet. In particular, web forums such as those found on the Jordanian website, mahjoob.com, are prime loci for the use of 3arabizi. This study has also showed that 3arabizi occurs with different topics than either Arabic or English do. In this regard, 3arabizi, as the most used code on the mahjoob.com website, was found to occur mainly within forums associated with Family and Friends, General discussions, Gender and Age groups, and Hobbies. In contrast, Arabic was found with forums related to Humour, Poetry, and Local culture, all posting areas where copied material from other websites seemed to be quite prevalent. English, despite being the designated language of the English site and its forums, was least prevalent among these three main codes, serving mainly for discussion of Work and Study-related topics. The relative scarcity of English messages challenges the notion of the primacy of English on the internet. On the other hand, the abundance in the corpus of a mixed code comprised of Vernacular Arabic written in Latin script along with elements from English suggests that the seemingly conflicting trends of globalization and localization of culture may, in fact, be working synergistically to produce new and fascinating hybrid forms of language, of which 3arabizi is a prime example.

References

Abdallah, S. (2008). Online Chatting in Beirut: Sites of Occasioned Identity-Construction. *Ethnographic Studies, 10*, 3-22.

Al-Tamimi, A., & Gorgis, D. T. (2007). Romanised Jordanian Arabic E-messages [Electronic Version]. *The International Journal of Language Society and Culture*. Retrieved March 25, 2009, from http://www.educ.utas.edu.au/users/tle/JOURNAL/

Al Share, B. (2005). A Sociolinguistic Analysis of Jordanian Netspeak (JNS). Unpublished Unpublished Master's Thesis. Jordan University of Science & Technology.

Anis, J. (2007). Neography: Unconventional Spelling in French SMS Text Messages. In B. Danet & S. C. Herring (Eds.), *The Multilingual Internet: Language, Culture, and Communication Online*. Oxford: Oxford University Press.

Auer, P. (1998). From Code-switching via Language Mixing to Fused Lects: Toward a Dynamic Typology of Bilingual Speech. *Interaction and Linguistic Structures, 6*.

Auer, P. (2008). The monolingual bias in bilingualism research - or: Why bilingual talk is (still) a challenge for linguistics. Retrieved 17 February, 2008, from www.germanistik.uni-freiburg.de/auer/?download=monolingual_bias_in_bilingualism_research.pdf

Bianchi, R. (2005). *Revolution or Fad? Latinized Arabic Vernacular.* Paper presented at the 11th TESOL Arabia Conference, Dubai.

Bentahila, A. (1983). *Language attitudes among Arabic-French bilinguals in Morocco*. Clevedon, Avon, England: Multilingual Matters.

Holes, C. (2004). *Modern Arabic: Structures, Functions, and Varieties* (Revised ed.). Washington, D.C.: Georgetown University Press.

Palfreyman, D. (2001a). Informal Latinized Orthographies [Electronic Version]. *Linguist List, 12*. Retrieved November 29, 2004, from http://www.ling.ed.ac.uk/linguist/issues/12/12-2760.html

Palfreyman, D., & Al Khalil, M. (2003). "A Funky Language for Teenzz to Use": Representing Gulf Arabic in Instant Messaging. *Journal of Computer-Mediated Communication, 9*(1).

Sakarna, A. K. (2006). The Englo-Arabic Language of Young Urban Jordanians: The Influence of the Internet, Mobile Phones, and TV Satellites. *Langues et Linguistique*(17), 65-80.

Warschauer, M., El Said, G. R., & Zohry, A. (2002). Language choice online: Globalization and identity in Egypt. *Journal of Computer-Mediated Communication, 7*(4).

Wodak, R., & Wright, S. (2007). The European Union in Cyberspace: Democratic Participation via Online Multilingual Discussion Boards. In B. Danet & S. C. Herring (Eds.), *The Multilingual Internet: Language, Culture, and Communication Online*. Oxford: Oxford University Press.