Eitan Grossman

From *rarum* to *rarissimum*: An unexpected zero person marker

DOI 10.1515/lingty-2016-0001
Received June 8, 2015; revised March 1, 2016

Abstract: This article addresses the problem of crosslinguistic rarity by mapping the types of diachronic factors that contribute to the rarity of a particular feature. It is proposed that a number of different diachronic factors may play a role, such as the rarity of source constructions, the rarity of particular types of change, the number of stages necessary for a particular feature to develop, and the number of pathways that can lead to a particular feature. This article looks at a *rarissimum* of person marking, namely, a zero-marked feminine 2nd singular person index in the Sahidic dialect of Coptic (Afroasiatic; Egypt). It is argued that such markers are rare because they presuppose rare input structures, and most processes of change would lead away from – rather than to – zero-marked 2SG. Furthermore, this study identifies a diachronic process in which a PART of a morpheme is reinterpreted as a segmentable morpheme (in this case, a person index), thereby leading to the loss of a zero person marker. This is the converse of the well-known "Watkins' Law", in which a segmentable person marker is reinterpreted as part of a base.

Keywords: Ancient Egyptian-Coptic, diachrony, inflection, markedness, paradigm, person, rara, syntax

1 Introduction

The present article addresses the issue of crosslinguistic rarity by focusing on a *rarissimum* of person marking. Assuming that the explanation of crosslinguistic rarities requires a diachronic dimension (Bybee 2008), any explanatory account of the problem involves two main questions. First, what are the constraints on language change that inhibit – but do not rule out – the development of the rare feature? Second, what are the developmental pathways that nonetheless lead to the emergence of the rare feature?

Linguists have given different answers to these questions, but several main types can be identified (see Table 1). First, a given feature may be (relatively)

Eitan Grossman, Department of Linguistics, Hebrew University of Jerusalem, Mount Scopus, Jerusalem 91905, Israel, E-mail: eitan.grossman@mail.huji.ac.il

Table 1: Source of crosslinguistic rarity.

Type	Factor	Rare feature	Documentation
Pathway	Few (vs. many) pathways	Closed syllables	Many languages
Stages	Many (vs. few) stages necessary	Endoclitics	Udi (Harris 2008)
Source	Rare (vs. common) source construction	Adverbial subordinator prefixes	Japhug (Grossman et al. 2015)
Type	Rare type of change	Coronal annihilation	Northwest Mekeo (Blevins 2009)

rare because there are fewer pathways that lead to the feature than away from it. Bybee (2001: 195–197) provides evidence for the argument that there are more open syllables than closed syllables, and only open syllables are (nearly) universal, because new open syllables are constantly being created by regular processes of language change (e. g., coda weakening and loss), while there are fewer processes that lead to closed syllables. Another possibility is that some rare features may necessitate numerous diachronic "steps" that occur in a certain order in order to develop, as in Harris's (2008) account of Georgian split case marking or Udi endoclitics. Yet another possibility is that a certain feature may require rare input structures, as in Grossman et al. (2015), which argues that adverbial subordinator prefixes are rare because they are facilitated by the relatively rare VSO order and case prefixes and are inhibited by other word orders or case suffixes. Fourth, some changes simply may be more frequent than others, as proposed by Greenberg (1978), a view adopted by Blevins (2009), who states that most languages have coronal segments because coronal maintenance and coronal creation are more frequent than "coronal annihilation". While the last might not seem like the sort of change one normally encounters in historical linguistics textbooks, it is nonetheless a documented change. Of course, it is generally thought that some changes are regular or require only a single step or causal mechanism, while others are sporadic or require multiple steps or causal mechanisms, e. g., assimilation vs. metathesis. Finally, some properties might simply be inherently unstable, and even though they could develop frequently in language, they might tend to develop into other categories or to disappear. For example, Blevins (2008) suggests that three-way vowel length distinctions may "push the perceptual envelope" and therefore be inherently prone to loss.[1]

1 Joan Bybee (personal communication) points out, for example, that hodiernal pasts may be rare because they tend to quickly become anteriors or general past tenses.

What is common to these diachronic explanations is that they all take crosslinguistic rarity as a problem to be dealt with, explicitly addressing the question how a given linguistic feature can be dispreferred for some reason and nevertheless develop and (perhaps) remain stable over time. It goes without saying that all of these explanations may make reference to a variety of additional factors, whether based on usage, acquisition, or general properties of human physiology and cognition. Importantly, these types of factors that contribute to crosslinguistic rarity are not mutually exclusive, and all things being equal, one would assume that a feature that is characterized by more of the above factors would be rarer than one that is characterized by fewer of them.

This article has three goals. The first is to document a *rarissimum* (Wohlgemuth & Cysouw (eds.) 2010a, 2010b) of person marking: in the Sahidic dialect of Coptic (Afroasiatic; Egypt), the 2SG.F person index is zero. While the facts of Coptic are well known to specialists, thanks to descriptive grammars (e. g., Layton 2004) and an in-depth study by Uljas (2009), the existence of the proposed *rarissimum* is unknown to typologists. There are no documented parallels in other languages, whether within or outside Afroasiatic.

The second goal is to provide a diachronic explanation for the rarity of this linguistic feature. Specifically, it is argued that the first reason that zero-marked 2SG.F is rare is that there are relatively few pathways that could lead to such a paradigmatic zero. This derives from two main facts: first, the SOURCE constructions from which such a feature could arise are relatively rare; and second, most PATHWAYS of change involving differential loss of overt markers, e. g., through phonological erosion, would not lead to zero-marked 2SG but rather to different types of homophonous or syncretic systems without a paradigmatic 2SG at all. Second, adding to the rarity of this feature is the fact that its source construction, i. e., gender marking in person forms, is itself crosslinguistically rare. Third, the feature documented here does not result from differential phonological erosion, based, e. g., on relative frequency, but rather from a series of conditioned sound changes that occurred in particular STAGES. One of which – depalatalization – is of a relatively rare TYPE. As such, the development of zero-marked 2SG.F in Coptic involves ALL of the factors discussed above that contribute to the crosslinguistic rarity of features.

A third goal is to document a rare pathway of change, which is the converse of the better-known "Watkins' Law", which leads to zero marking of 3SG through reanalyzing a 3SG marker as part of the verb stem. In the pathway of change documented here, a zero marker is reanalyzed away, with the result that an erstwhile PART of a morpheme is reanalyzed as a segmentable person marker. This process in effect leads away from the zero-marked 2SG.F, reinstating an overt segmental person marker.

Taken together, these three stages – the identification of a synchronic rarity, the identification of a language-specific diachronic explanation, and the identification of crosslinguistic inhibiting or facilitating factors – provide the necessary first steps of a diachronic account of a crosslinguistic rarity. This approach basically follows Greenberg's (1966a) line of reasoning with respect to explanation in linguistics.

The structure of this article is as follows. Section 2 presents some brief background about zero person marking in crosslinguistic perspective. Section 3 sketches the relevant Coptic data. Section 4 provides evidence that this feature is absent from other Afroasiatic languages, and cannot be treated as an inherited feature, arguing instead that this feature has a straightforward diachronic explanation, and raising questions about its diachronic stability. Section 5 discusses the reasons for the crosslinguistic rarity, arguing that there are few diachronic pathways through which such a feature could arise, while Section 6 argues that this feature is in fact diachronically unstable. Section 7 offers brief concluding remarks.

2 Zero person marking in crosslinguistic perspective

In one of the first discussions of universals of person marking, Uspensky (1972: 68, cited in Cysouw 2003: 57), claims that "[i]f a zero expression occurs in the form of a certain person in the indicative mood, then, included in the meanings thus expressed (i. e. by a zero mark) we find the meaning of the 3rd person or that of the 1st person". Cysouw (2003) corroborates this universal, noting several exceptions in which "Latin-type" paradigms, i. e., paradigms with no syncretism or homophony between any of the singular persons, have zero marking for the 2nd person. For example, in Bongo (Nilo-Saharan; Sudan), independent pronouns have bound ("clitic") realizations when nothing intervenes between the pronoun and the verb. Thus, independent 1SG *ma* is realized as *m-* (1a), independent 3SG *bu* is realized as *b-* (1b), but independent *i* disappears (1c).

(1) a. *m-ony* 1SG-eat 'I ate'
b. *b-ony* 3SG-eat 's/he ate'
c. *Ø-ony* 2SG-eat 'you ate'

However, Cysouw notes that zero 2nd person marking is extremely rare in his sample, even if it appears to be slightly more frequent than zero 1st person

marking. In her sample of 347 languages, Siewierska (2009: 429) finds that zero-marked 1st or 2nd person is "too infrequent to warrant sub-classification" into types of zero marking. On the other hand, the association of zero marking with 3rd person has often been noted and re-confirmed (e. g., Ariel 2000; Bybee 1985; Greenberg 1966b; Siewierska 2009), although this observation might need revision in light of a recent argument that there is in fact "no evidence for zeros to be more common among third than among non-third persons" (Bickel et al. 2015).

These statements refer to what Siewierska (2004: 24) calls zero "in the paradigmatic sense of the term", i. e., person paradigms in which there is a combination of zero and overt person markers. Siewierska (2013) shows that in a 380-language sample, zero 3rd person marking of the S argument is documented in 17 % of the languages (n = 66), and another 9 % (n = 21) could be analyzed as having obligatory zero marking in all 3rd persons, but as Siewierska points out, such languages could also be analyzed as having no 3rd person forms at all. Based on these data, one can conclude that while zero marking in the 3rd person is well documented, it is hardly the majority case.

Bickel et al. (2015) test the hypothesis that "zeros are assumed to develop and be preserved more commonly in third than in non-third person". They find no support for this hypothesis as a synchronic universal: only 35 % of languages have more zero markers in the 3rd person than in other persons, the other 65 % having what the hypothesis would consider "disfavored" structures. Furthermore, Bickel et al. (2015) find no evidence for the hypothesis as a diachronic universal pertaining to paradigms. However, when it is interpreted as a hypothesis about the distribution of zero person marking cutting across paradigms (i. e., "zero markers are more likely to be found in the third than in the first and second person, across all paradigms in a family"; Bickel et al. 2015:4), it is weakly supported.

Interestingly, to the best of my knowledge, large-scale typological studies of person marking have consistently left out gender distinctions in person markers, generally seeing them as a confound (Cysouw 2003; Bickel et al. 2015).

3 Background: Coptic person markers

Coptic[2] has a crosslinguistically unremarkable set of independent person markers, which distinguish singular and plural number in 1st, 2nd, and 3rd

2 Coptic is the latest stage of the Ancient Egyptian language, an independent branch of the Afroasiatic phylum. For overviews of Ancient Egyptian, see Grossman & Richter (2014), Haspelmath (2014a), Loprieno (1995), Loprieno & Müller (2012). The examples in this article, taken from the Sahidic dialect, are glossed following the Leipzig Glossing Rules

Table 2: Independent person markers.

	Singular	Plural
1	*anok*	*anon*
2SG.M	*ntok*	*ntôtn*
2SG.F	*nto*	
3SG.M	*ntof*	*ntoou*
3SG.F	*ntos*	

persons, and masculine and feminine gender in the singular 2nd and 3rd persons, as shown in Table 2. Such systems are common in Afroasiatic languages, a fact often pointed out in typological studies of person (Cysouw 2003; Siewierska 2004).

On the other hand, Coptic has a highly complex series of bound person indexes, which distinguish the same categories but show considerable allomorphy, depending on the morphosyntactic environment and the phonological properties of the host (Polotsky 1960; Layton 2004). For example, prefixed person indexes differ in a synchronically unpredictable way depending on the verbal template. Table 3 shows two paradigms of the verb *-me* 'love', the first of which shows the present tense, with verb-initial (prefixed) A/S person indexes, and the second of which shows the past tense, with verb-initial TAM/polarity markers followed by A/S person indexes. Note that while 2SG.M, 3SG.M, 3SG.F, and 2PL person indexes are identical in both paradigms, those for 2SG.F, 1PL, and 3PL (marked in bold) differ.

Table 3: Bound A/S markers in two verbal paradigms.

	Present	Past
1SG	***ti****-me*	*a-**i**-me*
2SG.M	*k-me*	*a-k-me*
2SG.F	***te****-me*	*a-**Ø**-me*
3SG.M	*f-me*	*a-f-me*
3SG.F	*s-me*	*a-s-me*
1PL	***tn****-me*	*a-**n**-me*
2PL	*tetn-me*	*a-tetn-me*
3PL	***se****-me*	*a-**u**-me*

(http://www.eva.mpg.de/lingua/resources/glossing-rules.php) and the *LT* style sheet, and are transliterated according to the Leipzig-Jerusalem system (Grossman and Haspelmath 2014).

For a given person, A/S and P indexes are identical in some cases and differ in others, again depending on the morphosyntactic environment and the phonological properties of the host. For example, consider the A/S and P 3SG.M indexes, which are identical, in (2a, b):

(2) a. *a-f-pôt*
PST-3SG.M.S-flee
'He fled.' (Besa, in Kuhn (ed.) 1956: 31, line 24)
b. *a-u-toš-f*
PST-3PL.A-appoint-3SG.M.P
'They appointed it.' (Besa, in Kuhn (ed.) 1956: 9, line 7)

However, some persons show allomorphy of the A/S marker depending on the morphosyntactic environment. For example, while the 3PL marker in the past verb form is *-u-* (ex. 2b), in the present it is *se-*, as in (3).

(3) *se-moute* *ero-Ø*
3PL.S.PRS-call to-2SG.F
'They call you.' (Besa, in Kuhn (ed.) 1956: 37–38, lines 33–1)

The allomorphy of the P indexes, on the other hand, is determined by the phonological environment. Simplifying matters somewhat, if the lexical verb ends in a short vowel, the bound P marker is *-i*, while if it ends in a consonant or a long vowel, it is *-t* (for a more precise description, see Layton 2004).[3]

(4) a. *nne-nai* *-taho-i*[4]
OPT.NEG-DEM.PL -touch-1SG.P
'May these things not happen to me.' (Ruth 1:16, in Worrell (ed.) 1942: 44, line 68)
b. *n-se-toms-t*
SEQ-3PL.A-bury-1SG.P
'And may I be buried.' (Ruth 1:17, in Worrell (ed.) 1942: 44, line 68)

3 Coptic also has Differential Object Indexing (DOI), since P arguments can be incorporated or case-marked if lexical noun phrases, or indexed on the verb or case-marked if pronominal (Grossman 2014). However, this does not have any bearing on the issue presented here, and will not be expanded on further.

4 The convention of a space before the hyphen follows Leipzig Glossing Rule 2A, which is here interpreted as an indication that the glossed item is a single morphosyntactic word but two phonological words.

Table 4: Zero-marked 2SG.F.

	3SG.M	2SG.F	Lexical NP
Past	*a-f-me*	*a-Ø-me*	*a-prôme -me*
Negative past	*mp(e)-f-me*	*mpe-Ø-me*	*mpe-prôme -me*
Aorist	*ša-f-me*	*šare-Ø-me*	*šare-prôme -me*
Negative aorist	*me-f-me*	*mere-Ø-me*	*mere-prôme -me*
Optative	*e-f-e-me*	*ere-Ø-me*	*ere-prôme -me*
Negative optative	*nne-f-me*	*nne-Ø-me*	*nne-prôme -me*

Focusing on 2SG.F A/S indexes in main clauses, we find that zero marking is the norm. Table 4 compares several TAM-marked verbal constructions: (i) the verb form with non-2SG.F person index, (ii) the verb form with 2SG.F, and (iii) the verb form with lexical noun phrase subject. The lexical verb used here is *-me* 'love', and the lexical NP subject used is *p-rôme* [DEF.M.SG-man] 'the man'.

In Table 4, we see that the 2SG.F person indexes in Coptic are often paradigmatic zeros, noting also that the TAM formatives occurring with the 2SG.F are precisely those that occur with lexical noun phrase subjects. In fact, 2SG.F is usually zero-marked, although it also has non-zero allomorphs. In some cases, these non-zero formatives are regular allomorphs, as in the present tense.

(5) *te-na-sei*
2SG.F-FUT-be.sated
'You will be sated.' (Besa, in Kuhn (ed.) 1956: 97, line 16)

In other cases, zero and non-zero formatives vary in the same construction, and can even occur as textual variants in different manuscripts, as in (6).

(6) a. *a-Ø-sôbe*
PST-2SG.F-laugh
'You laughed.' (Genesis 18:15, cited in Uljas 2009: 176)
b. *a-r-sobe*
PST-2SG.F-laugh
'You laughed.' (Genesis 18:15, cited in Uljas 2009: 176)

The distribution of zero and non-zero 2SG.F variants is complex, and a full account is beyond the scope of this article, but Uljas (2009) provides an

exhaustive description.[5] The important thing, for present purposes, is that Coptic has a number of constructions in which 2SG.F is represented by a paradigmatic zero.

Nonetheless, it is important to keep in mind two facts, to which we return in more detail in Section 6. First, Coptic 2SG.F has non-zero allomorphs. Second, Coptic has a number of constructions in which 2SG.F could be analyzed as zero, but could also be analyzed as non-zero, usually involving *-r(e)-*. This ambiguity results from the fact that TAM formatives that occur with 2SG.F reference are often homophonous with TAM formatives that occur with lexical noun phrase subjects. For example, the Aorist (basically, a habitual form) formative that occurs with most bound person markers is *ša-* (7a), but *šare-* when it occurs with 2SG.F (7b) or lexical noun phrase subjects (7c).

(7) a. *ša-f-ei* *ebol*
AOR-3SG.M.A-come out
'It comes out.' (Matthew 12:43)

b. *šare-Ø-tsio-Ø* *m-p-oeik* *mn-p-moou*
AOR-2SG.F.A-satisfy-2SG.F.P OBL-DEF.M.SG-bread with-DEF.M.SG-water
'You satisfy yourself with bread and water.' (Shenoute, in Leipoldt (ed.) 1908: 204)

c. *šare-p-at^{h}êt* *kmš-te-sbô* *m-pe-f-eiôt*
AOR-DEF.M.SG-fool mock-DEF.F.SG-wisdom of-POSS-3SG.M-father
'The fool mocks the wisdom of his father.' (Proverbs 15:5)

4 So where did Coptic get its zero 2SG.F?

First of all, we can safely rule out inheritance as the source of this rare feature. It is not reconstructible for Proto-Afroasiatic, since Proto-Semitic (Goldenberg 2013), Proto-Chadic (Frajzyngier 2012, personal communication), and Proto-Berber (Kossmann 2012, personal communication) have overt 2SG markers, usually involving a dental or alveolar consonant. Cushitic and Omotic languages tend to have distinctive 2SG person markers (Mous 2012, personal communication; Amha 2012). Most importantly, zero 2SG is not attested in the early stages of

5 An anonymous reviewer pointed out that Coptic, as a dead language, does not afford many instances of 2SG.F, since most texts do not involve women as addressees. While it is probably true that there are more tokens of 2SG.M than 2SG.F in the extant corpus, it should be noted that the Coptic corpus is very extensive, and comprises many texts of diverse sorts (narratives, sermons, legal texts, private letters, and more) with women as addressees.

Ancient Egyptian, which has an overt 2SG marker, /c/ (orthographically). As such, we can safely conclude that zero-marked 2SG in Coptic is not an inherited feature, but rather a language-internal development.

When seen from a diachronic perspective, it turns out that there are several documented processes by which a language can develop zero person marking. One such process is known as Watkins' Law, according to which "third person markers are reanalyzed as part of the verbal stem, giving thus rise to zero marking in the third person" (Bickel et al. 2015), a process that Watkins posited to account for the reanalysis of the 3SG ending *-t* as part of the verbal stem in the course of development from Proto-Iranian to Persian (Watkins 1962: 94; see Siewierska 2009 and Bickel et al. 2015 for a full discussion).

Another possibility is that person marking simply never developed for a particular person, but did for others (i. e., differential grammaticalization). Yet another is differential phonological erosion, perhaps due to frequency (Bybee 2007; Haspelmath 2008).

Zero 2SG.F person markers in Coptic developed due to sound change, but not in a way that is easily explained by most phonological accounts of the development of zero markers, which have looked mainly to frequency-driven explanations. In Coptic, the 2SG.F zero marker is the result of a long-term process of conditioned segmental sound changes, in which:

(i) original /*k/ was palatalized to /c/ before high front vowels,
(ii) in some environments, /c/ merged with /t/ by depalatalization or fronting of the place of articulation, and
(iii) /t/ was lost in a number of environments; it was almost invariably lost when word-final, perhaps first going through a stage of debuccalizing to /ʔ/.

The details of these changes are still poorly understood, but the main line of development is clear (Kammerzell 1998, 2005; Peust 1999), and crucially, word-final position is the position in which these changes consistently took place. Table 5 shows these developments; the changes involving word-final position (/_____#) do not mean that this is the only environment in which the

Table 5: Relevant sound changes.

	Change	Probable dating	Source
(i)	**k → c /_____i*	pre-Old Egyptian	Kammerzell (1998: 38, 2005: 182–192)
(ii)	*c→t/_____#*	Old Egyptian or Middle Egyptian	Kammerzell (2005: 230); Peust (1999: 123–124)
(iii)	*t → (ʔ →) ∅/_____#*	pre-Late Egyptian	Peust (1999: 152–155)

change took place, but rather a weaker claim, i. e., that in this environment the change took place.

Evidence for the first change (i), i. e., the palatalization of *k > /c/ before high front vowels, is based on reconstruction and on language-internal phonological analysis (Kammerzell 1998, 2005). Evidence for the second two changes is provided in Table 6 (Loprieno 1995).

Table 6: Stages of phonological change.

	Transliteration		**Phonological representation**
Stage 1	rmT	'man'	*/ra:mac/
Stage 2	rmt	'man'	*/ra:mat/
Stage 3	*rôme*	'man'	*/ro:mə/

Importantly, none of these sound changes are outlandish per se, although one, the depalatalization or fronting of /c/ to /t/, does seem to be relatively rare.

(i) The palatalization or coronalization (Telfer 2006) of velars is well documented in the world's languages (Bhat 1978; Telfer 2006; Bateman 2007).

(ii) Change from /c/ to /t/, while rarer, is also documented, both as the result of a synchronic alternation and as the result of postulated historical changes. For example, in Korean (isolate; North and South Korea), /c/ is neutralized to /t/ in coda position (Sohn 2001: 163–165). In Tarahumara (Uto-Aztecan; Mexico), the alveolopalatal affricate can be depalatalized and realized as an alveolar affricate before /a/ (Caballero 2008: 36).[6]

(iii) Finally, loss of *t*, whether via debuccalization, e. g., /t/ > /ʔ/ as in some English dialects in syllable-final position, as well as in Burmese, Kagoshima Japanese, and Yamphu (O'Brien 2010, 2012: 10), or via deletion is extremely well documented across the world's languages.

6 I thank Juliette Blevins (personal communication, 8 April 2015) and Ian Maddieson (personal communication, 4 April 2015) for drawing my attention to these examples. Furthermore, in response to a query on the LINGTYP mailing list (http://listserv.linguistlist.org/mailman/listinfo/lingtyp; 4 April 2015), Siva Kalyan and Martin Kümmel pointed out similar changes in the Indo-European languages Lycian, Cypriot Greek, Sindhi, some Spanish dialects, as well as the Finno-Ugric languages Udmurt and Khanty. Olle Engstrand and Guillaume Jacques drew my attention to similar changes in Proto-Algonquian and Proto-Wintun. Henning Andersen pointed out that the changes involved might be complex, involving additional interim stages. Mark Donohue provided additional examples of both /t/ > /c/ and /c/ to /t/. Sven Grawunder informed me that a search in the ASJP database (http://asjp.clld.org), as well as Brown et al. (2013), turned up a significant number of correspondences of / ʧ / <> /t/ and /c/ <> /t/, emphasizing that this does not tell us anything about the direction of change.

The only change that is apparently rare is the depalatalization or fronting of /c/ to /t/.

5 Why aren't 2SG person zero markers more common?

5.1 2SG zero markers: A quick typology

Given the diversity of diachronic routes that can lead to zero person marking, it is nonetheless curious that more such cases have not been diagnosed. However, on further reflection, its rarity makes sense. Consider that zero person marking in the strict paradigmatic sense is likely to be diagnosed only when it is opposed to other overt markers. As such, a dedicated zero marker for 2SG – and not a combined 1SG+2SG or 2SG+3SG marker – would be identifiable only in relatively specific circumstances.

(i) For example, assume that the verb stem *bata* in a hypothetical language means 'jump'. One possibility would be that the language does not have bound person markers, so the entire paradigm would be simply *bata*. This would not usually be considered a zero-marked 2SG, but rather a type of undifferentiated person, unless person is marked elsewhere in the system ("123" type in Siewierska 2004: 76).

(ii) Another possibility is that the language has one unmarked form for 1SG + 2SG, e. g., *bata*, but a marked form for 3SG, e. g., *bata-t* ("12 vs. 3" type in Siewierska 2004: 76), as in English *-s* for 3SG and zero for 1SG and 2SG. Such paradigms, as Cysouw (2003: 40) points out, do not have a distinction between speaker and addressee, but rather have a 3rd vs. non-3rd system.

(iii) A third possibility is that the language has an unmarked form for 2SG + 3SG (e. g., *bata*) but an overtly marked form for 1SG (e. g., *bata-m*), as in Atapaka (Penutian; U.S.A.) singular S/A suffixes, which have *-o:* for the 1SG and zero for 2SG and 3SG (Siewierska 2004: 77). Such paradigms have a 1st vs. non-1st system.

(iv) Another possibility is 1SG + 3SG homophony, with 2SG overtly marked, e. g., *bata* vs. *bata-s* ("13 vs. 2" type in Siewierska 2004: 76), as in the Spanish imperfect, which has *-s* for the 2SG but zero for 1SG and 3SG. Such paradigms have a 2nd vs. non-2nd system. This picks out the 2SG as a distinctive category, but does not have the zero-marked 2SG we are interested in.

(v) Yet another possibility is 1SG + 3SG overtly marked (e. g., *bata-t*), with zero marking for 2SG (e. g., *bata*), and as such, a 2nd vs. non-2nd system. This does pick out a distinctive zero-marked 2SG.

(vi) A final way for a language to have a zero-marked 2SG would be to have distinctive overt markers for 1SG and 3SG (e. g., *bata-m* vs. *bata-t*), and zero for 2SG (e. g., *bata*). This is the case of Coptic as described in this article.

Of course, there are more possible systems, but the point to be made here is that in many systems of bound person markers, even if the addressee is not overtly coded, it does not mean that the system has a distinctive zero-marked 2SG. This is likely to be one of the factors contributing to the crosslinguistic rarity of zero-marked 2SG.

5.2 Zero 2SG markers through loss

A possible explanation for the rarity of zero-marked 2SG is that overt (i. e., non-zero) person markers associated with the 2nd person, when lost, can lead to homophonous person marking rather than to a paradigmatic 2SG zero marker. Of the scenarios described in Section 5.1, (i) has no overt person markers to lose. Systems (ii) and (iii), if the overt non-2SG person marker were lost, would result in (i), which has no person marking. In (iv), the loss of the only overt person marker, 2SG, would also result in a paradigm without overt person markers, and for (v) or (vi), the loss of one of the person markers would result in a "12 vs. 3" or "1 vs. 23" system, neither of which has a specifically zero-marked 2SG.

Yet another type of pathway of differential loss leading to zero-marked 2SG could involve systems with overt but homophonous or syncretic person markers, in which the marker itself is lost only for the 2SG and not for one of the other persons. For example, the object markers of Chai (Surmic, Nilo-Saharan; Ethiopia) are a "12 vs. 3" system, in which all markers are overt; see Table 7. If the P suffixes *-in* or *-n^{y}* were lost only when the reference is 2SG (but not 1SG),

Table 7: Object markers in Chai (cited in Siewierska 2004: 76).

	Imperfective P suffixes	**Perfective P suffixes**
1SG	*-in*	*-n^{y}*
2SG	*-in*	*-n^{y}*
3SG	*-e*	*-u/-a*

then a zero-marked 2SG could be the result. However, I am unaware of any such process being documented, and it would be hard to explain.

One way for a zero-marked 2SG to develop is shown by a hypothetical future stage of Koiari (Papuan) singular person markers in the realis indicative mood, which have a "13 vs. 2" system, in which all markers are overt; see Table 8. If the 2SG markers were lost, then the result would be a zero-marked 2SG, although it would then be an "addressee vs. non-addressee" system.

Table 8: Singular person markers in the realis indicative mood in Koiari (cited in Siewierska 2004: 77).

	Present	Past
1SG	*-ma*	*-nu*
2SG	*-a*	*-nua*
3SG	*-ma*	*-nu*

All in all, there are relatively few ways for a PARADIGMATIC zero-marked 2SG index to develop, at least via loss. The primary one is likely to be one in which a "Latin-type" paradigm, i. e., one in which all persons are marked by different overt formatives, loses the explicit 2SG marker, although a "Koiari-type" system could also be a starting point for an eventual zero-marked 2SG.

5.3 Other paths to zero-marked 2SG

Other than differential loss, another type of possibility is differential grammaticalization of person markers, such that 1SG and 3SG markers develop, e. g., from pronouns, while the 2SG simply does not grammaticalize. However, there are to date no accounts that would support such a scenario. Thus, Siewierska (2009: 426–427) reviews the various accounts of the development of 3rd person zeros, and none of them predict that 2nd person should develop zero marking. For example, Ariel's (2000) Accessibility Theory predicts that due to their higher accessibility, "first and second person pronouns are more likely to undergo phonological reduction, cliticization and affixation than third person forms. Hence the frequent occurrence of third person zeroes as opposed to first or second person zeroes" (Siewierska 2009: 426). As such, differential grammaticalization is an unlikely explanation. Moreover, since zero-marked 2SG is so rare in person marking systems, we can take this is as prima facie evidence that such a historical pathway is rare.

5.4 Why is zero-marked 2SG.F so rare?

Having seen that there are relatively few pathways through which zero-marked 2SG indexes can develop, we now turn to the question of gender: why is it that zero-marked 2SG.F is so rare?

As Siewierska (2013) has shown, gender marking in independent pronouns is relatively rare, overall. Out of a sample of 378 languages, 254 had no gender distinctions at all, while where such distinctions exist, gender distinctions in the 3rd person far outnumbers such distinctions in other persons. All in all, gender distinctions in 2SG markers are rare: Siewierska (2004: 104) reports that of a 133-language sample, only 18 % ($n = 24$) had gender distinctions in the 2nd person. Moreover, gender distinctions in the 2nd person show strong genetic and areal biases: Siewierska (2004: 105) shows that gender distinctions in the 2nd person are found in few families and in few areas, mainly Afroasiatic languages, Ndu languages of Papua New Guinea, and in a few other outliers. Assuming that independent person markers are a significant diachronic source for bound person markers, it stands to reason that there are simply far fewer opportunities for gender distinctions in the 2SG to arise.

In fact, bound or dependent person markers usually show FEWER distinctions than independent markers, and "[t]he most common opposition completely absent in dependent forms as compared to their independent counterparts is gender" (Siewierska 2009: 113). As such, gender distinctions in bound 2SG bound markers are assumed to be even RARER than in independent 2SG person markers.

This is likely to be a part of any explanation for the rarity of zero-marked 2SG.F. Since (i) zero marking itself is rare, (ii) zero-marked 2SG is very rare, and (iii) gender distinctions in the 2SG are rare, it is only to be expected that zero-marked 2SG.F person markers would be exceedingly rare.

In fact, it is likely that the main pathway through which such situations could arise would be one in which an overt 2SG person marker in system with distinct overt person markers for all three persons, which also has a gender distinction in the 2SG, were lost as the result of a highly specific change, e. g., a sound change that incidentally targets the phonological material in the person marker.

6 A rarissimum: (In)stability

One of the alleged characteristics of rarissima is that they might be unstable, diachronically speaking (Wohlgemuth & Cysouw (eds.) 2010a). The Coptic zero

2SG.F marker corroborates this hypothesis. As Uljas (2009) shows, the situation in which the 2SG.F marker was zero, following a TAM formative identical to that which occurs prefixed to lexical noun phrase A/S – but crucially, not to most person indexes – was reanalyzed such that *-r(e)-* was reanalyzed as a 2SG.F person marker. Table 9 presents this process, taking the Aorist verb form (basically, a habitual) as exemplary. Admittedly, this is a highly idealized version of the situation. As Uljas (2009) shows, there is much synchronic variation between zero 2SG.F and overt 2SG.F markers, but a clear diachronic process can be identified, in which zero 2SG.F markers are replaced by overt ones involving *-r(e)-*.

Table 9: Reanalysis of TAM prefix as TAM prefix + person index.

	Lexical NP	**3SG.M**	**2SG.F**
Stage 1	*šare*-NP -V	*ša-f*-V	*šare*-Ø-V
	Reanalysis		
Stage 2	*šare*-NP -V	*ša-f*-V	*ša-re*-V

As Uljas (2009: 178–179) points out, this is a crosslinguistically unusual pathway of grammaticalization, leading from a sub-morphemic part of a TAM marker to a bound person marker or pronoun. What Uljas did not point out is that this is the converse of the phenomenon described by Watkins' Law, in which a segmentable morpheme (the person index) is reinterpreted as part of a morpheme (the base):

(8) a. Watkins' Law
BASE-INDEX ⇒ BASE-Ø
b. Coptic
BASE-Ø ⇒ BASE-INDEX

This instability itself is of interest. It is unlikely that zero-marked 2SG.F gave way to an overt 2SG.F marker because of the simple fact of its rarity across languages. It is hard to see how a crosslinguistically uncommon situation could affect the types of synchronic factors that are implicated in language change, unless one assumes that the linguistic knowledge of speakers of individual languages somehow captures crosslinguistic generalizations. Moreover, this kind of explanation runs into problems, since crosslinguistically uncommon structures can be stable over time (Harris 2010).

However, the process could be motivated by more general principles. One such motivation could be frequency, but it is hard to see what kind of prediction a token frequency-based account would make. Based on Bybee's "Preserving Effect", one could predict that structures with low token frequency would be more likely to undergo analogical change. However, there is little reason to think that the 2SG.F differs significantly from the 2SG.M in terms of token frequency, and there is similarly little reason to think that the 2SG in general had a low token frequency in actual everyday speech.

Another possibility is analogy, close to what Haspelmath (2014b) called "system pressure", in which "[r]ules of grammar generally target large classes of items, rather than individual expressions or small classes" (Haspelmath 2014b: 197). Here "system pressure" removes the anomaly of having three "unnatural" rules for a single verbal template. The allomorphy of the aorist TAM marker (the subscript numbers represent different allomorphs of the TAM marker) is:

(i)	TAM_1-lexical A/S-V	*šare-*	lexical NP subject
(ii)	TAM_1-V	*šare-*	2SG.F
(iii)	TAM_2-A/S index-V	*ša-*	all other person indexes

(9) Stage 1: Allomorphy

a. *šare-p-at^{h}êt* *kmš-te-sbô* *m-pe-f-eiôt*
AOR-DEF.M.SG-fool mock-DEF. F.SG-wisdom of-POSS-3SG.M-father
'The fool mocks the wisdom of his father.' (Proverbs 15:5)

b. *šare-Ø-tsio-Ø* *m-p-oeik* *mn-p-moou*
AOR-2SG.F-satisfy-2SG.F OBL-DEF.M.SG-bread with-DEF.M.SG-water
'You satisfy yourself with bread and water.' (Shenoute, in Leipoldt (ed.) 1908: 204)

c. *ša-f-ei* *ebol*
AOR-3SG.M-come out
'It comes out.' (Matthew 12:43)

As the result of reanalysis, this allomorphy is reduced, such that all bound person markers are preceded by the same allomorph of the Aorist TAM marker:

(i)	TAM_1-lexical A/S-V	*šare-*	lexical NP subject
(ii)	TAM_2-A/S index-V	*ša-*	all person indexes

Another way of looking at the relatively vague notion of "system pressure" is provided by Bybee's (1985, 2007) conception of the role of type frequency in

productivity. As Bybee points out, "when a construction is experienced with different items occupying a position, it enables the parsing of the construction" (Bybee 2007: 15), illustrating the claim with the following example:

> If *happiness* is learned by someone who knows no related words, there is no way to infer that it has two morphemes. If *happy* is also learned, then the learner could hypothesize that *-ness* is a suffix, but only if it occurs on other adjectives would its status as a suffix become established. Thus a certain degree of type frequency is needed to uncover the structure of words and phrases. In addition, a higher type frequency also gives a construction a stronger representation, making it more available or accessible for novel uses.

In the present case, it could be proposed that the higher type frequency of the shorter TAM allomorph (*ša-*), which occurs with all bound person markers except 2SG.F, provided the basis for reanalysis of the longer allomorph (*šare-*), which occurs only with the 2SG. F. The following examples repeat the sentences in (9) but are glossed differently to show the results of this process more clearly:

(10) Stage 2: Reduction of allomorphy and elimination of zero 2SG.F via reanalysis
   a. *šare-p-at^{h}êt* *kmš-te-sbô* *m-pe-f-eiôt*
      AOR-DEF.M.SG-fool mock-DEF.F.SG-wisdom GEN-POSS-3SG.M-father
      'The fool mocks the wisdom of his father.' (Proverbs 15:5)
   b. *ša-re-tsio-Ø* *m-p-oeik* *mn-p-moou*
      AOR-2SG.F-satisfy-2SG.F OBL-DEF.M.SG-bread with-DEF.M.SG-water
      'You satisfy yourself with bread and water.' (Shenoute, in Leipoldt (ed.) 1908: 204).
   c. *ša-f-ei* *ebol*
      AOR-3SG.M-come out
      'It comes out' (Matthew 12:43).

Further evidence for this analysis is provided by Uljas (2009), who shows that the overt 2SG.F marker *-r(e)-* spread to other verb forms via analogy or "affix pleonasm" (Haspelmath 1993). Example (11a) shows the older 2SG.F marker without *-r(e)-*, while (11b) shows an innovative 2SG.F marker:

(11) a. *ešče-mpe-Ø-souôn-t*
      COND-PST.NEG-2SG.F-know-1SG
      'If you knew me.' (Shenoute, in Leipoldt (ed.) 1908: 21)

b. *t-a-sône* *etbe-ou* *mp-r-čoo-s*
POSS.F.SG-1SG because-what PST.NEG-2SG.F-say-3SG.F
'My sister, why didn't you say it?' (Paese and Thekla 82: 6, cited in Uljas 2009: 180)

Note that this account, even though it eliminates zero 2SG.F A/S indexes, actually leaves zero 2SG.F in the system, as a P index (see 7b) above. As such, the reanalysis discussed here eliminates complexity in one part of the person-indexing system, i. e., in the allomorphy of TAM prefixes, but increases it in another, i. e., in the allomorphy of A/S vs. P indexes.

7 Conclusions

The evidence presented here indicates that zero-marked 2SG.F is a *rarissimum*, i. e., rare – perhaps unique – in the world's languages and rare – perhaps unique – within Afroasiatic. Of the four types of diachronic scenarios that contribute to the crosslinguistic rarity of linguistic features, the development of Coptic zero-marked 2SG.F shows all four. This feature developed in an ecology of a rare constellation of SOURCE construction features, i. e., a "Latin-type" system, one in which all singular persons are overtly marked; even rarer is the gender distinction in the 2SG. Moreover, few PATHWAYS of regular language change would lead to a situation in which there is a specifically zero-marked 2SG. Upon this rare situation operated a series of sound changes, one of them of a crosslinguistically rare TYPE, leading to the ultimate loss of the overt 2SG.F marker, creating a zero-marked 2SG.F. Altogether, the development of the zero-marked 2SG.F involved multiple STAGES that had to occur in a particular order on particular input structures. Strikingly, this situation was diachronically unstable, and a sub-morphemic part of preverbal TAM prefixes was reanalyzed as an overt 2SG.F marker *-r(e)-*. This leads to the possibility that yet another type of diachronic factor may be involved in cross-linguistic rarity, namely, the inherent instability of certain features. While this may smack of teleology, it may nonetheless be the case that certain features are indeed relatively unstable. For example, Blevins (2008) observes that the "three-way contrast between oral, weakly nasalized, and fully nasalized, documented for Palantla Chinantec is also extremely rare, occurring only in this language, where it appears to be unstable", noting that three-way length contrasts in Estonian, Saami, and Dinka may also be relatively unstable.

The case of the zero-marked 2SG.F in Coptic is yet another corroboration of the generalization that crosslinguistically rare linguistic structures are not ruled out by general principles of universal grammar, i. e., they are learnable and transmissible throughout generations. Rather, synchronically rare linguistic structures can and do arise as the result of regular processes of language change. However, even regular processes of language change operate on synchronic structures, and rare synchronic structures can be preserved or lost, or they can give rise to *rarissima*, which in turn can be preserved or lost. This article is intended to provide evidence for the latter, in hopes of one day having enough crosslinguistic data on pathways of change to ask what conditions favor the preservation or elimination of crosslinguistically rare features.

Acknowledgments: I would like to thank the many linguists (listed in Footnote 6) who answered my request for information on the LINGTYP mailing list (4 April 2015), as well as Juliette Blevins, Joan Bybee, and Michael Cysouw for answering specific questions about various matters. I apologize if I have left anyone out. I would also like to thank the anonymous reviewers of this article for their helpful and wise criticism. Of course, the responsibility for any mistakes in this article is my own. Thanks are also due to Ariel Shisha–Halevy, my first teacher of Coptic, who first pointed out to me the crosslinguistic rarity of zero-marked 2SG.F in classes during 1999–2001.

Abbreviations: 1/2/3 = 1st/2nd/3rd person; A = agent-like argument of canonical transitive verb; AOR = aorist; COND = conditional; DEF = definite; DEM = demonstrative; F = feminine; FUT = future; GEN = genitive; M = masculine; NEG = negation; OBL = oblique; OPT = optative; P = patient-like argument of canonical transitive verb; PL = plural; POSS = possessive; PST = past; S = single argument of canonical intransitive verb; SEQ = sequential verb form; SG = singular.

References

Amha, Azeb. 2012. Omotic. In Frajzyngier & Shay (eds.). 2012. 423–504.

Ariel, Mira. 2000. The development of person agreement markers: From pronouns to higher accessibility markers. In Michael Barlow & Suzanne Kemmer (eds.), *Usage-based models of language*, 197–260. Stanford, CA: CSLI Publications.

Bateman, Nicoleta. 2007. *A crosslinguistic investigation of palatalization*. San Diego, CA: University of California, San Diego doctoral dissertation.

Bhat, D. N. S. 1978. A general study of palatalization. In Joseph H. Greenberg (ed.), *Universals of human language*, Vol. 2: *Phonology*, 47–92. Stanford, CA: Stanford University Press.

Bickel, Balthasar, Alena Witzlack-Makarevich, Taras Zakharko & Giorgio Iemmolo. 2015. Exploring diachronic universals of agreement: Alignment patterns and zero marking across person categories. In Jürg Fleischer, Elisabeth Rieken & Paul Widmer (eds.), *Agreement from a diachronic perspective*, 29–52. Berlin: De Gruyter Mouton.

Blevins, Juliette. 2008. Natural and unnatural sound patterns: A pocket field guide. In Klaas Willems & Ludovic De Cuypere (eds.), *Naturalness and iconicity in language*, 121–148. Amsterdam: Benjamins.

Blevins, Juliette. 2009. Another universal bites the dust: Northwest Mekeo lacks coronal phonemes. *Oceanic Linguistics* 48. 264–273.

Brown, Cecil H., Eric W. Holman & Søren Wichmann. 2013. Sound correspondences in the world's languages. *Language* 89. 4–29.

Bybee, Joan L. 1985. *Morphology*. Amsterdam: Benjamins.

Bybee, Joan L. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.

Bybee, Joan L. 2007. *Frequency of use and the organization of language*. Oxford: Oxford University Press.

Bybee, Joan L. 2008. Formal universals as emergent phenomena: The origins of structure preservation. In Good (ed.), 108–121.

Caballero, Gabriela. 2008. *Choguita Rarámuri (Tarahumara) phonology and morphology*. Berkeley, CA: University of California, Berkeley doctoral dissertation.

Cysouw, Michael. 2003. *The paradigmatic structure of person marking*. Oxford: Oxford University Press.

Frajzyngier, Zygmunt. 2012. Chadic. In Frajzyngier & Shay (eds.). 2012. 236–341.

Frajzyngier, Zygmunt & Erin Shay (eds.). 2012. *The Afroasiatic languages*. Cambridge: Cambridge University Press.

Goldenberg, Gideon. 2013. *Semitic languages: Features, structures, relations, processes*. Oxford: Oxford University Press.

Good, Jeff (ed.). 2008. *Linguistic universals and language change*. Oxford: Oxford University Press.

Greenberg, Joseph. 1966a. Synchronic and diachronic universals in phonology. *Language* 42. 508–517.

Greenberg, Joseph. 1966b. *Language universals, with special reference to feature hierarchies*. Den Haag: Mouton.

Greenberg, Joseph. 1978. Diachrony, synchrony and language universals. In Joseph H. Greenberg, Charles A. Ferguson & Edith Moravcsik (eds.), *Universals of human language*, Vol. 1: *Method & theory*, 61–91. Stanford, CA: Stanford University.

Grossman, Eitan. 2014. No case before the verb in Coptic. In Grossman et al. (eds.). 2014. 203–225.

Grossman, Eitan & Martin Haspelmath. 2014. The Leipzig-Jerusalem transliteration of Coptic. In Grossman et al. (eds.). 2014. 145–153.

Grossman, Eitan, Guillaume Jacques & Anton Antonov. 2015. A cross-linguistic rarity in synchrony and diachrony: Adverbial subordinator prefixes exist. Under review.

Grossman, Eitan & Tonio Sebastian Richter. 2014. The Egyptian-Coptic language: Its setting in space, time and culture. In Grossman et al. (eds.). 2014. 69–101.

Grossman, Eitan, Martin Haspelmath & Tonio Sebastian Richter (eds.). 2014. *Egyptian-Coptic linguistics in typological perspective*. Berlin: De Gruyter Mouton.

Harris, Alice C. 2008. On the explanation of cross-linguistically unusual structures. In Good (ed.) 2008, 54–76.

Harris, Alice C. 2010. Explaining typologically unusual structures: The role of probability. In Wohlgemuth & Cysouw (eds.) 2010b, 92–103.
Haspelmath, Martin, 1993. The externalization of inflection. *Linguistics* 31. 279–309.
Haspelmath, Martin. 2008. Creating economical morphosyntactic patterns in language change. In Good (ed.) 2008, 185–214.
Haspelmath, Martin. 2014a. A grammatical overview of Egyptian and Coptic. In Grossman et al. (eds.) 2014, 103–143.
Haspelmath, Martin. 2014b. On system pressure competing with economic motivation. In Brian MacWhinney, Andrej L. Malchukov & Edith A. Moravcsik (eds.), *Competing motivations in grammar and usage*, 197–208. Oxford: Oxford University Press.
Kammerzell, Frank. 1998. The sounds of a dead language: Reconstructing Egyptian phonology. *Göttinger Beiträge zur Sprachwissenschaft* 1. 21–41.
Kammerzell, Frank. 2005. Old Egyptian and Pre-Old Egyptian: Tracing linguistic diversity in Archaic Egypt and the creation of the Egyptian language. In Stephan Seidlmayer (ed.), *Texte und Denkmäler des ägyptischen Alten Reiches*, 165–247. Berlin: Achet.
Kossmann, Maarten. 2012. Berber. In Frajzyngier & Shay (eds.). 2012. 18–101.
Kuhn, K. H. (ed.). 1956. *Letters and sermons of Besa*. Leuven: Durbecq.
Layton, Bentley. 2004. *A Coptic grammar*. Wiesbaden: Harrassowitz.
Leipoldt, Johannes (ed.). 1908. *Sinuthii Archimandritae vita et opera omnia*, Vol. 3. Leuven: Peeters.
Loprieno, Antonio. 1995. *Ancient Egyptian: A linguistic introduction*. Cambridge: Cambridge University Press.
Loprieno, Antonio & Matthias Müller. 2012. Ancient Egyptian and Coptic. In Frajzyngier & Shay (eds.). 2012. 102–144.
Mous, Maarten. 2012. Cushitic. In Frajzyngier & Shay (eds.). 2012. 342–422.
O'Brien, Jeremy. 2010. Perception and English t-glottalization. Los Angeles: University of Southern California qualifying paper. http://jeremypobrien.nfshost.com/papers/obrien_tglot.pdf
O'Brien, Jeremy. 2012. *An experimental approach to debuccalization and supplementary gestures*. Los Angeles: University of Southern California doctoral dissertation.
Peust, Carsten. 1999. *Egyptian phonology: An introduction to the phonology of a dead language*. Göttingen: Gutschmidt & Peust.
Polotsky, Hans-Jakob. 1960. The Coptic conjugation system. *Orientalia* 29. 392–422.
Siewierska, Anna. 2004. *Person*. Oxford: Oxford University Press.
Siewierska, Anna. 2009. Person asymmetries in zero expression and grammatical functions. In Franck Floricic (ed.), *Essais de linguistique generale et de typologie linguistique offerts au professeur Denis Creissels à l'occasion de ses 65 ans*, 425–438. Paris: Presses de l'École Normale Supérieure.
Siewierska, Anna. 2013. Third person zero of verbal person marking. In Matthew Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max-Planck-Institut für evolutionäre Anthropologie. http://wals.info/chapter/103 (accessed on 15 April 2015)
Sohn, Ho-Min. 2001. *The Korean language*. Cambridge: Cambridge University Press.
Telfer, Corey S. 2006. *Coronalization as assibilation*. Calgary: University of Calgary doctoral dissertation.
Uljas, Sami. 2009. The forms of the Coptic 2nd person feminine singular pronouns. *Zeitschrift für ägyptische Sprache und Altertumskunde* 136. 173–188.

Uspensky, Boris A. 1972. Subsystems in language, their interrrelations and their correlated universals. *Linguistics* 88. 53–71.
Watkins, Calvert. 1962. *Indo-European origins of the Celtic verb*, Vol. 1: *The sigmatic aorist*. Dublin: Institute for Advanced Studies.
Wohlgemuth, Jan & Michael Cysouw (eds.). 2010a. *Rara and rarissima: Documenting the fringes of linguistic diversity*. Berlin: De Gruyter Mouton.
Wohlgemuth, Jan & Michael Cysouw (eds.). 2010b. *Rethinking universals: How rarities affect linguistic theory*. Berlin: Mouton de Gruyter.
Worrell, William H. (ed.). 1942. *Coptic texts in the University of Michigan collection*. Ann Arbor, MI: University of Michigan Press.