

Werk

Titel: 1

Ort: Wiesbaden

Jahr: 2017

PURL: https://resolver.sub.uni-goettingen.de/purl?666048797_0021 | LOG_0004

Kontakt/Contact

[Digizeitschriften e.V.](#)
SUB Göttingen
Platz der Göttinger Sieben 1
37073 Göttingen

✉ info@digizeitschriften.de

Turkic Languages

Edited by
Lars Johanson

in cooperation with
Hendrik Boeschoten, Bernt Brendemoen,
Éva Á. Csató, Peter B. Golden, Tooru Hayasi, László Károly,
Astrid Menz, Dmitrij M. Nasilov, Irina Nevskaya,
Sumru A. Özsoy, Abdurishid Yakup

21 (2017) 1

Harrassowitz Verlag · Wiesbaden

The journal *TURKIC LANGUAGES* is devoted to linguistic Turcology. It addresses descriptive, comparative, synchronic, diachronic, theoretical and methodological problems of the study of Turkic languages including questions of genealogical, typological and areal relations, linguistic variation and language acquisition. The journal aims at presenting work of current interest on a variety of subjects and thus welcomes contributions on all aspects of Turkic linguistics. It contains articles, review articles, reviews, discussions, reports, and surveys of publications. It is published in one volume of two issues per year with approximately 300 pages.

Manuscripts for publication, books for review, and all correspondence concerning editorial matters should be sent to Prof. Dr. Dr. h.c. Lars Johanson, Turkic Languages, Department of Slavistics, Turcology and Circum-Baltic Studies, University of Mainz, 55099 Mainz, Germany. The email address johanson@uni-mainz.de may also be used for communication.

Books will be reviewed as circumstances permit. No publication received can be returned.

Subscription orders can be placed with booksellers and agencies. For further information please contact: Harrassowitz Verlag, 65174 Wiesbaden, Germany; fax: 49-611-530999; e-mail: verlag@harrassowitz.de.

© Otto Harrassowitz GmbH & Co. KG, Wiesbaden 2017

This journal, including all of its parts, is protected by copyright. Any use beyond the limits of copyright law without the permission of the publisher is forbidden and subject to penalty. This applies particularly to reproductions, translations, microfilms and storage and processing in electronic systems.

Printing and binding by Hubert & Co., Göttingen

Printed on permanent/durable paper

Printed in Germany

www.harrassowitz-verlag.de

ISSN 1431-4983

Contents

Turkic Languages, Volume 21, 2017, Number 1

Editorial note by Lars Johanson	1
---------------------------------------	---

Articles

Martine Robbeets: Transeurasian core structures in Turkic	3
Klára Agyagási: Kazan Tatar as a dominant language of the Volga-Kama region. A case study of lexical intermediation	36
Annette Herkenrath & Birsel Karakoç: Two questionable candidates for subordinatorship: <i>-mİşLIK</i> and <i>-mAzLIK</i> in Turkish	46
Gerjan van Schaaijk: Place nouns heading relative clauses with focal subjects.....	79
Sema Aslan Demir: Some remarks on viewpoint operators in Turkmen	107
Delio Vania Proverbio: On the phonetic unpredictability denoted by some Old Turkic texts written in Syriac script. Or the encoding ambiguity intrinsic to the Aramaic writing	115

Report

Éva Á. Csató: <i>Turkic Linguistics: The State of the Art</i> . Workshop at the University of Mainz in March 2016	152
----------------------------------------------------------------------------------------------------------------------------	-----

Editorial note

Turkic Languages, Volume 21, 2017, Number 1

This issue of *TURKIC LANGUAGES* presents a collection of papers on widely different topics.

In “Transeurasian core structures in Turkic”, Martine Robbeets examines to what extent proto-typical features of Turkic might go back to Proto-Transeurasian structures. After looking into a number of core structures of Japonic, Koreanic, Tungusic, Mongolic, and Turkic, the author asks how these languages may have come to share the features in question. Areal diffusion, universal tendencies, genealogical relationships, and combinations of these factors are considered as possible explanations.

In “On the phonetic unpredictability denoted by some Old Turkic texts written in Syriac script”, Delio Vania Proverbio deals with the encoding ambiguity intrinsic to the Aramaic writing system. The paper starts with an examination of an East Old Turkic manuscript in Syriac script and describes a number of graphotactic regularities found in the text. The author concludes that a rigorous formal account of the graphotactic constraints is only possible to a very limited extent because of the low complexity of the graphemic set in this offshoot of the Aramaic alphabet.

Klára Agyagási presents a paper on “Kazan Tatar as a dominant language of the Volga-Kama region” as a case study of lexical intermediation. After an overview of the emergence of the political, cultural, and linguistic dominance of Kazan Tatar in the region, the role of Islam in the Tatar culture is discussed. The main part of the paper is a historical areal study of the Arabic word *şabi* ‘boy, male child’ with a discussion of how this lexical element spread among the languages of the Kazan Tatar khanate: Arabic ⇒ Kazan Tatar ⇒ Viryal Chuvash ⇒ Mari dialects.

In “Some remarks on viewpoint operators in Turkmen”, Sema Aslan Demir discusses Turkmen postterminal markers. Although Turkmen is an Oghuz language, it shares some areal features with the Kipchak and Karluk branches of Turkic, which can be observed in the inventory of postterminal markers and interpreted as a deviation from the Oghuz typology. The study focuses on the postterminal markers {-An} vs. {-(I)pdIr} and their negative counterparts {-An däldir} vs. {-mAndIr}.

In “Two questionable candidates for subordinators: *-mİşLIK* and *-mAzLIK* in Turkish”, Annette Herkenrath and Birsal Karakoç investigate a number of morpho-syntactic, semantic, and functional features of two infrequently used complex verb forms in modern Turkish. On the basis of corpus-linguistic methods, the potential of these markers to serve as subordinators is discussed. It is concluded that they have the capacity to expand into clause-like structures, even though some contradictory patterns are found.

In “Place nouns heading relative clauses with focal subjects”, Gerjan van Schaaijk directs the attention to a construction that has puzzled many linguists: the

distribution of the so-called subject participle and the object participle in Turkish relative clauses. The author provides a pragmatic analysis of the object participle construction. He concludes that the subject of the relative clause is a non-referential noun phrase placed in preverbal focus position and that the head noun of the relative clause can without exception be interpreted as a noun denoting location. Such structures are presentative constructions providing new information, and they are related to existential constructions since they express “places where things happen”.

In the report “Turkic linguistics: The state of the art”, Éva Á. Csató gives an account of an international workshop organized at the University of Mainz in March 2016 on the occasion of the incorporation of the Department of Oriental Studies (Seminar für Orientkunde) into the newly established Department of Slavistics, Turcology, and Circum-Baltic Studies. All of the more than fifty participants had some relation to the Mainz chair of Turcology, as former doctoral students, research fellows, or project participants.

Lars Johanson

Transeurasian core structures in Turkic

Martine Robbeets

Robbeets, Martine 2017. Transeurasian core structures in Turkic. *Turkic Languages* 21, 3–35.

In this paper I investigate to what extent the core structures of Turkic might be inherited from those of proto-Transeurasian. The label “Transeurasian” is used for the grouping that includes the Japonic, Koreanic, Tungusic, Mongolic, and Turkic languages. The term “core structures” refers to a concentration of proto-typical linguistic features that delimit a group of languages vis-à-vis their neighbors. After evaluating 20 core structures of the Transeurasian languages, I argue how the languages may have come to share these features, considering as possible explanations areal diffusion, universal tendencies, genealogical relationships or an interaction of these factors.

Martine Robbeets, Max-Planck-Institut für Menschheitsgeschichte, Department of Linguistic and Cultural Evolution Kahlaische Strasse 10, room 022, DE-07745 Jena, Germany. E-mail: robbeets@shh.mpg.de

1. Introduction

After fifty-five years of dedication to the past and present of the Turkic languages, Johanson (2015) takes a moment’s pause to reflect upon the specific core structures of Turkic. Observing that Turkic shares most of its basic typological properties with the Mongolic, Tungusic, Koreanic and Japonic languages, Johanson (2015: 591) asks: “Are the core structures of Turkic common Transeurasian structures?” and thereby touches upon the still open question of whether the core structures of Turkic have been inherited from those of proto-Transeurasian.

In the present paper, I will delimit a number of core structures shared by the Transeurasian languages, paying attention to the extent to which these are reflected in the Turkic languages and weighing different historical motivations that may account for them. One of my aims is to show that there are a number of core structures that delimit the Transeurasian languages from the Uralic languages, with Turkic behaving like a proto-typical Transeurasian branch under Uralic influence. This view goes against Janhunen’s (2014: 13) claim that “it is also relevant to emphasize once more that speaking of “Altaic” instead of “Ural-Altaic” is a misconception, for there are no areal or typological features that would be specific only to “Altaic” without Uralic”.

The term “Transeurasian” refers to a large group of geographically adjacent languages in Northern Eurasia. They stretch from the Pacific in the East to the Baltic and the Mediterranean in the West and include up to five different linguistic families: Japonic, Koreanic, Tungusic, Mongolic, and Turkic (Johanson & Robbeets

2010: 1–2). I distinguish “Transeurasian” from the more traditional term “Altaic”, which I reserve for the linguistic grouping consisting of Tungusic, Mongolic and Turkic languages only.

The question of whether all similarities between the Transeurasian languages should be accounted for by language contact or whether some are the residue of a common ancestor is one of the most debated issues of historical comparative linguistics. However, reserving the term “core structures” for a concentration of prototypical linguistic features that delimit a group of languages from neighboring languages, independent of how these features developed historically, I will refrain from excluding features *a priori* because they may be the result of code-copying or inheritance. Only after evaluating 20 structural features shared across the Transeurasian languages, will I consider how the insights from the data are relevant for historical statements about the ways by which the languages may have come to share these features, considering as possible explanations areal diffusion, universal tendencies, genealogical relationship or an interaction of these factors.

As representatives of the contemporary varieties of Transeurasian, I will use Turkish (Turkic), Khalkha Mongolian (Mongolic), Evenki (Tungusic), Korean and Japanese (Japonic). For retrieving linguistic data underlying the feature values, I consult Göksel & Kerslake (2005) for Turkish; Janhunen (2012) for Khalkha Mongolian; Bulatova & Grenoble (1999) and Nedjalkov (1997) for Evenki; Martin (1992) and Sohn (1994) for Korean and, Martin (1988), Kaiser et al. (2001) and Iwasaki (2006) for Japanese. To allow for a diachronic perspective, I will supplement the contemporary languages with the oldest reliable historical varieties, consulting Erdal (2004) for Old Turkic, Street (1957), Weiers (1966) and Rybatzki (2003) for Middle Mongolian, Poppe (1954) for Written Mongolian, Gorelova (2002) for Manchu, Martin (1992) and Lee & Ramsey (2011) for Middle Korean and, Vovin (2005, 2009) and Frellesvig (2010) for Old Japanese. In order to delimit external boundaries, I have included Ainu (Ainuic), Nivkh (Amuric) and Rukai (Austronesian) as adjacent languages to the east, Mandarin Chinese (Sino-Tibetan) to the south, and Kolyma Yukaghir (Yukaghiric), Ket (Yenisseian) and Eastern Khanty (Uralic) to the north; see Figure 1. For retrieving linguistic data underlying the feature values, I use Gruzdeva (1998) for Nivkh; Maslova (2003) for Kolyma Yukaghir; Werner (1997), Vajda (2004) and Georg (2007) for Ket; Filchenko (2007) for Eastern Khanty; Li & Thompson (1989) for Mandarin; Zeitoun (2007) for Mantaurean Rukai; and Shibatani (1990) and Tamura (2000) for Ainu. Some of the data underlying this paper are drawn from Robbeets (2017), but the approach taken here differs from that paper in that it is focused on Turkic, the concept of core-grammar and the typological distinction between Uralic and Transeurasian.



Figure 1: The distribution of the contemporary Transeurasian languages and neighboring languages included in the sample (generated with WALS tools)

The paper is structured as follows. In Sections 2 to 5, I address different core structures of Transeurasian on the level of phonology, lexico-semantics, morphology and syntax respectively. In Section 6, I deal with core-patterns of grammaticalization shared by the Transeurasian languages. In Section 7, I summarize the presence of the 20 examined features in the selected languages, using a tabular overview and assess different reasons that could account for the correlations such as universal tendencies, copying and common ancestorship. Finally, I conclude this paper in Section 8.

2. Phonological core structures

CS 1. Presence of tongue root vowel harmony

The most general harmony phenomenon in Turkic is palatal harmony, which prompts all vowels within a domain to be exclusively front or back (e.g. Tk. *ip-ler* [rope-PL] ‘ropes’ vs. *pul-lar* [stamp-PL] ‘stamps’); see Johanson (1993). Most Turkic languages also employ a rounded vs. unrounded harmony, which causes neutralization of the roundness distinction in high vowels. Certain languages also apply this harmony to suffixes with non-high vowels. Palatal harmony is also found in most Uralic languages, such as in Khanty. Since the western Mongolic languages Oirat and Kalmuck display palatal harmony as well, it has been proposed that the original system of Mongolic harmony was palatal (Poppe 1955, Svantesson 1985). However Ko (2012) demonstrated that the original vowel harmony in Mongolic was in fact based on the opposition between the advanced vs. retracted position of the tongue root, rather than on a palatal contrast. He argued that the tongue root retraction system in Khalkha (e.g. *od-o:s* [feather-ABL] vs. *ɔd-ɔ:s* [star-ABL]) represents

retention rather than innovation. Furthermore, he supported the view that Tungusic vowel harmony is RTR based, as it is in Manchu and Evenki, and that the reduced vowel harmony in contemporary Korean derives from a tongue-root based system in Middle Korean. Contemporary Japanese and Ryukyuan languages do not have vowel harmony. In Old Japanese, however, there is a restriction on the shape of root morphemes, whereby the vowel o_2 cannot occur in a root together with the vowels u , o_1 or a . This phenomenon, known as Arisaka's law, has been taken as a kind of vowel harmony, but it has been excluded from comparisons with other Transeurasian languages because it does not reflect palatal harmony, the type of harmony which was attributed to the Transeurasian languages until recently (Frellesvig 2010: 44). In the light of the reconstruction of a 7-vowel system in proto-Japonic by Frellesvig and Whitman (2008), however, the harmony-like opposition in Old Japanese implies an underlying opposition between pJ $*i$, $*ə$ and $*u$, $*o$, $*a$, which does not exclude an original RTR based contrast. Therefore, it appears that RTR harmony may have been a core structure of proto-Transeurasian, with proto-Turkic perhaps having shifted to a palatal contrast under the influence of the neighboring Uralic languages. Whereas Vovin (1993: 50–51) and Bugaeva (2015: 465–467) reconstruct palatal harmony in Ainu, Shibatani (1990: 15) speculates that the Ainu opposition between o and u , a might have its origin in tongue root harmony, but here the indications are even weaker than in the Japanese case. According to Maslova (2003: 35), Yukaghir might be more appropriately described as having tongue root harmony than palatal harmony. Chukchi also displays tongue-root harmony. Although Gruzdeva (1998: 10) suggests that Nivkh leaves traces of height harmony, Janhunen (1981) and Ko, Whitman and Joseph (2014) interpret the system in terms of tongue root harmony. Cross-linguistically, tongue root harmony seems to be rather rare (Ko 2012: 11–12). A rough estimate would be that less than 10% of the world's languages have a tongue root vowel harmony system.

CS 2. Absence of r - in initial position

Across the Transeurasian languages, the consonant r - is not allowed to occur word-initially, except in copies (e.g. J *rajio*, K *latiwo*, Even *radio*, Khal. *radio*, Tk. *radyo* 'radio'). This is also true for Kolyma Yukaghir. Ket lacks a phoneme $/r/$ altogether. Although initial $*r$ - is not reconstructed for proto-Uralic, Khanty is atypical in this sense, e.g. *rayta* 'to drop, slide' and *räy* 'garbage'. Nivkh, Ainu, Mandarin and Rukai also have native words in initial r -. Outside Uralic and Transeurasian, a fair number of languages such as proto-Indo-European, Basque, some Melanesian languages, Efic (Niger-Congo), ancient Caucasian, Susu, Diyari (Australian), and Piro (Arakawan) lack r - in word initial position. A rough estimate would be that less than 30% of the world's languages lack r - in initial position

CS 3. Absence of initial velar nasal

In most Turkic languages, as well as in Mongolic languages and Korean, the velar nasal η - cannot appear in word-initial position. Japanese lacks a velar nasal phoneme. In the Tungusic languages, with exception of Manchu however, η - can appear word-initially, but generally it is restricted to a specific phonological environment, notably when it is followed by the sonorants n, r, l, m, y , e.g. Evk. *ηene-* ‘to go’, Ma. *genu-* ‘to go together’, Evk. *ηe:le-*, Ma. *gele-* ‘to fear’. According to Poppe (1964: 4) the initial velar nasal in Tungusic is the result of secondary assimilation of pTg * g -, which implies that originally * η - was absent in Tungusic as well. The assimilation was probably triggered by influence from languages in the Siberian area, such as Nivkh, which allow initial velar nasals (Anderson 2006). It is under the same influence that initial η became allowed in Dolgan (Turkic), e.g. *ηassa* ‘pipe’. In Khanty, Ket, Kolyma Yukaghir, Ainu and Mandarin η - does not occur in word-initial position. Rukai allows an initial velar nasal, e.g. *ηa|ai* ‘saliva’. In Anderson’s (2005: 42) sample of 468 languages, 69% lack an initial velar nasal. Among the languages of the world that have a velar nasal phoneme, as is the case with most Transeurasian languages, only 35% do not use it in word-initial position.

CS 4. Presence of voicing distinction for stops

Turkic, Mongolic and Tungusic languages share a voiced-voiceless opposition for stops, and voicing distinction can be reconstructed for proto-Transeurasian. In Contemporary and Middle Korean, stops display an opposition between lax (p), aspirated (ph) and tensed (p’). Even if the lax stops become lightly voiced between voiced sounds, there is no phonemic voicing distinction. The Japanese and Ryukyuan voicing distinction for stops is a secondary development, as voiced stops derive from prenasalized voiceless stops. Therefore, the ancestor of Japanese lacked voicing distinction. Khanty lacks voicing distinction for stops, a feature characteristic of proto-Uralic, although many contemporary Uralic languages have developed an original singleton-geminate contrast into a voicing distinction. Ket and Yukaghir display a voicing distinction, but languages on the northeast Pacific Coast such as Ainu, Nivkh and Chukchi do not. Mandarin, like Nivkh, has a distinction between aspirated and unaspirated stops, but lacks a voiced-voiceless opposition. Characteristic of most Austronesian languages, Rukai also displays voicing distinction for stops. In Maddieson’s (2005: 24) sample of 566 languages, 61% display a voicing distinction for stops.

3. Lexical and semantic core structures**CS 5. Preference for non-verbal strategy for (extra-family) verbal copies**

As far as the mechanisms of loan verb accommodation are concerned, most recipient languages can be categorized into two distinct groups: one where copied verbs arrive as verbs needing no formal accommodation, and one where they arrive as non-

verbs and need formal accommodation (Wohlgemuth 2009). The Turkic languages can be assigned to the second category because their copied verbs need formal accommodation by a suffix or a light verb; copying the English verb ‘to click (with a mouse)’, for instance, Turkish integrates the loanverb by applying either a verbalizer, e.g. Tk. *klik-le-* or a light verb, e.g. Tk. *klik et-*. A similar strategy is reported for Mongolic, southern Tungusic Korean and Japanese Khal. *zee-l-* << Mandarin *zhài* ‘borrow, lend’; Ud. *tancewa-la-* << Russian *tancewa-t* ‘to dance’; K *coking ha-*, J *zyogingu suru* ‘to jog’ << English *jog*; J *demo-r-* << English *demonstrate*. The northern Tungusic languages, however, prefer to borrow verbs through direct insertion, e.g. Evk. *vypolňaj-* << Russian *vypolňja-t* ‘to fulfill, carry out’. Since we have no information about verb borrowing in the historical stages, I mark them with +/- . In contrast to the Transeurasian languages, Ainu, Sinitic languages such as Mandarin, Uralic languages such as Khanty, and Austronesian languages such as Rukai show a strong preference for direct insertion (Wohlgemuth 2009: 158, 161; Tamura 2000: 267). Yukaghir and Nivkh did not integrate any recognizable verbal borrowings from Russian or other foreign languages into their lexicons. In Wohlgemuth’s (2009: 157) sample, 55% of languages worldwide are found to use direct insertion, while the remainder prefer non-verbal strategies such as indirect insertion and the light verb strategy.

CS 6. Presence of a two-way proximal-distal distinction in demonstrative pronouns

Although Old Turkic displays a two-way distinction in its demonstratives, i.e. OT *bo/bun-* ‘this’ vs. *ol/an-* ‘that’, many contemporary Turkic languages such as Turkish make a three-way distinction, e.g. Tk. *bu* ‘this’, *şu* ‘that’, *o* ‘that (over there)’. Demonstrative pronouns in earlier and contemporary varieties of Mongolic and Tungusic exhibit a proximal-distal distinction: MMo. *ene* ‘this’ vs. *tere* ‘that’, Khal. *e-* ‘this’ vs. *te-* ‘that’, Ma. *ere* ‘this’ vs. *tere* ‘that’ and Evk. *er(i)* ‘this’ vs. *tar(i)* ‘that’. Demonstrative pronouns in Contemporary and Middle Korean, however, show a proximal-mesial-distal opposition: K *i* ‘this’, *ku* ‘that’, *ce* ‘that over there’ and MK *i* ‘this’, *ku* ‘that’, *tye* ‘that over there’. This is also true for Contemporary Japanese: J *ko-* ‘this’, *so-* ‘that’, *a-* ‘that over there’. In contrast to most accounts of Old Japanese demonstratives, which posit a three-way contrast between OJ *ko*₂ ‘this’, *so*₂ ‘that’ and *ka* ‘that over there’, Frellesvig (2010: 139–142) argued that OJ *ka* was not a productive member of the demonstrative system and that pre-Old Japanese had a simple proximal-distal distinction. Similarly, in Yaeyama Ryukyuan the opposition between demonstratives is restricted to proximal *kuri* ‘this’ vs. distal *uri* ‘that’ (Aso 2015: 429). While Khanty distinguishes between proximal *timi* ‘this (here)’ and distal *tomi* ‘that (there)’, Yukaghir, Ket and Ainu have a three-way opposition, with each demonstrative pronoun denoting a different degree of proximity: Yukaghir *tiŋ* ‘this’ (proximal), *adiŋ* ~ *ediŋ* ‘that’ (mesial), *taŋ* ‘that’ (distal); Ket *tu-* ‘this, that’ (neutral), *ki-* ‘this, that’ (proximal); *qa-* ‘this, that’ (distal) and Ainu *ta an* ‘this’

(distal), *ne an* ‘that’ (mesial), *to an okai* ‘that over there’ (distal). Nivkh makes as many as five distinctions: *tyd* ‘this’ (near and visible), *hyd* ‘this, that’ (distant), *ad* ‘that’ (more distant and visible), *aixnt* ‘that’ (most distant), *kud* ‘that’ (absent).¹ Rukai distinguishes four demonstrative pronouns in terms of visibility and distance: *ina* ‘this’ (proximal), *ana* ‘that’ (mesial), *ona* ‘that over there’ (distal but visible), *dhona* ‘that over there’ (distal and invisible). Mandarin has a two-way distinction between proximal *zhè(ge)* ‘this’ and distal *nà(ge)* ‘that’, which developed from a three way-distinction in Classical Chinese between neutral, proximal and distal. In Diessel’s (2005: 170–173) sample of 234 languages, 54% exhibit a two-way distance contrast in demonstratives, while 38% exhibit a three-way contrast.

CS 7. Property words are verbally and nominally encoded such that some property words exhibit switched encoding

Cross-linguistically, adjectives have no prototypical encoding strategy of their own; they will align themselves either with verbs or with nominals. The large majority of property words in the contemporary Turkic languages are nominally encoded. Originally, in proto-Turkic, the encoding of property words appears to have been mixed because, at least in Old Turkic, both the nominal and the verbal strategy was used. There seems to be a tendency to apply the verbal strategy in the case of time-unstable properties such as OT *bädü-* ‘to be(come) big, great’, OT *isi-* ‘to be hot’, OT *kat-* ‘to be hard, firm, tough’, OT *kiz-* ‘to be red’, OT *tumlü-* ‘to be cold’, OT *tünči-* ‘to be(come) putrid, smell foul’, OT *us-* ‘to be thirsty’, OT *yeni-* ‘to be(come) light’, Otk. *tigra-* ‘to be tough’, Otk. *iglä-* ‘to be(come) ill’. Contemporary Turkic languages maintain a few reflexes of these verbal property words, for instance for ‘to be(come) big’, Tk. *büyü-*, SUig. *pezi-*, Az. *böyü-*, Khalaj *bidi-*, Tuva *bedi-*, Gag. *bü:-*, Karaim *büyü-* and for ‘to be(come) red, red-hot’, Turkm. *Giz-*, Tur. *kiz-*, Yak. *kī:s-*, but in the majority of cases, the earlier verbal property word has been derived with a deverbal noun suffix and became lexicalized as a nominal property word, e.g. Otk *bädük* ‘big, great; greatness’ and Tk. *büyük* ‘big’; see also Johanson (2006).

As in Turkic, most property words in Mongolic and Tungusic languages are nominally encoded, but some are verbally encoded (e.g. WMo. *qala-* ‘to be(come) warm’; Khal. *ayu:-* ‘be afraid’, Ma *aka-* ‘to be sad’, Evk. *buli:-* ‘to be sad’). In Japanese and Korean many property words are verbally encoded, but others are nominally encoded (J *sizuka*, OJ *siduka* ‘quiet’, J/OJ *taka-* ‘to be high’, K *kippu-* ‘to be happy’, *phikon ha-* ‘be tired’). In line with most Uralic languages, property words in Khanty are exclusively nominally encoded. This is also true for Ket. In Yukaghir, Ainu and Nivkh, however, property words are exclusively verbally encoded. As in the case of most Transeurasian languages, Ainu property verbs express both the

1 Note that this analysis deviates from the feature values given for distance contrasts in demonstratives by Diessel (2005: 170–173), since he marks Ainu, Nivkh, Yukaghir and Turkish as having a two-way contrast.

property and the process leading to the property, e.g. *pirka* ‘to be(come) good’. In line with Mainland Southeast Asian and Austronesian languages, Mandarin and Rukai use verbal encodings for property words.

Generally, the mixed encoding of adjectives in the Transeurasian languages is split in the sense that most property words have only a single encoding option. Turkish *güzel* ‘beautiful’, for instance, has nominal encoding and cannot be inflected as a verb. However, in Old Turkic, some doublets such as OT. *aç* ‘hungry’ / *aç-* ‘to be hungry’, OT. *keç* ‘late, slow’ / *keç-* ‘to be late, slow’, OT *köp* ‘abundant’ and OTk *köp-* ‘to swell, boil over’ and OTk. *karı* ‘old’ and *karı-* ‘to become old’ exhibit traces of original switching, whereby the same property word can have both nominal and verbal encoding; see Doerfer’s (1982: 104–112) list of so-called ‘Nomenverba’. Similar traces of switching are found in the other Transeurasian languages, especially in the earlier varieties, e.g. MMo. *bulqa* ‘hostile; hostility’ / *bulqa-* ‘to be hostile’; Ma. *jalu* ‘full’ / *jalu-* ‘to be full’, Ma. *sula* ‘loose, free’ / *sula-* ‘to be loose, be free’; MK *toso-* vs. MK *toso ho-* ‘to be warm’; OJ *taka* ‘high’ / *taka-* ‘to be high’, OJ *opo* ‘big’ / OJ *opo-* ‘to be big’. None of the neighboring languages exhibits such behavior. In Stassen’s (2005b: 478–481) sample of 386 languages, 27% have mixed encoding in predicative adjectives. Logically, the proportion of languages exhibiting mixed and switched encoding will be lower.

CS 8. Partial emphatic reduplication of nominal property words

Partial emphatic reduplication is a phenomenon whereby the first consonant (if present) and vowel of a nominal property word are repeated with the addition of another consonant to indicate the presence of the property to the utmost degree. In Turkic languages the phenomenon is widespread, e.g. Tk. *bem-beyaz* ‘snow white’, *up-uzun* ‘extremely long’, OT *kap-kara* ‘quite black’. Whaley and Li (2000) found that it is also recurrent in Mongolic and Tungusic, e.g. Khal. *xob-xoldu*: ‘frozen through’, WMo. *ub-ulaγan* ‘completely red’, Evk. *ab-aya* ‘very good’. I am unable to find examples in Manchu, but the phenomenon is present in Sibe, a currently spoken variety of Manchu, e.g. *fak-farxun* ‘extremely dark’. In Tungusic, emphatic reduplication is restricted to Sibe, Kile-Nanai, Solon Evenki and Oroqen, i.e. the languages spoken on Chinese soil, which have been under strong influence from the Mongolic languages Khalka and Dagur. On the basis of this distribution, and because the greatest flexibility (in terms of both the number of reduplicated words and the type of concepts they denote) is found in Turkic, Whaley and Li (2000: 358) argued for a diffusion of the feature from Turkic to Mongolic to Tungusic. Japanese, Korean and the neighboring languages under examination do not display partial emphatic reduplication. In Rukai, however, descriptive verbs are partially reduplicated in comparative constructions (see CS 17).

4. Morphological core structures

CS 9. Inflectional morphology is predominantly suffixing

Bound units in Turkic are postponed; i.e. they are suffixes rather than prefixes or infixes. Across the strongly suffixing Transeurasian languages, prefixation is rare and is restricted to derivational morphology, such as the partial emphatic reduplication in Core Structure 8 and some derivational prefixes in Korean (e.g. K *yel-* ‘young, new’ in *yel-cwungi* ‘a chick out of its shell’) and in Japanese (e.g. *ma-* intensive in *ma-siro* ‘snow white’). As is the case for most Uralic languages, Khanty is strongly suffixing, as is Yukaghir. Nivk is considered to be weakly suffixing. In Ket, nominal inflectional morphology is strongly suffixing, whereas verb inflection is predominantly prefixing. In Ainu and Rukai, inflection makes use of both prefixes and suffixes. Probably due to Transeurasian influence, Mandarin is hard to assign unequivocally to either the isolating or weakly suffixing type, but Sinitic varieties in general tend towards the isolating pole. In Dryer’s (2005a: 110–113) sample of 894 languages, 43% are strongly suffixing.

CS 10. Absence of obligatory numeral classifiers

Although in Turkic and Mongolic some nouns of low countability may be accompanied by a unit of measure by means of which they can be counted, e.g. Tk. *sekiz bardak su* (8 glass water) ‘eight glasses of water’, OT *yeti tutum talkan* (7 handful parched.grain) ‘seven handfuls of parched grain’, Khal. *gourben debter nom* (3 volume book) ‘3 volumes of books’, these languages do not make use of sortal numeral classifiers. The same is true for the Tungusic languages, except Manchu. Under Chinese influence, Manchu has developed about 70 sortal numeral classifiers, such as *fesin*, which is used for objects equipped with a handle, e.g. *ilan fesin loho* (3 CLAS sword) ‘three swords’. However, the use of these classifiers is not obligatory in Manchu. *Loho ilan* (sword 3) ‘three swords’, for instance, is equally possible. Whereas the standard pattern in Middle Korean was to modify a noun with a preposed numeral, e.g. *twu kalh* (2 knife) ‘two knives’, under Chinese influence Contemporary Korean increased its use of classifiers, e.g. *pus se:k calwu* (writing.brush three CLAS) in which *calwu* denotes long objects with handles. However, the original pattern surfaces in expressions such as K *twu nala* ‘two countries’ and the use of classifiers remains optional in Korean, e.g. *kalh hana-ka issta* (knife one-NOM be.present) ‘there is one knife’.² While there is an extensive list of obligatory classifiers in Contemporary Japanese, e.g. *enpitu san-bon* (pencil three-CLASS) ‘three pencils’, the use of classifiers is much less developed and is not obligatory in Old Japanese, where Chinese influence is restricted to a minimum. Numeral classifiers are absent in Uralic languages such as Khanty, as well as in Yukaghir and Ket.

2 Note that my evaluation differs from Gil’s (2005: 228–229) interpretation that Korean has obligatory numeral classifiers.

Ainu and Nivkh make use of a set of obligatory classifiers. The obligatory use of classifiers is a widespread feature shared by Mandarin and the languages of South-east Asia, but the use of classifiers in Classical Chinese was the exception rather than the rule. In Rukai, the use of classifiers is optional in the sense that it uses a set of unaffixed numerals without classifiers as well as a set of bound numerals, which combine with five different sortal classifiers to form verbs. In Gil's (2005: 226–229) sample of 400 languages, 80% lack obligatory numeral classifiers.

CS 11. Presence of *mi-Ti* opposition in first vs. second singular personal pronouns

Nichols (2012) observes that *m-T* pronominal paradigms with first person labial nasal *m* and second person apical or palatal obstruent *t*, *c*, *s*, etc. are much more common in northern Eurasia than elsewhere in the world. Janhunen (2013: 213) adds that there is a smaller group of *mi-Ti* languages extending from Uralic in the west, to Turkic, Mongolic and Tungusic in the east, and Yukaghir in the north, in which not only the initial consonant but also the root vowel of the singular stems shows a basic similarity in that it contains a non-low unrounded front vowel *i* or *e*. Although *m* is absent in the nominative first person singular in the Turkic, Mongolic and Tungusic languages, e.g. Tk *ben*, OT *ben*, Khal. *bii*, MMo. *bi*, Ma. *bi*, Evk. *bi:*, it has developed in oblique forms in assimilation to the nasal oblique suffix *-n*, e.g. OT *min-*, Khal. *min-ii* [GEN], MMo. *mi-nu* [GEN], Ma. *min-*, Evk. *min-*. The second person singular forms all reflect a voiceless dental T, e.g. Tk. *sen*, OT *sen*, Khal. *cii*, MMo. *ci*, Ma. *si*, Evk. *si:*. The Korean pronouns are first singular K/MK *na* and second singular K/MK *ne* among others. In Japanese, J *watasi* and OJ *wa* among others are used in the first singular, while a variety of contemporary pronouns and OJ *na* are used in the second singular. Although the proto-Uralic first and second singular pronouns **mun* and **tun* reflect not a *mi-Ti* though still an *m-T* distinction (Janhunen 1982: 35), Khanty is deviant in having first singular *mă* and second singular *nöŋ*. In Yukaghir, however, the *mi-Ti* opposition is present in first singular *met* vs. second singular *tet*. In Nivkh, the distinction is absent in the singular pronouns, first person *n'i* vs. second person *či*, but it is present in the opposition between the first plural inclusive *mir/mer* and the second plural pronoun *čij*. The opposition is not found in Ket, Ainu, Chinese and Rukai. In Nichols and Peterson's (2005: 546–551) sample of 230 languages, 13% display an *m-T* opposition in first vs. second person pronouns. Logically, languages reflecting a *mi-Ti* opposition will represent an even smaller proportion.

CS 12. Formation of a secondary nasal oblique stem in personal pronouns

In most contemporary Turkic languages, the nominative and oblique forms of the personal pronouns have merged, e.g. Tk. *ben* for the first singular nominative and oblique, but in Old Turkic the first singular nominative *ben* is distinguished from the oblique stem *min-*, which can be derived from an original pTk **bi-n-* [1SG-OBL-].

Similarly, the Mongolic and Tungusic languages derive oblique pronominal stems from the nominative roots through a nasal suffix, for instance in the first person plural pronouns MMo. *ba* [NOM] vs. *man-* [OBL] and Khal. *bid* [NOM] vs. *bidn-* [OBL] and in the first person singular pronouns Ma. *bi* [NOM] vs. *min-* [OBL], Evk. *bi:* [NOM] vs. *min-* [OBL]. There are no oblique pronominal stems in Contemporary Japanese, but in Old Japanese traces remain of an oblique nasal suffix in some case forms, e.g. in the Eastern OJ first person singular dative *wa-nu-ni* in alternation with Western OJ *wa-ni*. Vovin (2005: 229–230) further found that an original Japonic pronominal oblique **-n-* is well supported by Northern Ryukyuan dialects where the first person pronoun uses *waa-* as the nominative and genitive base and extended *waN-* in the oblique cases. Hence, with the exception of Korean, the Transeurasian languages share a tendency of forming a secondary oblique stem of the personal pronouns by means of a suffix, which can be identified phonologically as the dental nasal *-n-*. The oblique nasal suffix is an important element in the Uralic pronominal paradigm as well, e.g. the Khanty first person pronoun *mä* [NOM] vs. *män-* [OBL]. Ket, Yukaghir, Ainu and Mandarin, however, do not derive secondary oblique stems. The third person singular pronoun in Nivkh has both regular and suppletive case forms, e.g. *if-øn* [3SG-NOM] vs. *if-toX* ~ *e-rx* [3SG-DAT/ADD], but here the oblique form is not derived from the nominative base. Rukai personal pronouns have different shapes for nominative, topic, genitive and oblique cases, e.g. the first person singular *-lrao* [NOM], *ilrae* [TOP], *-li* [GEN] vs. *-iae* [OBL], in which the oblique seems to be formally derived from the nominative base by means of the same *i-* ...-*e* marking as in the topic form.

5. Syntactic core structures

CS 13. Dependent-marking of clause arguments

In the clause, the verb is the head and the arguments are dependents. Morphological marking, reflecting the syntactic relations in the clause, may be located on the head, on the dependent, on both, or on neither. Even if they may have subject-verb agreement on the verb, the Turkic languages are dependent-marking because they tend to mark agreement and case government more on dependents than on verbs; see the subject-verb and case agreement in the Turkish example in (1). Most Mongolic languages are strongly dependent-marking, as they mark case and lack verb agreement; see the Khalkha example in (2). Tungusic languages have case and verbal agreement with the subject; see the Evenki example in (3). Having case and lacking verb agreement, Korean and Japanese are strongly dependent marking, as illustrated in (4) and (5).

- (1) Turkish
Bu ev-i Ahmet-e yap-tı-m.
 this house-ACC Ahmet-DAT make-PF-1SG
 ‘I built this house for Ahmet.’ (Göksel & Kerslake 2005: 146)
- (2) Khalkha
Öwgön-iig ger-t-ee ury-jee.
 old.man-ACC home-DAT-REFL invite-PF
 ‘He [the tiger] invited the old man to his home.’ (Janhunen 2012: 296)
- (3) Evenki
Nungan eri gule-ve o:-ra-n.
 he this house-ACC make-PF-3SG
 ‘He built this house.’ (Nedjalkov 1997: 83)
- (4) Korean
Minca-nun halapeci-kkey ton-ul tuli-ess-eyo.
 Minca-TOP grandfather-DAT money-ACC give-PST-POL
 ‘Minca gave her grandfather some money.’ (Sohn 1994: 84)
- (5) Japanese
Miki-ga Yamada sensei-ni tegami-o mise-ru.
 Miki-NOM Yamada teacher-DAT letter-ACC show-NPST
 ‘Miki shows the letter to professor Yamada.’ (Iwasaki 2006: 122)

Having subject-verb and case agreement, Khanty is weakly dependent-marking. It illustrates the tendency of gradually changing from double-marking to dependent marking in Uralic languages (Nichols 1986: 89). Yukagir and Mandarin Chinese are dependent-marking. However, Nivkh, Ainu and Ket, the isolates in Eurasia, are all head-marking, as is Rukai. In Nichols and Bickel’s (2005a: 98–101) sample of 235 languages, 27% are dependent-marking in the clause.

CS 14. Dependent-marking in possessive noun phrases

In possessive noun phrases, the possessed noun is the head and the possessor is the dependent. Morphological marking, reflecting the syntactic relation between the possessor and the possessed, may be located on the head, on the dependent, on both, or on neither. The Turkic languages are double marking, e.g. *Ali-nin oğl-u* [Ali-GEN son-3SG.POSS] ‘Ali’s son’, *oda-nin kapı-sı* [room-GEN door-3SG.POSS] ‘the door of the room’. Old Turkic is commonly double-marking, e.g. OT *ton-nuñ bit-i* (clothe-GEN louse-3SG.POSS) ‘clothes’ louse’, but there are also cases in which it is head-marked, e.g. *köl tegin atı-sı* [Köl Tegin nephew-3SG.POSS] ‘Köl Tegin’s nephew’ or unmarked, e.g. *balık kapag* [city gate] ‘city gate’. The Mongolic languages are dependent marking, e.g. Khal. *min-ii eej* [I-GEN mother] ‘my mother’, WMo. *šabi-yin nom* [pupil-GEN book] ‘the pupil’s book’. Except for Manchu, where possessive

relations are dependent-marked, e.g. *ama i bo*: [father GEN house] ‘father’s house’, genitive case is absent in most Tungusic languages, since possessive relations are head-marked, e.g. Even *svinija ulrə-n* [swine meat-3SG.POSS] ‘swine’s meat, pork’. Both Contemporary and Middle Korean are dependent-marking, e.g. K *na-uy yenphil* [I-GEN pencil] ‘my pencil’, MK *mo-l-oy hyang* [horse-GEN scent] ‘the scent of horses’. This is also true for Contemporary and Old Japanese, e.g. J *anata no atama* [you-GEN head] ‘your head’ OJ *Yamato-no kuni* [Yamato-GEN land] ‘the land of Yamato’. Proto-Uralic was originally head-marking like Khanty, e.g. Khan. *qul-əm* [fish-2SG.SG] ‘your fish’.³

The loss of head-marking patterns and extension of dependent-marked ones in western Uralic languages results from the influence of the Indo-European languages to the west. Given the fact that Turkic is double marking in the possessive noun phrase, while it has the proto-typical Transeurasian pattern of dependent marking in the clause, it is likely that it acquired double marking under the influence of the proto-typical Uralic pattern. This is supported by the observation that Turkic nominal possessive suffixes precede case suffixes (Johanson 2002: 22–23), while possessive suffixes were word final in proto-Uralic (Nichols 1986). This indicates that the Turkic possessive marker lies more toward the derivational than the inflectional end of the continuum, and that it functions as a semantic component of the noun rather than an agreement marker as in Uralic. Yukaghir and Mandarin are dependent marking, e.g. Yuk. *tude kerewe-d ugurce* [3SG cow-GEN leg] ‘the legs of his cow’ and Man. *bāba de máma* [father GEN mother] ‘the mother of father’. However, the Eurasian isolates are prototypically headmarking, e.g. Ainu *eci-siki-hi* [2PL-eye-GEN] ‘your eyes’, Nivkh *vit-yanj* [2SG-book] ‘your book’ and Ket *b-a:m* [1SG-mother] ‘my mother’. Rukai is head-marking as well, e.g. Ruk. *tolropongo-ni dhipolo* [hat-2SG.GEN Dhipolo] ‘Dhipolo’s hat’. In Nichols and Bickel’s (2005b: 98–101) sample of 235 languages, 42% are dependent-marking in the possessive noun phrase.

CS 15. Extensive use of converb

Converbs, also known as gerunds or adverbial participles, can be defined as nonfinite verb forms whose main function is to mark adverbial subordination (Haspelmath 1995: 3). Originally coined by the Altaic scholar Ramstedt, the term converb was adopted from Transeurasian linguistics to denote a cross-linguistic category. All Transeurasian languages are converb-prominent languages in the sense that they use converbs rather than adverbial subordinators as found in many European languages; see the examples below.

3 In Khanty, the possessive suffix makes reference to the number and person of the possessor, as well as to the number of the entity possessed (Filchenko 2008: 80).

- (6) Turkish
Ali gel-ince şaşır-dı
 Ali come-CONV be.surprised-PST3SG
 ‘When Ali came, he was surprised.’ (Johanson 1995: 314)
- (7) Khalkha
Ger-ees-ee gar-aad
 house-ABL-REFL exit-PFV.CONV
deuc-en jil-iin daraa ol-d-lao.
 forty-ADN year-GEN after find-PASS-FIN
 ‘She went away from home and was found forty years later.’ (Janhunen 2012: 280)
- (8) Evenki
əmə-mme:n iri-l-i-m.
 come-CONV cook-INCH-NPST-1SG
 ‘As soon as I arrive, I will start cooking.’ (Bulatova & Grenoble 1999: 44)
- (9) Korean
Kiho-nun nol-ko ca-ss-eyo.
 Kiho-TOP play-CONV sleep-PST-POL
 ‘Kiho played and then slept.’ (Sohn 2009: 300)
- (10) Japanese
Taroo-ga bangohan-o tabe-te furo-ni hai-ta.
 Taroo-NOM dinner-ACC eat-CONV bath-DAT enter-PST
 ‘Taroo took a bath after he ate dinner.’ (Alpatov & Podlesskaya 1995: 473)

Although the Uralic languages are characterized by extensive use of converbs, Khanty is rather atypical in this sense because it has only a single converb in *-min*, which is the least frequent nonfinite verb form. Yukaghir and Nivkh also use a variety of converbs to link clauses. Ainu, however, employs subordinating conjunctions. Ket has no converbs or serial verb constructions of any kind. In Mandarin, verbs or verbal phrases are merely juxtaposed, the relation between the items being largely unmarked. Rukai marks adverbial subordination through a variety of means such as subordinating conjunctions, changes in word order and nominalized verb forms.

CS 16. Use of locative existential constructions to encode predicative possession

The Transeurasian languages show a clear preference for expressing the concept ‘X has Y’ on the basis of an existential sentence, whereby the possessed noun phrase functions as the grammatical subject of the ‘exist’-predicate, while the possessor noun phrase is in a dative-locative case form. Although locative possessive constructions were standard in Old Turkic, Turkish uses genitive existential sentences as well as locative existential sentences. ‘I have a book’, for instance, can be expressed by *Ben-de bir kitab var* [I-LOC a book exist] or by *Ben-im bir kitab-ım var*

[I-GEN a book-1SG.POSS exist]. Middle Mongolian and Khalkha make use of either a conjunctive possessive which construes the possessor noun phrase as the grammatical subject of the copula and marks the possessed with the comitative *-tai*, e.g. Khalkha *Bi nom-tai bai-n'* [I book-COM be-DUR], or a locative possessive, e.g. *Naded nom bai-n'* [I-DAT book be-DUR]. As is the case for most Tungusic languages, Manchu and Evenki employ locative existential constructions, e.g. Evk. *Min-du: kniga bisi-n* [I-DAT book be-3SG]; Ma. *Min-de bithe bi* [I-DAT book be]. Korean uses a locative existential construction, e.g. K *Na-hanthey chayk-i issta* [I-LOC book-NOM exist], but the possessor can also be construed as the topic of the noun phrase, e.g. *Na-nun chayk-i issta* [I-TOP book-NOM exist]. This is also true for Japanese, e.g. *Watashi-ni hon-ga aru* [I-DAT book-NOM exist] and *Watashi-wa hon-ga aru* [I-TOP book-NOM exist]. Topic possessives may have developed under the influence of Chinese, since they represent the standard strategy in Mandarin. Among the strategies used to encode predicative possession in the Uralic languages, we find locative possession, as in Finnish and Hungarian, genitive possession, as in Nenets, and possession encoded by a transitive verb 'to have', as in Khanty. Whereas Yukaghir employs a conjunctive possessive and Ainu a 'have'-possessive, Ket and Nivkh use locational possessives. Although many Austronesian languages employ topic possessives, Rukai makes use of locative and genitive possessive constructions. In Stassen's (2005a: 474–477) sample of 240 languages, 20% use a locative existential construction to encode predicative possession.

CS 17. Use of the ablative case form to encode predicative comparison

The Transeurasian languages all form comparative constructions in which the standard noun phrase is constructed in the ablative case form, e.g. Tk. *bu araba-dan daha büyük* [this car-ABL more big] 'bigger than this car', OT *barča-da üzä-räk* [everything-ABL high-COMP] 'higher than anything else', Khal. *ene xun-ees iluu* [this person-ABL good] 'better than this person', MMo. *qola-sa qola* [far-ABL far] 'farther than far', Evk. *oron-duk gugda-tmar* [deer-ABL tall-COMP], Ma. *ere niyalma ci sain* [this person ABL good] 'better than this person', OJ *ware-yo₁ri mo₂ mantusi-ki₁ pi₁to₂* [I-ABL PT be.poor-ADN person] 'people poorer than me' and J *chikyu:-yori omoi* (globe-ABL be.heavy) 'heavier than the globe'. In literary Korean, the ablative marker *eyse* 'from' can be used in comparative constructions, e.g. K *i eyse te khu-n salang* [this ABL more be.big-ADN love] 'a greater love than this', but it is more common to use a comparative particle *pota* 'than', e.g. K *kicha pota ppaluta* [train PT be.fast] 'faster than a train', MK *nyey pwota thak.wel hota* [past PT superior be] 'superior to the past'. The Uralic languages differ from one another with regard to comparative constructions; languages to the west, such as Finnish and Hungarian, use more particle comparatives as in European languages, languages to the east, such as Nenets and Udmurt, mark the comparative standard with the ablative case ending, as in the Transeurasian languages. In Khanty, the marker of comparison is a postposition *niŋə* 'since, from', which has ablative-like semantics but differs from

the standard ablative case ending *-oy* or the ablative-relative ending *-i*. Yukaghir and Ket mark the comparative standard with the ablative case ending. In Nivkh, the comparative suffix *-yk* is traditionally considered a separate case form, as there is no evidence to relate it to the formally similar locative-ablative suffix *-(u)ye*; *-(u)x* (Gruzdeva p.c.). Ainu forms comparative constructions by means of the particle *kasuno* ‘than’. In comparative constructions in Mandarin, the standard noun phrase is constructed as the direct object of a verb ‘to exceed’. In Rukai, a comparative construction is formed through partial reduplication (CVV) of the descriptive verb stem. In Stassen’s (2005c: 490–493) sample of 167 languages, 47% use locational comparatives, but the proportion of languages that specifically use the ablative case form to encode predicative comparison is logically expected to be lower.

6. Core-grammaticalizations

CS 18. Direct insubordination

The Transeurasian languages display a recurrent tendency to reanalyze non-finite suffixes as finite ones without the omission of a specific matrix predicate, a tendency which I call “direct insubordination” (Robbeets 2015, 2016). Comparative evidence indicates that these markers originated as deverbal noun suffixes, marking a derivational process at the lexical level, which were then extended to function as (ad)nominalizers in dependent clauses at the syntactic level, and eventually—through a pragmatic role in discourse—were extended still further to mark finite forms in independent clauses. For instance, deverbal noun suffixes such as OTk *-(A)r* in OTk. *tug-* ‘to be born, to rise (of sun) (intr.)’ → *tugar* ‘sunrise, east’; MMo. *-m* in MMo. *quri-* ‘to come together (intr.)’ → *qurim* ‘feast’; Ma. *-rA* in *mute-* ‘to be able’ → *mutere* ‘ability’; MK *-(u/o)m* in *yel-* ‘to bear’ → *yelum* ‘fruit’ and OJ *-sa* in *naga-* ‘to be long’ → *nagasa* ‘length’ develop over intermediate stages of clausal nominalizers and relativizers into finite suffixes, as illustrated in examples (11) to (15).

(11) Old Turkic

<i>Ölüm-tä</i>	<i>oz-upan</i>	<i>ögir-ä</i>	<i>savin-ü</i>	<i>yorï-r.</i>
death-ABL	escape-CONV	rejoice-CONV	be.happy-CONV	go.on-FIN

‘Having been saved from death it happily goes on with its life.’ (Erdal 2004: 325)

(12) Middle Mongolian

<i>Udurit-</i>	<i>basu</i>	<i>ber</i>	<i>ulu busire-m.</i>
guide-COND	PT	NEG	believe-FIN

‘Even if you guide them, they don’t believe.’ (Weiers 1966: 144)

- (13) Manchu
Si nene-me isinji-ci uthai sin-de bu-re.
 you be.first-CONV come-CONV at.once you-DAT give-FIN
 ‘If you come first, I shall give [it] to you straight away.’ (Gorelova 2002: 256)
- (14) Korean
Onul-un swuep-i eps-um.
 today-TOP class-NOM not.exist-FIN
 ‘No class today.’
- (15) Old Japanese
Punapi₁to₂-wo mi₁-ru-ga to₂mo₂si-sa.
 boat.people-ACC see-NML-GEN be.envidable-FIN
 ‘How enviable it is to see the boat-people!’ (Wrona 2008: 206)

The Uralic languages also display a recurrent tendency toward direct insubordination. Deverbal noun suffixes such as proto-Uralic **-k*, **-pÁ*, **-mə* and **-sÁ* are thought to have developed into finite markers for present-day (**-k*, **-pÁ*) and past (**-mə*, **-sÁ*) tense, either in proto-Uralic or after the separation of the daughter languages (Collinder 1965: 110–115; Janhunen 1982: 36–37). Eastern Khanty preserves only a faint trace of this development since the finite form of the negative verb can be marked with the perfective participle *-əm*, as illustrated in example (16).

- (16) Eastern Khanty
Məta wəjəj lök ənt-im.
 some animal track NEG-FIN
 ‘There is not a single animal track.’ (Filchenko 2007: 429)

In Nivkh the deverbal action noun and infinitive suffix *-d'* has developed into a finite form *-d'*, as illustrated in example (17). However, rather than being a case of “direct insubordination”, Gruzdeva (2016: 196) attributes this development to the lexicalization of a modally marked form of the copular verb *ha-* ‘do so’. When the modally marked copula was lexicalized into a modal particle, *-d'* was reanalyzed as a finite form.

- (17) Nivkh
If hum-d' hyjm-d'.
 he live-NML know-FIN
 ‘He knows the living one/(his) life.’ (Malchukov 2013: 200)

As in Nivkh, clausal nominalization in construction with a copula is the main source for developing new finite constructions in Yukaghir and Mandarin (Yap & Matthews 2008: 20, Malchukov 2013: 192–195). Ket displays yet another strategy for developing finite markers, namely to reduce the matrix predicate to an affix on the

former dependent verb (Malchukov 2013: 196–197). In Ainu, deverbal noun suffixes appear to function as both derivational suffixes and syntactic clausal nominalizers, but there is no indication that they have developed into finite endings. Ainu lacks other nonfinite markers such as participial or converb affixes that could be open to developing into finite markers. Similarly, Rukai does not exhibit traces of direct insubordination.

CS 19. Grammaticalization from negative verb to verbal negator over a construction comprising an inflected negative auxiliary and an invariant lexical verb

In Turkic, we find indications that the verbal negative suffix OT *-mA-* originated as an inflecting negative auxiliary verb plus an invariant derivationally complex lexical verb. These include the stem-internal position of negation in Turkic, sandwiched between derivation and inflection; the occurrence of the adnominal negative suffix OT *-mA-z*, which seems to have developed from an aorist *-r* in unaccented position; the observation that Chuvash *mar* acts as an independent negative verb, taking an nominal argument in petrified constructions such as (18); and the analysis of the Chuvash optative first person singular as an auxiliary negative form, whereby inflection has shifted to the lexical verb, e.g. *vula-m mar* [read-1SG NEG] ‘I will not read’. In addition, we can find independent lexical cognates for pTk **ma-* ‘not to exist’ in the other Transeurasian languages (Robbeets 2015: 203–204).

- (18) Chuvash
- | | |
|---------------------------|---------|
| <i>Epě kil-melle mar.</i> | |
| I come-DEB | NEG |
|
 | |
| <i>*Kel-me-lle ma-r.</i> | |
| come-NML-DIR | NEG-FIN |
| ‘I don’t have to come.’ | |

This proto-typical Transeurasian grammaticalization cycle is more explicitly recoverable in the Tungusic languages; see examples (19) and (20). In (19a) Evenki *e-* is an independent negative verb ‘not to exist, not to live’. In example (19b), the negative verb acts as a finite auxiliary to the lexical verb, which assumes an invariant adnominal form, and in (21c) the negation has shifted to a postposed position. The Nanai example in (20) represents the final stage of the negative cycle, i.e. fusion, whereby the auxiliary negative verb has assumed the status of derivational suffix on the lexical verb and its phonological form is reduced to lengthening of the stem-final vowel.

(19) Evenki

a.

Esile e-dyeli-m tadu-gla.
 now NEG-FUT-1SG there-ENCL

‘Now I will not be (live) there.’ (Nedjalkov 1994: 27)

b.

Nungan nekun-mi e-ce-n
 he younger.brother-POSS.REFL NEG-PST-3SG
suru-v-re.

go.away-CAUS-ADN

‘He did not lead his younger brother away.’ (Nedjalkov 1994: 11)

c.

Nungan songo-ro e-ce-n.
 he cry-ADN NEG-PST-3SG

‘He did not cry [—what’s the use of crying?].’ (Nedjalkov 1994: 8)

(20) Nanai

Xola:-ci-si.

read.NEG-PST-2SG

‘You didn’t read.’

Similarly, Middle Mongolian *ese-* acts as an independent negative verb, meaning ‘not to be, not to exist’ inflected with past *-be* in (21a), but gradually the negative auxiliary came to be used as an invariant form, transferring its entire inflection to the lexical verb; e.g. the past marker *-be* is attached to *ire-* ‘to come’ in example (21b).

(21) Written Mongolian

a.

Ükü-be-üü ese-be-üü.
 die-PST-INTER NEG-PST-INTER

‘Did [he] die or did [he] not?’ (Poppe 1954: 175)

b.

Mamu bayši ese ire-be.
 our teacher NEG come-PST

‘Our teacher did not come.’ (Poppe 1954: 175)

Old and Contemporary Japanese use an independent negative existential adjective *na-* ‘to be non-existent, not to exist’, which is thought to derive from the same origin as the Old Japanese negative suffix *-(a)n-*. The Korean verbal negator MK *a-ni*, *K an(i)* can also be derived from an original negative verb **an-* and the suffix MK *-i* that derives both nouns and adverbs from verbs (Robbeets 2014).

Similar to the Transeurasian languages, one of the characteristics of the Uralic languages is the expression of negation by means of a construction comprising a fully inflected negative auxiliary and a largely invariant lexical verb (Comrie 1981; Janhunen 1982: 37; Payne 1985: 215–221; Honti 1997; Suihkonen 2002: 173). In the case of the Khanty negative particles *əntə*, the negative auxiliary has become totally free of inflections and turned into an invariant verbal negator, which recalls the situation in Mongolic in (22b). However, there are no examples in Uralic in which the negative auxiliary ultimately becomes a suffix, as it does in Turkic, Tungusic and Japanese. In Yukaghir there are no language-internal indications that the proclitic clausal negation *el-* originated in a negative verb or auxiliary. For Nivkh and Chinese, we find indications that the verbal negator originated as an independent verb, but this did not follow the same pathway, through a construction comprising a fully inflected negative auxiliary and a largely invariant lexical verb. For Khanty, Nivkh and Chinese, we find indications that the verbal negator originated as an independent verb, but only in Uralic did this happen along the same pathway, through a construction comprising a fully inflected negative auxiliary and a largely invariant lexical verb. Ainu uses a negative particle that precedes the verb and cannot be derived from a verb. Neither are there indications that the Ket negative particle *bə:n* or the Rukai negative suffix *-ka* originated from a negative verb. Worldwide, the expression of negation via negative auxiliaries is a minor type to begin with, being found in only 40 (17%) out of 240 languages in Dahl's (1979) sample, which is areally biased towards Uralic and Altaic languages, in 45 (4%) out of 1011 languages in Dryer's (2005b) sample, and in 16 (5%) out of the 297 languages in Miestamo's (2005) sample. As a consequence, the particular development of negative verbs to auxiliaries to particles or suffixes is even rarer.

CS 20. Grammaticalization of plural/collective markers to express inclusive/exclusive distinction on first person pronoun

With regard to the hypothesis of Transeurasian relatedness, Johanson (2002: 154) notes that “an inclusive vs. exclusive distinction might have belonged to the old affinities. ... In the Turkic languages still containing traces of the distinction, as we have seen, it can hardly be considered a result of contact-induced copying”. In my view, this argument becomes even stronger in the light of a particular grammaticalization pattern shared across the Transeurasian languages, by which plural and collective suffixes on the first person pronoun grammaticalize into an inclusive/exclusive distinction. Among the Turkic languages, there are no unique pronominal forms that distinguish inclusive from exclusive person forms, although traces of the distinction remain in the imperative paradigms of Yakut, Tofa, Tuvan, Turkmen, Khakas, Shor, Altay and Chulym (Schönig 1987, Nevskaya 2010: 122). Nevertheless, Old Turkic and most currently spoken varieties of Turkic distinguish between a first person plural (Tk./OT *biz* ‘we’) and an augmented plural form (Tk./OT *biz-ler* ‘we (as a group)'). Nevskaya (2010: 124) argues for a collective interpretation of the

augmented plural, denoting “an isolated group of people who want to oppose themselves to the others”, rather than an inclusive interpretation as suggested by Grönbech (1936: 81). However, the collective interpretation seems to be an intermediate stage on the way from augmented plural to inclusive/exclusive distinction, because in the imperative paradigms of Khakas, Shor, Altay and Chulym, the inclusive is derived from the first plural marker augmented with the plural marker *-LAR*. As the person endings on verbs have grammaticalized from original person pronouns, it seems safe to assume that the augmented plural first person markers developed into inclusive markers at some stage in pre-Old Turkic.

The Middle Mongolian distinction between exclusive *ba* and inclusive *bida* is formally preserved in the Khalka oblique paradigm in the variation between formally exclusive *man-* and formally inclusive *bidn-*, but the functional distinction has been lost. Etymologically, the Middle Mongolian inclusive *bida*, reflected in the Khalkha formally inclusive oblique *bidn-*, derives from the first person singular MMo. *bi* ‘I’ and a plural suffix *-da*, which also occurs in the plural demonstratives pronouns MMo. *e-de* ‘these’ vs. *te-de* ‘those’ (Doerfer 1985: 2; Domii 2006; Nevskaya 2010: 119).⁴ Domii argues that originally **ba* and **bi-da* complemented each other as plural pronouns, and that the distinction between exclusive and inclusive meaning was a secondary development.

In the Tungusic languages, the inclusive-exclusive opposition is generally well preserved, e.g. exclusive Ma. *be*, Evk. *bu* vs. inclusive Ma. *muse*, Evk. *mut* ~ *mit*.

The Tungusic exclusives Evk. *bu* and Ma. *be* can be derived from the first person plural pTg **bö* and an augmented plural **bö-(x)e*, respectively (Doerfer 1978: 81–83, 95–96; Janhunen 2013: 217), whereas the inclusive Evk. *mut* ~ *mit* may go back to pTg **bö* plus the collective suffix pTg **-ti* (Benzing 1955: 1020), and the inclusive Ma. *muse* may be an extension of this root with the collective suffix *-sa* (Benzing 1955: 1017–1018).

Similar to the Turkic languages, Middle and Contemporary Korean distinguish between a first person plural (K/MK *wuli* ‘we’) and an augmented plural form (K *wuli-tul*, MK *wuli-tolh* ‘we (as a group)’) in which K *tul*, MK *·tolh* is a collective marker. A similar tendency can be found in the history of Japanese, where the first person singular/plural OJ *wa-* ‘I, we’ coexists with the same form augmented by a collective marker OJ *wa-re* ‘we’, a form which in its turn was later augmented into *ware-ra* ‘we’. Like contemporary Japanese, Old Japanese lacks a real inclusive-exclusive distinction, but the distinction is well preserved in the Ryukyuan languages; for example, in Kikai (Amami), Sesoko (Okinawa), Ikema (Miyako), Irabu (Miyako), Tarama (Miyako), Hateruma (Yaeyama), and Yonaguni (Yaeyama) (Shimoji 2014). In most cases the exclusive can be derived from the first person singular plus a plural suffix, while the inclusive is based on the same form plus a collective suffix,

4 An alternative analysis, deriving the inclusive MMo *bida* from the first singular pronoun **bi* ‘I’ plus the second plural pronoun **ta* ‘you (many)’ is proposed by Janhunen (2013: 215), but the voicing of the medial dental stop would represent an irregular development.

e.g. the Kikai exclusive suffix *-naa* is also used as a plural suffix in the second and third person plural pronouns, while the inclusive suffix *-tjaa* is also used as a collective suffix on human nouns. Therefore, it seems sensible to assume a recurrent tendency whereby plural and collective pronouns grammaticalized into an inclusive-exclusive distinction in proto-Japonic.

As is the case in many Uralic languages, Khanty marks a dual distinction, but not an inclusive-exclusive distinction, on its person pronouns. While Ket and Yukaghir lack the distinction, Nivkh distinguishes between exclusive *n'yŋ* and inclusive *mer ~ mir*, and Ainu between exclusive *cóka* and inclusive *aoka*, but there is no indication that the distinction derives from augmented plural or collective marking.⁵

The inclusive-exclusive distinction found in the first person plural pronouns between exclusive *wómen* and inclusive *zánmen* ‘we’ of Beijing and certain other northern Chinese dialects may be due Transeurasian influence. Such a distinction was not found in Old Chinese, and it began to appear in North China during the period of Altaic rule. It is significant in this regard that both Middle Mongolian, spoken under the Yuan dynasty, and Manchu, spoken under the Qing dynasty, distinguish exclusive and inclusive forms. Rukai distinguishes exclusive *-nai ~ nai* [NOM] from inclusive *-mita ~ ta* [NOM], a feature characteristic of Austronesian languages, but the forms in Rukai are not derivable from each other. In Cysouw’s (2005: 166–167) sample of 200 languages, 31% differentiate inclusive and exclusive in independent pronouns. Therefore, the percentage of languages that developed the distinction from plural/collective marking on first person pronouns will logically be much smaller.

7. How to explain the Transeurasian core structures in Turkic

Table 1 lists the languages under discussion as horizontal comparison points and the 20 selected core structures as vertical comparison points. The presence of a given feature is indicated with plus (+), absence with minus (-) and uncertainty with (+/-). In the final row, the number of plus values are added up and the percentage of core structures present in a particular language is calculated.

Table 1: Feature values for selected Transeurasian languages along with their historical stages and representative neighboring languages

Frequency world-wide	CS	Tk	(pre-) OT	Khal	(pre-) MMo	Evk	(pre-) Ma	K	(pre-) MK	J	(pre-) OJ	Khan	Ket	Yuk	Niv	Ain	Ch	Ruk
< 10%	01	-	-	+	+	+	+	-	+	-	+/-	-	-	+	+	+/-	-	-
< 30%	02	+	+	+	+	+	+	+	+	+	+	-	-	+	-	-	-	-
35%	03	+	+	+	+	-	+	+	+	+	+	+	+	+	-	+	+	-

5 In Robbeets (2017), I mistakenly analyzed Ainu as lacking an inclusive-exclusive distinction on first person plural pronoun.

61%	04	+	+	+	+	+	+	-	-	+	-	-	+	+	-	-	-	+
45%	05	+	+/-	+	+/-	-	+/-	+	+/-	+	+/-	-	-	-	-	-	-	-
54%	06	-	+	+	+	+	+	-	-	-	+	+	-	-	-	-	+	-
< 27%	07	-	+	-	+	-	+	-	+	-	+	-	-	-	-	-	-	-
	08	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-
43%	09	+	+	+	+	+	+	+	+	+	+	+	-	+	-	-	-	-
80%	10	+	+	+	+	+	+	+	+	-	+	+	+	+	-	-	-	+
< 13%	11	-	+	+	+	+	+	-	-	-	-	-	-	+	-	-	-	-
	12	-	+	+	+	+	+	-	-	-	+	+	-	-	-	-	-	-
27%	13	+	+	+	+	+	+	+	+	+	+	+	-	+	-	-	+	-
42%	14	-	-	+	+	-	+	+	+	+	+	-	-	+	-	-	+	-
	15	+	+	+	+	+	+	+	+	+	+	-	-	+	+	-	-	-
20%	16	+	+	+	+	+	+	+	+	+	+	-	+	-	+	-	-	+
< 47%	17	+	+	+	+	+	+	+	-	+	+	-	+	+	-	-	-	-
	18	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-
	19	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-
	20	-	+	+	+	+	+	-	-	-	+	-	-	-	-	-	-	-
	+ 13	18	19	20	16	19	12	13	12	17	8	5	11	3	2	4	3	
	% 65	90	95	100	80	95	60	65	60	85	40	25	55	15	10	20	15	

7.1. Turkic in the middle

In accordance with the definition given in the introduction, the 20 feature values listed in Table 1 can be regarded as “core structures” because they represent a concentration of proto-typical linguistic features that delimit the Transeurasian languages from their neighbors. This is clear from the fact that 60% or more of the structures are represented in the Transeurasian languages compared to 55% and less in the neighboring languages. Although Khanty (40%) and the Uralic languages in general show more typological similarity with the Transeurasian core structures than do other neighboring languages, it is still possible to delimit the Transeurasian languages in relation to their Uralic neighbors. This seems to contradict Janhunen’s statement quoted in the introduction.

Among the core structures that enable us to delimit the Transeurasian languages in relation to their Uralic neighbors are CS 1 presence of tongue root vowel harmony in Transeurasian vs. palatal harmony in Uralic; CS 4 presence of voicing distinction for stops vs. original singleton-geminate distinction in Uralic; CS 5 preference for non-verbal strategy for (extra-family) verbal copies vs. direct insertion in Uralic; CS 7 mixed and switched encoding of property words in Transeurasian vs. mixed and switched encoding of property words in Transeurasian vs. nominal

encoding in Uralic; CS 8 partial emphatic reduplication vs. none in Uralic; CS 11 absence of a primary initial *m* in the nominative first person singular vs. presence in Uralic, as well as presence of a (secondary) *mi-Ti* opposition in first vs. second singular person pronouns vs. *m-T* distinction in Uralic; CS 13 dependent-marking of clause arguments in Transeurasian vs. original double-marking in Uralic; CS 14 dependent-marking in possessive noun phrases vs. original head-marking in Uralic and; CS 20 Grammaticalization of plural/collective markers to inclusive/exclusive distinction on first person pronoun, compared to Uralic's lack of an inclusive-exclusive distinction on its person pronouns.

For core structures 16 and 17, namely the use of locative existential constructions to encode predicative possession, Uralic makes use of a larger variety of strategies than the Transeurasian languages, where all languages uniformly use locative possession or ablative comparatives. For CS 19, grammaticalization from negative verb to verbal negator with a construction comprising a fully inflected negative auxiliary and a largely invariant lexical verb, the Transeurasian negative markers tend to develop into suffixes bound to the verb, while the development in Uralic stops when the negative reaches the status of an unbound and invariant verbal negator.

Along the margins of the Transeurasian continuum, we can observe a gradual loss of Transeurasian features in the western and eastern peripheries. Examples of original Transeurasian features in Turkic probably changing under Uralic influence include CS 1 Transeurasian tongue root harmony, which aligns with the Uralic languages as palatal harmony in Turkic; CS 7 gradual loss of verbal encoding of property words—mirroring Uralic nominal encoding—as one proceeds from older to contemporary varieties and from Tungusic in the east to Turkic in the west; CS 11 the secondary development of *m*-initials yielding an *mi-Ti* opposition in first vs. second singular person pronouns in Turkic, Mongolic and Tungusic; and CS 14 the fact that Turkic is double marking in the possessive noun phrase, while it has the proto-typical Transeurasian pattern of dependent marking in the clause, under the influence of the proto-typical double marking in Uralic. Therefore, we can say that sandwiched between the Uralic languages to the west and the other Transeurasian languages to the east, Turkic behaves like a proto-typical Transeurasian branch under Uralic influence.

7.2. Eliminating universal tendencies in linguistic structuring

A critic could object that the core structures of Transeurasian are universally so common that their parallel occurrence in Turkic and several other branches of Transeurasian is purely coincidental. This is certainly not the case, however, because the structures are clearly concentrated in the Transeurasian area in contrast to neighboring areas, and because many features are relatively infrequent cross-linguistically. Table 1 includes an estimation of the frequency of 14 out of 20 core structures. Four features are not very common (CS 5, 9, 14, 17) in the sense that they

occur in less than half but more than a third of the languages worldwide. Six features are relatively uncommon in the sense that they occur in less than a third of the languages worldwide (e.g. CS 1, 2, 7, 11, 13, 16). Only 3 out of 20 (15%) of the core structures are common in the sense that they occur in more than half of the languages worldwide.

7.3. Eliminating code-copying

Johanson (2002: 148) aptly remarks that: “typological parallelisms do not necessarily imply genetic relationship If, for instance, Japanese and Korean ... are very similar today and exhibit typically “Ural- Altaic” syntactic features, this might be the result of early continental contacts between them as well as interactions with older forms of Tungusic and Mongolic (perhaps also Gilyak)”. Indeed, the affiliation of the Transeurasian languages remains a subject of debate, but even critics such as Janhunen (1996: 220) would agree that before the first millennium B.C. the homelands of the individual language families concerned were concentrated in a compact area in southern Manchuria, adjacent to the homelands of Nivkh (Amuric) and Sinitic speakers. This geographical situation provided an opportunity for prehistorical contact. Although some of the core structures discussed here, such as CS 8, partial emphatic reduplication of nominal property words, and CS 11, presence of *mi-Ti* opposition in first vs. second singular person pronouns are almost certainly contact-induced, others appear to be the residue of common ancestral structures, as suggested by the following observations.

1. Geography: Isolated position of Japanese

Although the Sea of Japan and the Tsushima Strait form a strong geographical boundary separating Japanese from the other Transeurasian languages, Japanese is typologically closer to the Transeurasian languages than less geographically isolated languages such as Ket, Yukaghir, Nivkh and Chinese. Even if we assume that Japonic was once present on the continent, this indicates that the Transeurasian characteristics in Japonic did not exclusively arise through code-copying because Ket, Yukaghir, Nivkh and Chinese were also present in the region.

2. History: Core structures decrease with contact over time

The typological coherence seems to be greater for historical than for contemporary stages of the languages investigated: 90% in (pre-)Old Turkic vs. 65% in Turkish, 100% in (pre-)Middle Mongolian vs. 95% in Khalkha, 95% in (pre-)Manchu vs. 80% in Evenki, 65% in (pre-)Middle Korean vs. 60% in Korean, and 85% in (pre-)Old Japanese vs. 60% in Japanese. Therefore, it is fair to say that Transeurasian core-structures have decreased over the last millennium, mainly at the peripheries, under the influence of Chinese, Siberian or Uralic features.

3. Distribution: Increase of features in Japanese

With 85% of the core structures reflected, Old Japanese is closer to the proto-type than Middle Korean, with 65% structural uniformity. Within a scenario of gradual diffusion of features, we would expect the positive values in Japanese to be lower than in Korean. It is further difficult to explain how some Transeurasian features, such as the nasal oblique pronominal stems in CS 12 could show a gap in Korean, the nasal suffix having diffused into Japanese without a Korean intermediary.

4. Areal dominance: core structures borrowed into other families

In his description of core structures of Turkic, Johanson (2015: 586–587) notices that “these specific core structures have turned out to be dominant in all the numerous family-external contact situations without being overruled by copying. They have all maintained a high genealogical stability and resisted areal influence”. Interestingly, some of the Transeurasian core structures discussed here have left a clear mark on the linguistic structure of surrounding language families, such as Sinitic. Examples include CS 6, the development of a two-way distinction in demonstratives as compared to the three way-distinction in Classical Chinese; CS 9, the weak suffixing tendency of Mandarin as opposed to other Sinitic languages; and CS 20, the development of an inclusive-exclusive distinction in first person plural pronouns in Beijing and certain other northern Chinese dialects, which was not found in Old Chinese.

5. Core dynamics: Core structures renewed by new morphological means

Johanson (2015: 588) also points out that “the stable place occupied by the aspect-tense categories in the structure of Turkic is also proven by the typical tendency to renew them by new morphological means”. The last three core features involve shared patterns of grammaticalization. They are particularly good candidates for being inherited because they appear recurrently in different forms and at various chronological stages of the same language. Aikhenvald (2013) characterized contact-induced grammaticalization as “change against the grain” or atypical grammaticalization, while she regarded genealogically motivated grammaticalization as “change that reinforces similarities” because it tends to maintain uniformity between related languages. Given that languages tend to renew their formal encodings in cyclic processes of grammaticalization while maintaining their inherited grammatical categories, new forms are thus expected to grammaticalize along shared conceptual pathways to restore old categories (Heath 1998: 729, Robbeets 2013). Consequently, genealogically motivated grammaticalization is expected to recur on different formal encodings at various points in time, while contact-induced grammaticalization is expected to be restricted to a single formal encoding (or to a very limited number of encodings) during a certain period of contact. The repeated waves of grammaticalization and replacement involved in features 18 to 20 imply that the parallel patterns are genealogically motivated.

6. *Isomorphism: core structures combine with formal correspondences*

The observation that some core structures shared among the Transeurasian languages combine with a formal correspondence of the marker reflecting the particular feature is also indicative of genealogical retention. This is for instance the case for CS 5, the non-verbal strategy of verbal borrowing employing a deverbal noun suffix of the common shape **-lA-* (Tk. *-lA-*, Khal. *-l-*, Ud. *-lA-*, J *-r(a)-*) to accommodate verbal borrowings (Robbeets 2015); CS 12, the formation of a secondary oblique stem of personal pronouns through a common suffix **-n-* in all Transeurasian languages except Korean; and in CS 18 to 20, the grammaticalizations of finiteness, negation and inclusive-exclusive distinctions (Robbeets 2014, 2016). In such cases, the core structure is likely to be genealogically motivated.

8. Conclusion

As an attempt to answer Johanson's 2015 open question about the Transeurasian origin of certain core structures in Turkic, I delimited a number of core structures shared by the Transeurasian languages and focused on how these are reflected in the Turkic languages. My study shows that the cluster of core structures in Transeurasian is clearly delimited in relation to its neighbors to the west (Uralic), the north (Yeniseic, Yukaghiric), the east (Nivkh, Ainu) and to the south (Sinitic, Austronesian). Although my observations are in line with Janhunen's findings about a certain internal uniformity within the larger Ural-Altai belt, they contradict his claim that there are no typological features that would be specific only to "Altaic" independent from Uralic. My research indicates that Turkic typically behaves as a language of the Transeurasian type, but deviates from the prototype under the influence of Uralic.

Admitting that many features shared between Turkic and the other Transeurasian languages ought to be a result of code-copying, my study further suggests that it is more parsimonious to attribute the majority of core structures discussed here to inheritance than to code-copying. The structural evidence advanced in this paper thus complements the evidence in support of Transeurasian genealogical continuity, which I proposed earlier on the basis of lexical and morphological comparison (Robbeets 2005, 2015). As a result, I believe in response to Johanson's question "Are the core structures of Turkic common Transeurasian structures?" it is legitimate to answer "partly yes".

Abbreviations

Linguistic forms

ABL	ablative	INTER	interrogative
ACC	accusative	LOC	locative
ADD	additive	NEG	negative
ADN	adnominalizer	NML	nominalizer
CAUS	causative	NOM	nominative

CLASS	classifier	NPST	non-past
COM	comitative	OBL	oblique
COMP	comparative	PF	perfect
CONV	converb	PFV	perfective
COND	conditional	PL	plural
CS	core structure	POL	polite
DIR	directive	POSS	possessive
DAT	dative	PROC	processive
DUR	durative	PST	past
ENCL	enclitic	PT	particle
FIN	finite	REFL	reflexive
FUT	future	SG	singular
GEN	genitive	TOP	topic
HON	honorific	→	derivation
INCH	inchoative	>>	copy

Languages

Ain.	Ainu	Niv	Nivkh
Ch.	Mandarin Chinese	WMo.	Written Mongolian
Evk.	Evenki	OJ	Old Japanese
J	Japanese	OT	Old Turkic
K	Korean	pJ	proto-Japonic
Khal.	Khalkha	pK	proto-Koreanic
Khan.	Khanty	pMo	proto-Mongolic
Ket	Ket	pTg	proto-Tungusic
Ma.	Manchu	pTk	proto-Turkic
MK	Middle Korean	Ruk.	Mantauran Rukai
MMo.	Middle Mongolian	Yuk.	Yukaghir

Acknowledgements

I wish to thank Nataliia Neshcheret for her help in collecting some data underlying this article. The research leading to these results has received funding from the European Research Council under the Horizon 2020 Program/ERC Grant Agreement n. 646612 granted to Martine Robbeets.

References

- Aikhenvald, Alexandra 2013. Areal diffusion and parallelism in drift: Shared grammaticalization patterns. In: Robbeets, Martine & Cuyckens, Hubert (eds.) 2013. *Shared grammaticalization: With special focus on the Transeurasian languages* (Studies in Language Companion Series 132.) Amsterdam: Benjamins. 23–42.
- Alpatov, Vladimir & Podlesskaya, Vera 1995. Converbs in Japanese. In: Haspelmath, Martin & König, Ekkehard (eds.) 1995. *Converbs in cross-linguistic perspective. Structure and meaning of adverbial verb forms: Adverbial particles, gerunds*. Berlin: Mouton de Gruyter. 465–486.

- Anderson, Gregory 2005. The velar nasal (η). In: Haspelmath, Martin et al. (eds.) 2005. *The world atlas of language structures*. Oxford: Oxford University Press. 42–45.
- Anderson, Gregory 2006. Towards a typology of the Siberian linguistic area. In: Matras, Yaron & McMahon, April & Vincent, Nigel (eds.) *Linguistic areas. Convergence in historical and typological perspective*. Basingstoke: Palgrave Macmillan. 266–300.
- Aso, Reiko 2015. Hateruma Yaeyama grammar. In: Heinrich, Patrick & Miyara, Shinsho & Shimoji, Michinori (eds.) 2015. *Handbook of the Ryukyuan languages. History, structure and use*. (Handbooks of Japanese Language and Linguistics 11.) Berlin: De Gruyter Mouton. 423–488.
- Benzing, Johannes 1955. *Die tungusischen Sprachen. Versuch einer vergleichenden Grammatik*. (Akademie der Wissenschaften und der Literatur, Abhandlungen der geistes- und sozialwissenschaftlichen Klasse 11.) Wiesbaden: Steiner.
- Bulatova, Nadežda Ja. & Grenoble, Leonore A. 1999. *Evenki*. (Languages of the World/Materials 141.) Munich: Lincom.
- Bugaeva, Anna 2015. Causative constructions in Ainu: A typological perspective with remarks on the diachrony. *STUF-Language Typology and Universals* 68: 4, 439–484.
- Collinder, Björn 1965. *An introduction to the Uralic languages*. Berkeley: University of California Press.
- Comrie, Bernard 1981. Negation and other verb categories in the Uralic languages. In: Ikola, Osmo (ed.) *Congressus Quintis Internationalis Fenno-Ugristarum* 6. Turku: Suomen Kielen Seura. 350–355.
- Cysouw, Michael 2005. Inclusive/exclusive forms for ‘we’. In: Haspelmath, Martin et al. (eds.) 2005. *The world atlas of language structures*. Oxford: Oxford University Press. 162–169.
- Dahl, Östen 1979. Typology of sentence negation. *Linguistics* 17, 79–106.
- Diessel, Holger 2005. Distance contrasts in demonstratives. In: Haspelmath, Martin et al. (eds.) 2005. *The world atlas of language structures*. Oxford: Oxford University Press. 170–173.
- Doerfer, Gerhard 1978. Urtungusisch \ddot{o} . In: Doerfer, Gerhard & Weiers, Michael (eds.) 1978. *Beiträge zur nordasiatischen Kulturgeschichte* (Tungusica 1.) Wiesbaden: Harrassowitz. 66–116.
- Doerfer, Gerhard 1982. Nomenverba in Türkischen. In: Gallotta, Aldo & Marazzi, Ugo (eds.) 1982. *Studia turcologica memoriae Alexii Bombaci dicata*. Naples: Istituto Universitario Orientale. 101–114.
- Doerfer, Gerhard 1985. *Mongolo-Tungusica*. Wiesbaden: Steiner.
- Domii, Tumurtogoo 2006. The inclusive and the exclusive in Mongolian. In: Shagdarsuren, Ts. (ed.) 2006. *Mongol ulsin ix sürgүүлйн. Erdem šinžilgeenij bičig*. (Acta Mongolica 6.267). Ulanbaatar: National University of Mongolia. 77–78.
- Dryer, Matthew S. 2005a. Prefixing versus suffixing in inflectional morphology. In: Haspelmath, Martin et al. (eds.) 2005. *The world atlas of language structures*. Oxford: Oxford University Press. 110–113.
- Dryer, Matthew 2005b. Negative morphemes. In: Haspelmath, Martin et al. (eds.) 2005. *The world atlas of language structures*. Oxford: Oxford University Press. 454–457.
- Erdal, Marcel 2004. *A grammar of Old Turkic*. Leiden: Brill.
- Filchenko, Andrey Yury 2007. A grammar of Eastern Khanty. Houston: Rice University PhD. Dissertation.

- Frellesvig, Bjarke 2010. *A history of the Japanese language*. Cambridge: Cambridge University Press.
- Frellesvig, Bjarke & Whitman, John 2008. Evidence for seven vowels in proto-Japanese. In: Frellesvig, Bjarke & Whitman, John (eds.) 2008. *Proto-Japanese: Issues and prospects*. (Amsterdam studies in the theory and history of linguistic science. Series 4. Current issues in linguistic theory 249.) Amsterdam: Benjamins. 15–41.
- Gil, David 2005. Numeral classifiers. In: Haspelmath, Martin et al. (eds.) 2005. *The world atlas of language structures*. Oxford: Oxford University Press. 226–229.
- Georg, Stefan 2007. *A descriptive grammar of Ket (Yenisei-Ostyak) 1. Introduction, phonology, morphology*. Folkestone: Global Oriental.
- Gorelova, Liliya 2002. *Manchu grammar*. Leiden: Brill.
- Göksel, Asli & Kerslake, Celia 2005. *Turkish. A comprehensive grammar*. London: Routledge.
- Grönbech, Karl 1936. *Der türkische Sprachbau*. Copenhagen: Levin & Munksgaard.
- Gruzdeva, Ekaterina 1998. *Nivkh*. (Languages of the World Materials 111.) Munich: Lincom.
- Gruzdeva, Ekaterina 2016. Epistemic modality and related categories in Nivkh. *Studia Orientalia* 117, 171–198.
- Haspelmath, Martin 1995. The converb as a cross-linguistically valid category. In: Haspelmath, Martin & König, Ekkehard (eds.) 1995. *Converbs in cross-linguistic perspective. Structure and meaning of adverbial verb forms: Adverbial particles, gerunds*. Berlin: Mouton de Gruyter. 1–56.
- Heath, Jeffrey 1998. Hermit crabs: Formal renewal of morphology by phonologically mediated affix substitution. *Language* 74, 728–759.
- Honti, László. 1997. Die Negation im Uralischen 1–3. *Linguistica Uralica* 2, 81–96, 161–176, 241–252.
- Iwasaki, Shoichi 2006. *Japanese. Revised edition*. (London Oriental and African Language Library 17.) Amsterdam: Benjamins.
- Janhunen, Juha 1981. Korean vowel system in North Asian perspective. *Hangeul* 172, 129–146.
- Janhunen, Juha 1982. On the structure of proto-Uralic. *Finnisch-ugrische Forschungen* 44, 23–42.
- Janhunen, Juha 1996. *Manchuria. An ethnic history*. (Mémoires de la Société Finno-Ougrienne 222.) Helsinki: Suomalais-Ugrilainen Seura.
- Janhunen, Juha 2012. *Mongolian*. (London Oriental and African Language Library 19.) Amsterdam: John Benjamins.
- Janhunen, Juha 2013. Personal pronouns in Core Altaic. In: Robbeets, Martine & Cuyckens, Hubert (eds.) 2013. *Shared grammaticalization: With special focus on the Transeurasian languages* (Studies in Language Companion Series 132.) Amsterdam: Benjamins. 211–226.
- Janhunen, Juha 2014. Ural-Altaic. The polygenetic origins of nominal morphology in the Transeurasian zone. In: Robbeets, Martine & Bisang, Walter (eds.) 2014. *Paradigm change in the Transeurasian languages and beyond*. (Studies in Language Companion Series 161.) Amsterdam: Benjamins. 311–335.
- Johanson, Lars 1993. Graphie und Phonologie im Türkischen: Probleme der Lautharmonie. In: Werner, Otmar (ed.) 1993. *Probleme der Graphie* (Script Oralia 57.) Tübingen: Narr. 83–94.

- Johanson, Lars 1995. On Turkic converb clauses. In: Haspelmath, Martin & König, Ekkehard (eds.) 1995. *Converbs in cross-linguistic perspective. Structure and meaning of adverbial verb forms: Adverbial particles, gerunds*. Berlin: Mouton de Gruyter. 313–348.
- Johanson, Lars 2002. *Structural factors in Turkic language contacts*. Richmond: Curzon.
- Johanson, Lars 2006. Adjectives and nouns in South Siberian. In: Erdal, Marcel & Nevskaya, Irina (eds.) *Exploring the eastern frontiers of Turkic in South Siberia*. (Turcologica 60.) Wiesbaden: Harrassowitz. 57–78.
- Johanson, Lars 2015. So close and yet so distant ... On Turkic core structures, genealogical and typological grouping of varieties, and mutual intelligibility. In: Zeyrek, Deniz & Şimşek, Çiğdem Sağın & Ataş, Ufuk & Rehbein, Jochen (eds.) *Ankara papers in Turkish and Turkic linguistics*. Wiesbaden: Harrassowitz. 583–592.
- Johanson, Lars & Robbeets, Martine 2010. Introduction. In: Johanson, Lars & Robbeets, Martine (eds.) 2010. *Transeurasian verbal morphology in a comparative perspective: genealogy, contact, chance*. (Turcologica 78.) Wiesbaden: Harrassowitz. 1–5.
- Kaiser, Stefan & Ichikawa, Yasuko & Kobayashi, Noriko & Yamamoto, Hirofumi 2001. *Japanese: A comprehensive grammar*. London: Routledge.
- Ko, Seongyeon 2012. Tongue root harmony and vowel contrast in Northeast Asian languages. New York: Cornell University Ph.D. dissertation.
- Ko, Seongyeon & Whitman, John & Joseph, Andrew 2014. Comparative consequences of the tongue root harmony analysis for proto-Tungusic, proto-Mongolic, and proto-Korean. In: Robbeets, Martine & Bisang, Walter (eds.) 2014. *Paradigm change in the Transeurasian languages and beyond* (Studies in Language Companion Series 161.) Amsterdam: Benjamins. 141–176.
- Lee, Ki-Mun & Ramsey, Robert 2011. *A history of the Korean language*. Cambridge: Cambridge University Press.
- Li, Charles N. & Thompson, Sandra A. 1989. *Mandarin Chinese. A functional reference grammar*. Berkeley: University of California Press.
- Malchukov, Andrej 2013. Verbalization and insubordination in Siberian languages. In: Robbeets, Martine & Cuyckens, Hubert (eds.) 2013. *Shared grammaticalization with special focus on the Transeurasian languages*. (Studies in Language Companion Series 132.) Amsterdam: John Benjamins. 177–208.
- Maddieson, Ian 2005. Voicing in plosives and fricatives. In: Haspelmath, Martin et al. (eds.) 2005. *The world atlas of language structures*. Oxford: Oxford University Press. 22–25.
- Martin, Samuel Elmo 1988. *A reference grammar of Japanese*. Tokyo: Tuttle.
- Martin, Samuel Elmo 1992. *A reference grammar of Korean*. Tokyo: Tuttle.
- Maslova, Elena 2003. *A grammar of Kolyma Yukaghir*. (Mouton Grammar Library 27.) Berlin: Mouton de Gruyter.
- Miestamo, Matti 2005. *Standard negation: The negation of declarative verbal main clauses in a typological perspective* (Empirical Approaches to Language Typology 31.) Berlin: Mouton de Gruyter.
- Nedjalkov, Igor V. 1994. Negation in Evenki. In: Kahrel, Peter & Van den Berg, René (eds.) 1994. *Typological studies in negation*. Amsterdam: John Benjamins. 1–34.
- Nedjalkov, Igor V. 1997. *Evenki. Descriptive grammar*. London: Routledge.
- Nevskaya, Irina 2010. *Inclusive and exclusive in Altaic languages*. In: Johanson, Lars & Robbeets, Martine (eds.) 2009. *Transeurasian verbal morphology in a comparative perspective: genealogy, contact, chance*. (Turcologica 78.) Wiesbaden: Harrassowitz. 115–128.

- Nichols, Johanna 1986. Head-marking and dependent-marking grammar. *Language* 62: 1, 56–119.
- Nichols, Johanna 2012. Selection for *m*: T pronominals in Eurasia. In: Johanson, Lars & Robbeets, Martine (eds.) 2012. *Copies versus cognates in bound morphology*. (Brill's Studies in Language, Cognition and Culture 2.) Leiden: Brill. 47–70.
- Nichols, Johanna & Bickel, Balthasar 2005a. Locus of marking in the clause. In: Haspelmath, Martin et al. (eds.) 2005. *The world atlas of language structures*. Oxford: Oxford University Press. 98–101.
- Nichols, Johanna & Bickel, Balthasar 2005b. Locus of marking in possessive noun phrases. In: Haspelmath, Martin et al. (eds.) 2005. *The world atlas of language structures*. Oxford: Oxford University Press. 102–105.
- Nichols, Johanna & Peterson, David 2005. Personal pronouns. In: Haspelmath, Martin et al. (eds.) 2005. *The world atlas of language structures*. Oxford: Oxford University Press. 546–553.
- Payne, John R. 1985. Negation. In: Shopen, Timothy (ed.) 1985. *Language typology and syntactic description* 1. *Clause structure*. Cambridge: Cambridge University Press. 197–242.
- Poppe, Nicholas 1954. *Grammar of Written Mongolian*. Wiesbaden: Otto Harrassowitz.
- Poppe, Nicholas 1955. *Introduction to Mongolian comparative studies*. (Mémoires de la société Finno-Ougrienne 110.) Helsinki: Suomalais-Ugrilainen Seura.
- Poppe, Nicholas 1964. Der altaische Sprachtyp. In: Spuler, Bertold et al. (eds.) 1964. *Mongolistik* (Handbuch der Orientalistik 5.2.) Leiden: Brill. 1–16.
- Robbeets, Martine 2005. *Is Japanese related to Korean, Tungusic, Mongolic and Turkic?* (Turcologica 64.) Wiesbaden: Harrassowitz.
- Robbeets, Martine 2013. Genealogically motivated grammaticalization. In: Robbeets, Martine & Cuyckens, Hubert (eds.) 2013. *Shared grammaticalization: With special focus on the Transeurasian languages* (Studies in Language Companion Series 132.) Amsterdam: Benjamins. 147–175.
- Robbeets, Martine 2014. The development of negation in the Transeurasian languages. In: Whaley, Lindsay & Suihkonen, Pirkko (eds.) *Typology of languages of Europe and Northern and Central Asia*. (Studies in Language Companion Series 164.) Amsterdam: Benjamins. 401–420.
- Robbeets, Martine 2015. *Diachrony of verb morphology: Japanese and the Transeurasian languages*. (Trends in Linguistics. Studies and Monographs 291.) Berlin: Mouton-De Gruyter.
- Robbeets, Martine 2016. Insubordination and the establishment of genealogical relationship. In: Evans, Nicholas & Watanabe, Honore (eds.) *Dynamics of insubordination*. (Typological Studies in Language.) Amsterdam: Benjamins. 209–245.
- Robbeets, Martine 2017. The Transeurasian languages. In: Hickey, Raymond (ed.) *The Cambridge handbook of areal linguistics* (Cambridge Handbooks in Language and Linguistics.) Cambridge: Cambridge University Press. 586–626.
- Rybatzki, Volker 2003. Middle Mongol. In: Janhunen, Juha (ed.) 2003. *The Mongolic languages*. London: Routledge. 57–82.
- Schönig, Claus 1987. Diachronic and areal approach to the Turkic imperative paradigm. In: van Damme, Mark & Boeschoten, Hendrik (eds.) 1987. *Utrecht papers on Central Asia* (Utrecht Turkological Series 2.) Utrecht: University of Utrecht. 205–222.

- Shibatani, Masayoshi 1990. *The languages of Japan*. (Cambridge Language Surveys.) Cambridge: Cambridge University Press.
- Shimoji, Shimoji 2014. Dual in Ryukyuan languages: typology and diachrony. Presentation held at the Language variation and change research forum, Kyushu University 24.05. 2014.
- Sohn, Ho-min 1994. *Korean*. London: Routledge.
- Sohn, Ho-min 2009. The semantics of clause linking in Korean. In: Dixon, R[obert] M. W. & Aikhenvald, Alexandra 2009. *The semantics of clause linking. A cross-linguistic typology*. Oxford: Oxford University Press. 285–317.
- Suihkonen, Pirkko 2002. The Uralic languages. *Fennia* 180: 1–2, 165–176.
- Stassen, Leon 2005a. Predicative possession. In: Haspelmath, Martin et al. (eds.) 2005. *The world atlas of language structures*. Oxford: Oxford University Press. 474–477.
- Stassen, Leon 2005b. Predicative adjectives. In: Haspelmath, Martin et al. (eds.) 2005. *The world atlas of language structures*. Oxford: Oxford University Press. 478–481.
- Stassen, Leon 2005c. Comparative constructions. In: Haspelmath, Martin et al. (eds.) 2005. *The world atlas of language structures*. Oxford: Oxford University Press. 490–493.
- Street, John 1957. *The language of the Secret History of the Mongols*. New Haven: American Oriental Society.
- Svantesson, Jan-Olof 1985. Vowel harmony shift in Mongolian. *Lingua* 67: 4, 283–327.
- Tamura, Suzuko 2000. *The Aimu language*. (ICHEL Linguistic Studies 2.) Tokyo: Sanseidō.
- Vajda, Edward J. 2004. *Ket*. (Languages of the World/Materials 204.) Munich: Lincom.
- Vovin, Alexander 1993. *A reconstruction of Proto-Aimu*. (Brill's Japanese Studies Library 4.) Leiden: Brill.
- Vovin, Alexander 2005. *A descriptive and comparative grammar of Western Old Japanese 1: Sources, script and phonology, lexicon, nominals*. (Languages of Asia 3.) Folkestone: Global Oriental.
- Vovin, Alexander 2009. *A descriptive and comparative grammar of Western Old Japanese 2: Adjectives, verbs, adverbs, conjunctions, particles, postpositions*. (Languages of Asia 8.) Folkestone: Global Oriental.
- Weiers, Michael 1966. Untersuchungen zu einer historischen Grammatik des präklassischen Schriftmongolisch. Bonn: Rheinischen Friedrich-Wilhelms-Universität Ph.D dissertation.
- Werner, Heinrich. 1997. *Die ketische Sprache*. Wiesbaden: Harrassowitz.
- Whaley, Lindsay J. & Li, Fengxiang 2000. Emphatic reduplication in Oroqen and its Altaic context. *Linguistics* 38: 2, 355–372.
- Wohlgemuth, Jan 2009. *A typology of verbal borrowings*. (Trends in Linguistics. Studies and Monographs 211.) Berlin: Mouton de Gruyter.
- Wrona, Janick 2008. The nominal and adnominal forms in Old Japanese: Consequences for a reconstruction of pre-Old Japanese syntax. In: Frellesvig, Bjarke & Whitman, John (eds.) 2008. *Proto-Japanese. Issues and prospects*. (Current Issues in Linguistic Theory 294.) Amsterdam: Benjamins. 193–215.
- Yap, Foong Ha & Matthews, Stephen. 2008. The development of nominalizers in East Asian and Tibeto-Burman languages. In: María José López-Couso & Elena Seoane (eds.) *Rethinking grammaticalization: New perspectives* (Typological Studies in Linguistics 76.) Amsterdam: John Benjamins. 309–341.
- Zeitoun, Elizabeth 2007. *A grammar of Mantauran (Rukai)*. (Language and Linguistics Monograph Series A4–2.) Taipei: Institute of Linguistics, Academia Sinica.

Kazan Tatar as a dominant language of the Volga–Kama region. A case study of lexical intermediation

Klára Agyagási

Agyagási, Klára 2017. Kazan Tatar as a dominant language of the Volga-Kama region. A case study of lexical intermediation. *Turkic Languages* 21, 36–45.

In the first part of the paper, a short overview is given of the emergence of the political, cultural and language dominance of Kazan Tatar (Kipchak) ethnicity in the Volga-Kama region. In the second part, the role of Islam in the Tatar culture is demonstrated. Then, through a historical areal study of the Arabic word *şabi* ‘boy, male child’, it is discussed how this lexical element spread among the dominated languages of the Kazan Tatar Khanate: Arab → Kazan Tatar → Middle Chuvash (Viryal dialect) → Mari dialects.

Klára Agyagási, Debrecen University, Institute of Slavic Studies, H-4032 Debrecen, Egyetem tér 1. E-mail: agyagasi.klara@arts.unideb.hu

1. Introduction

As is documented in historical sources, Kazan Tatars had already been present in the Middle Volga Region for several centuries by the 15th century when they gained a political role as a leading ethnicity in their own state.

The medieval political history of the Central Volga Region begins with the founding of the Volga Bulgarian state in the early 10th century (for details, see Zimonyi 1990, and Golden 1992: 253–258). The center of power of the Volga Bulgarian state was built at the confluence of the Volga and Kama rivers, but the state extended considerably to the north and east as well, mostly along the left bank of the Volga, subjugating the various neighboring ethnic groups, most of which were of Finno-Ugric origin. The two and a half century period of prosperity of the Volga Bulgarian empire was brought to an abrupt end by the Mongolian invasion, with the Mongolian army devastating the Bulgarian capital in 1236. The Volga Bulgarian power elite immediately submitted to Mongolian rule, whereas the other ethnic groups of the empire fled their original territory of settlement.

As a result, the ethnic structure of the region changed completely. The conquering Mongolians formed their own state in the region among the newly restructured population, calling it Joči Ulus (Golden 1992: 297–302).¹ One of the main characteristics of this state was that Mongolians were outnumbered in it (limited mostly to the center of power), whereas the majority of the population consisted of the

1 This state thus formed was later called the Golden Horde.

Kipchaks who were the first to submit to the Mongolian rule on the Eastern European steppe at the time of the Batu campaigns. (The first sepulchers to Kipchak Turkic people in the Volga–Kama region date back to the 14th century, with the number of sepulcher inscriptions in Kipchak increasing steadily over time.) The other important characteristic was that this state continued to exercise control over the Finno-Ugric and other ethnic groups which lived on the peripheries of the new geographical arrangement. By the end of the rule of the Golden Horde the Mongolians were Turkified (cf. Golden 1992: 292), and the successor state of the Mongolian empire, the Kazan Khanate (beginning in 1437) was unequivocally dominated by Kipchak Turks.

The ethnogenesis of the Middle Volga Region outlined briefly here can be traced in the changes of the linguistic relations of the people living here. From the 10th until the 16th century, this region always belonged to a single political unit which, however, was ruled by three different elites (Volga Bulgarian, Mongolian, and then Kazan Tatar) over the centuries. This means that the various language groups were not separated by political borders and, thus, that language change could spread freely among the languages of the region.²

The discipline of contact linguistics has proposed various typologies for the description of changes of extralinguistic origin (e.g. Thomason & Kaufman 1988, Campbell 1998, Muysken 2000, Hickey 2012). A common characteristic of these is that they differentiate between a dominant language (i.e. the language of the people exercising political power), which is the source of language change, and subordinated languages, whose historical relationship to each other can be formed according to various scenarios.

In the rest of this paper, a case study will be presented tracing the fate of an Arabic word borrowed by the dominant Kipchak (Kazan Tatar) language as a result of the spread of Islam, which then spread (i.e. from the center of the linguistic area) to the Cheremis language, where it completely lost its connection with Islamic culture.

2. The case of Ar. *şabi* ‘boy, male child’

Arabic and Persian loanwords in the languages of the Volga Region were the result of the spread of Islam and Islamic culture. They were borrowed directly into the languages of the region from written language use, but because only a few of the ethnic groups in the Volga Region converted to Islam, the circumstances of borrowing need to be examined in every single case.

2 The Volga-Kama Region was defined as a linguistic area (Sprachbund) as early as in the 1970s (cf. Serebrennikov 1972), with the relevant linguistic processes (Bereczki 1983, Helimski 2003) and the mechanisms of the Sprachbund (Bereczki 2005, Agyagási 2012, Johanson 2013: 656–657) being described.

2.1. Islam in the Volga-Kama Region

It is a well-known fact that Islam first appeared in the Volga Region in the early 10th century, namely, in 921/922, when the sovereign of the Volga Bulgarians and his court converted to Islam (cf. Ibn Fadlan, Golden 1992: 256–257). Muslim sources report that Volga Bulgarian tribes living along the “bank of the river”³ all converted to Islam and had mosques, muezzins, imams, and Koranic schools (Golden 1992: 254, Róna-Tas 1982a: 155). At the same time, it is reasonable to assume that Islam did not become the exclusive faith among the peoples of the Volga Bulgarian empire—only among their tribal leaders and those in their immediate circles, and only by the late 10th century. The Chuvash language (which is not descended from a central dialect, but from another Volga Bulgarian dialect, cf. Agyagási 2007) preserves only very few words associated with the early Islamic influence that could have been borrowed directly in Chuvash.⁴

The Arabic and New Persian loanwords in Chuvash were borrowed via Kazan Tatar (Schermer 1977), and well after the 10th century. It is a view generally held among scholars that Islam affected Chuvash only starting from the 16th century, and only in a very cursory fashion (Braslavskij 1997: 16–22). This is also supported by the fact that the very rich traditional belief system has been preserved up to our time, and the folk customs associated with it were still alive in the early 20th century (Mesaroš 2000, Salmin 2007).

The second wave of Islamization reached the Volga-Kama Region under the reign of Özbek Khan (1312–1342). This is the period when Islam prevailed in the Golden Horde as well (Golden 1992: 298, Vásáry 2003), and from which the spread of Islamic culture from dominant Kazan Tatar to the peripheries via various “avenues” is dated.

2.2. Arabic loanwords in Cheremis dialects

The Arabic loanwords in the Cheremis lexicon were first pointed out by Budenz (1864: 398–400). This study uses Cheremis dialect forms from Reguly’s Mountain Cheremis lexicon and discusses 17 Cheremis words of Arabic origin, alongside which Reguly systematically presents the Chuvash and Tatar variants of the Arabic borrowings as well. He states that the Arabic words entered the Cheremis lexicon “via Tatar-Chuvash”. Budenz’s choice of words (“Tatar-Chuvash”) is simply a collective designation referring to the Volga Turkic people. A phonologically based analysis of the relationship of source vs. recipient language words was not Budenz’s aim in this paper.

3 This is a reference to the Volga and Kama rivers, that is, the central region of the empire.

4 Relevant loanwords of Arabic origin include *dūnya* → Chuvash *tēnče* (Róna-Tas 1982b), and of New Persian origin, *χ^waja* → Old Chuvash *χuja* (Róna-Tas 1988: 761), and, despite Erdal’s claim (Erdal 1993: 139) most likely *āđīna* (Agyagási 1982: 11–12).

In his 1923 work, Räsänen discusses the Tatar loanwords of the Cheremis language, 61 words of Arabic origin from all over the entire Cheremis language area which were borrowed into Cheremis via Tatar. In his monograph published three years earlier (Räsänen 1920), he mentioned 13 words which were borrowed from Chuvash into Cheremis. These were borrowings of Arabic origin in Chuvash, and every Chuvash word of Arabic origin has a reflex⁵ in Tatar. Räsänen's 1969 Turkic etymological dictionary does not present anything new regarding Arabic loanwords in Cheremis.

Isanbaev (1989: 33–35) presents an evaluative summary of the topic, mostly based on sources by Russian authors. He estimates the number of Arabic loanwords in Cheremis at 200, but does not provide an exhaustive listing. He regards most of the Arabisms to have been borrowed into Cheremis via Tatar.

As is clear from the above overview, in studying Arabic loanwords in Cheremis dialects, researchers have always recognized the fact that these were borrowed via Tatar, but they did not connect it with the mechanisms at work in the Volga-Kama linguistic area, and did not analyze the phonological details of this intermediary stage of borrowing regarding the specific forms either.

2.3. Arabic *şabi* → Kazan Tatar *sabïy* → Cheremis *şuβo* ~ *şjβâ*. The issue of intermediation

The Cheremis word which I will associate with the Arabic form *şabi* 'young boy, young child' below is not used by itself in Cheremis, but only as the second member of coordinate structures (for details, see the etymological database below). The first element of this structure also means 'boy' (*erɣe* and *iɣe*) and is part of the basic Finno-Ugric vocabulary preserved in the Cheremis lexicon (cf. Bereczki & Agyagási 2013: 17, 20).

Of the Volga-Kama Turkic languages, only Kazan Tatar has the Arabic word. The sound correspondences between the Arabic and the Kazan Tatar variants can easily be established, since the Arabic word is part of the Classical Arabic language and thus is found in the Koran (for details, again, see the etymological database below), where the diacritics needed to reproduce the exact pronunciation are included. This unequivocally shows that the Arabic word had to be pronounced [ʃabi]. This form was likely borrowed through written language use, after the wider spread of Islam began around the 14th century (at the time of Middle Kipchak), by one of the Kipchak dialects which relocated to the Volga Region; this was the dialect that

5 The Chuvash forms quoted here (and some others as well) were later shown, by Scherner, to have been borrowed into Chuvash from Tatar (for a full listing, see Scherner 1977: 195), which had borrowed them from Arabic.

later served as the basis for the central dialect of Kazan Tatar: Arabic *ṣabi* → Middle Kipchak *sabīy*.⁶

The Middle Kipchak form, however, could not have been the direct source of the Cheremis form, since in the history of Cheremis dialects following the breakup of the Proto-Cheremis language there was never a sound change $a > u$, or $a > i$ in the initial syllable⁷ that could have affected the Middle Kipchak word (Räsänen 1923: 14). Also, the borrowed (i.e. foreign origin) word-final cluster $-īy$ never underwent a $-īy > o \sim \hat{a}$ change in loanwords in Cheremis dialects (cf. Agyagási 2013a: 205).

The sound changes $a > u$ and $a > i$ are, however, well-known in the history of Chuvash. They affected the phoneme /a/ of the Oghur variety of Western Old Turkic.⁸ Oghur /a/ was realized as either an illabial [ã] or a labial [â] allophone already before the Mongol invasion of Europe, which can be demonstrated by the Old Turkic loanwords in Hungarian (Róna-Tas & Berta 2011: 1114–1117), and also by the Old Russian loanword in Volga Bulgarian (Agyagási 2005: 68–69). The two allophones developed into separate phonemes in the Middle Chuvash period, following the Mongol invasion of Europe—phonemes that can be described by different distinctive features: [ã] > /i/, [â] > /o/ -/u/. The Middle Chuvash timeline of the early stages of the process is traceable in the Middle Mongol loanwords in the Chuvash language, e.g. Middle Mongol *malayai* ‘cap’ → Middle Chuvash **mālaḡay* > *molaḡay* > *mulaḡay* (Róna-Tas 1982b: 99–100, 117). The change [ã] > /i/ occurred later and affected fewer words.

The phonological difference between the two allophones of the phoneme /a/ became, in the Middle Chuvash period, a feature serving as a basis for the classification of further dialects of the Chuvash variety of Volga Bulgarian; the [ã] > /o/ -/u/ change characterized the Northern (or Viryal) dialect, whereas the [â] > /i/ change, the Southern (or Anatri) dialect.

Oghur **saw* (c. EOT *sāg* ‘sound, healthy’, Clauson 1972: 803) > VB_{Chuv} **sāw* ~ **sāw*
 VB_{Chuv} **sāw* > Middle Chuvash *sov* > V *suv(ǎ)* (Ašmarin 11: 163)
 VB_{Chuv} **sāw* > Late Middle Chuvash **siv(ǎ)* > A *siv(ǎ)* (c. *sivǎ* ‘zdorovyj’
 Ašmarin 11: 105)

6 In the Tatar form, the secondary word-final $-y$ was added in order to avoid morphological ambiguity; in Middle Kipchak dialects word final $i\sim i$ functioned as a third person possessive marker.

7 In the Cheremis dialect data documented by Beke at the JO V and K collection points, the word underwent regular reduction of the u in the initial syllable (cf. Agyagási 2002).

8 The term Western Old Turkic was introduced by Róna-Tas (1998); for descriptions of the characteristics of i in the Oghur variety (called Bulgarian-Turkic earlier) on the basis of Old Turkic loanwords in Hungarian, see Róna-Tas & Berta (2011), as well as Agyagási (2013b).

Since it was the Viryal dialect that participated in Chuvash-Cheremis contacts, it is natural that words containing initial syllable Middle Chuvash /a/ had the labial allophone of /a/, and when these words were borrowed, Cheremis dialects typically borrowed them with the labial reflex (o) of [â]. Thus, Cheremis dialects often preserved this initial syllable labial sound of the Viryal type after the illabial variants of the Anatri based literary language forced them out of use in the Modern Chuvash period, cf.:

Oghur **talkïy* ‘hemp breaker’ (c. EOT *talku*, Róna-Tas & Berta 2011: 906) >
 VB_{Chuv} **tâlî ~ tãlî*
 VB_{Chuv} **tâlî* > Middle Chuvash *tolđ* > V **tulã* → Cher. P B M UP *tole*, UJ C Ć
tule, V K *tule* ‘Hanfbreche’ (Beke 8: 2788)
 VB_{Chuv} **tâlî* > Late Middle Chuvash *tĩlã* > A *tĩlã* ‘mjalka’ (Ašmarin 13: 315), lit.
tĩlã

It is a well-known fact, however, that in the Sundyr and Morgaush-Yadrin sub-dialects of the Viryal dialect, that is, in the area of contact between Chuvash and Cheremis, alongside the normative Viryal *â* > *u* change, the *â* > *ĩ* change also occurred next to labial consonants (Sergeev 2007: 353–356)—which can be regarded as a secondary dissimilation phenomenon.

Oghur **bar-* (c. EOT *bar-* ‘to go’, Clauson 1972: 354) > VB_{Chuv} **bãr- ~ *bãr-*
 > **pãr- ~ *pãr*
 VB_{Chuv} **pãr-* > Late Middle Chuvash **pĩr-* > A *pĩr-* ‘idti, xodit’ (Buinsk,
 Samara, Ašmarin 9: 184)
 VB_{Chuv} **pãr-* Middle Chuvash **por* > V *pur-* ‘prixodit’ (Ašmarin 9: 293) ~ *pĩr-*
 (Abaševo, Čeboksarskij rajon, Ašmarin 9: 185)

With regard to all these facts, in the further history of Arabic *šabi* → Middle Kipchak *sabïy*, it is plausible to posit that the Tatar word was borrowed from Kazan Tatar into the Northern dialect of the Middle Chuvash language area, where its initial syllable vowel became labial, and its word-final *-ïy* diphthong participated in the Middle Chuvash *-ïy* > *i* change.⁹ Since, in the Middle Chuvash variety, *b* no longer occurred intervocalically (for details, see Agyagási 2010: 207–208), word internal *b* was replaced by a voiceless lenis in Middle Chuvash: Middle Kipchak *sabïy* → Middle Chuvash *sãbi*. After this, the form *sãbi* underwent the *â* > *o* > *u* change cha-

9 The monophthongization of diphthongs was a general tendency in the Middle Chuvash variety. In native words the diphthong *-ay* occurred word finally, changing into *i* in Middle Chuvash. The results of this change are visible in Cheremis in Middle Chuvash loanwords, cf. Oghur **buγday* ‘wheat’ (EOT *bugday*) > Middle Chuvash **puray* > *põri* → Mountain Cheremis (Kozmodemjansk) *põri* ‘spelt’ (Róna-Tas & Berta 2011: 186, Räsänen 1920: 177).

racteristic of the Viryal dialect (*sâBi* > *soBj* > *suBj*). Since the change happened in a labial environment, the delabialization of *u* in the Sundryr variety of the Viryal dialect produced the variant *sîBi*. The Mari dialects likely borrowed both variants of this form from here at the time when Proto Mari broke up and the split into eastern and western dialects occurred (both groups of dialects having the word-initial *s* > *š* change). (This change began before the breakup of Proto Mari and ended in the first half of the 17th century, cf. Berczki 1965: 73.)¹⁰ The Viryal (Sundryr) dialect variants were later lost from the lexicon of this dialect.

On the basis of the above, the history of Arabic *šabi* in the Volga-Kama Region language area can be reconstructed as follows:

Ar. *šabi* → Kazan Tatar *sabjy* → Middle Chuvash V *sâBi* > *soBD* > *suBD* → Cher. dial. *šuβo*

V_{Sundryr} *suBD* > *sîBi* → Cher. dial. *šiβç*.

The fact that this word forms coordinate structures with elements of Finno-Ugric origin in Cheremis dialects demonstrates that, at the time when it was borrowed, Chuvash-Cheremis bilingualism already existed.

Etymological database

Arabic data:

صَبِيّ: 1. ‘a youth, boy or male child; a young male child’; 2. ‘a young woman, girl or female child’ (Lane 1968: 1650); صَبِيّ ‘page, jeune homme servant auprès d’un prince’ (Dozy 1881: 817)¹¹

Turkic data:

Osm. *sabi* ‘Kind, Knabe’, Krč. *sabi* ‘Säugling’, Trkm. *sabjy* ‘mladenec’, Kzk. *sâBi* ‘ein kleines Kind’ (Räsänen 1969: 391)

Tat. *sabdy* ‘mladenec, ditja, rebenok; nezrelyj, zelenyj, ne dostigšij soveršennoletija’ (TRSl 459); صَبِيّ ‘kečkenä ir bala’ [young male child] (Khamzin & Makhmutov & Sayfullin 1965: 493).

Cheremis data:

P B M UJ *iye* C Č JT *iye*, MK JO V K *iγə*, UP *iγə* ‘Kind, Junge eines Tieres’ (Beke 1997: 304.)

UJ *èrɣè-šuβo* ‘Knabe, männlicher Nachfolge’, P B M UJ *iye-šuβo*, MK *iγə-šijβə*, UP *iγə-šjβə*, JO V *iγə-šəβə*, JP *iγə-šjβú*, B *ikšuβo*, CÜ CK Č *ik-šəβə*, K *ik-šəβə*, JT *iγə-šəγo* ‘Nachkommen, Kind’ (Beke 2001: 2517).

Ob₁ *ik-šuβo*, OB₂ *iγə-šuβo*, Okr. *iye-šuβo*, *iγəšuβo*, (MalK) *iγi-šijβə*, *iγi-šuβo*,

¹⁰ This was a change that occurred generally in the Chuvash and Volga Bulgarian loanwords in Cheremis dialects.

¹¹ I am grateful to István Zimonyi for his help in accessing Arabic sources.

Ok *i·γ̣-ṣ̌iβ̣ə*, Ms *i·γ̣ə-ṣ̌u·β̣ə ~ ṣ̌u·γ̣ə*, Mm₁ *i·k-ṣ̌əβ̣ə*, *i·k'ṣ̌əβ̣ə*, Mm₂ *i·k-ṣ̌əβ̣ə*, *ikṣ̌əβ̣ə*, Mmu *i·kṣ̌əβ̣e* ‘Kinder’ (Moisio & Saarinen 2008: 149).

Literature

- Agyagási, K. 1982. On the edition of Chuvash literary sources. In: Róna-Tas, A. (ed.) *Chuvash Studies*. Budapest: Akadémiai Kiadó. 7–17.
- Agyagási, K. 2002. K voprosu o pojavlenii sverxkratnix labial'nyx glasnyx v marijskix dialektax. In: Luutonen, J. (ed.) *Volgan alueen kielikontaktit. Turun Yliopiston Soumalaisen ja Yleisen kielitieteen laitoksen julkaisuja 70* [On the appearance of the extra-short vowels in Mari dialects]. Turku: Department of Finnish and General Linguistics of the University of Turku. 70–86.
- Agyagási, K. [Adjagaši, K.]. 2005. *Rannie russkie zaimstvovanija tjurkskix jazykov Volgo-Kamskogo areala* [Early Russian loanwords in the Turkic languages of the Volga-region]. (Studies in linguistics of the Volga-region 2.) Debrecen: University Press.
- Agyagási, K. 2007. Mittelbulgarische Dialekte. Mittelbulgarischer Sprachzustand. In: Boeschoten, H. & Stein, H. (eds.) *Einheit und Vielfalt in der türkischen Welt. Materialien der 5. Deutschen Turkologenkonzferenz Universität Mainz*. (Turcologica 69.) Wiesbaden: Harrassowitz. 24–36.
- Agyagási, K. 2010. Loanwords as data in historical linguistics. An areal linguistic study. *Sprachtheorie und Germanistische Linguistik* 20: 2, 197–222.
- Agyagási, K. 2012. Language contact in the Volga-Kama Area. In: Kincses-Nagy, É. & Bicsi, M. (eds.) *The Szeged meeting. Proceedings of the Fifteenth International Conference on Turkish Linguistics held on August 20–22, 2010 in Szeged*. (Studia Uralo-Altaica 49.) Szeged: Department of Altaic Studies. 21–37.
- Agyagási, K. 2013a. O trex redkix dialektizmax, legšix v osnovu nekotoryx russkix zaimstvovaniy v marjskom jazyke [On three archaic Russian dialectal words borrowed by the Mari dialects]. *Slavica* 42, 201–211.
- Agyagási, K. 2013b. Nyugati ótörök és magyar kapcsolatok: Tanulságok az ogur hangtörténet számára [Contacts between Western Old Turkic and Hungarian: Consequences for the historical phonology of Oghur]. In: Agyagási, K. & Hegedűs, A. & É. Kiss, K. (eds.) *Nyelvelmélet és kontaktológia* [Linguistic theory and language contact]. Pázmány Péter University. 155–172.
- Ašmarin, N. I. 1928–1950. *Slovar' čuvaškogo jazyka* 1–17 [Chuvash dictionary 1–17]. Kazan' & Čeboksary: Narodnyj Komissariat po Prosveščeniju Čuvašskoj ASSR.
- Beke, Ö. 1997–2002. *Mari nyelvjárási szótár* 1–9 [Cheremis dialect dictionary 1–9]. Revised edition by G. Bereczki et al. (Bibliotheca Ceremissica Savariae 4.) Szombathely: Savariae.
- Bereczki, G. 1965. Wichtigere lautgeschichtliche Lehren der russischen Lehnwörter im Tscheremissischen. In: *Congressus Secundus Internationalis Fenno-Ugristarum Helsingae habitus* 23–28. 8, 1. *Acta Linguistica* (Helsinki, Societas Fenno-Ugrica), 70–78.
- Bereczki, G. 1983. A Volga-Káma vidék nyelveinek areális kapcsolatai [Areal contacts among the languages of the Volga–Kama region]. In: Balázs J. (ed.) *Areális nyelvészeti tanulmányok* [Studies in areal linguistics]. Budapest: Tankönyvkiadó.
- Bereczki, G. 2005. Vzaimosvjazi jazykov volgo-kamskogo areala [Linguistic contacts in the Volga-Kama region]. *Congressus Internationalis Fenno-Ugristarum* (Joškar-Ola) 10:1, 1–49.

- Bereczki, G. & Agyagási, K. 2013. *Etymologisches Wörterbuch des Tscheremissischen (Mari). Der einheimische Wortschatz. Nach dem Tode des Verfassers herausgegeben von Klára Agyagási und Eberhard Winkler.* (Veröffentlichungen der Societas Uralo-Altaica Band 86.) Wiesbaden: Harrassowitz Verlag.
- Braslavskij, L. Ju. 1997. *Islam v Čuvašii* [Islam in Chuvashia]. Čeboksary: Izdatel'stvo Čuvašija.
- Budenz, J. 1864. Cseremisiz tanulmányok [Cheremis studies]. *Nyelvtudományi Közlemények* 3, 397–470.
- Campbell, L. 1998. *Historical linguistics. An introduction.* Edinburgh: Edinburgh University Press.
- Clauson, sir G. 1972. *An etymological dictionary of pre-thirteenth century Turkish.* Oxford: Clarendon Press.
- Dozy, R. 1881. *Supplément aux dictionnaires arabes* 1. Leyden: E. J. Brill.
- Erdal, M. 1993. *Die Sprache der wolgabulgarischen Inschriften.* (Turcologica 13.) Wiesbaden: Harrassowitz.
- Golden, P. B. 1992. *An introduction to the history of the Turkic peoples. Ethnogenesis and state-formation in Medieval and Early Modern Eurasia and the Middle East.* (Turcologica 9.) Wiesbaden: Harrassowitz.
- Helimski, E. 2003. Areal groupings (Sprachbünde) within and across the borders of the Uralic language family: A survey. *Nyelvtudományi Közlemények* 100, 156–168.
- Hickey, R. 2012. Language contact: Reconsideration and reassessment. In: Hickey, R. (ed.) *The handbook of language contact.* Oxford: Wiley-Blackwell. 1–28.
- Isanbaev N. I. 1989. *Marijsko-tjurkskie jazykovye kontakty* 1. *Tatarskie i baškirske zaimstvovanija* [On Mari–Turkic language contacts. Tatar loanwords in Bashkir]. Joškar-Ola: Marijskoe knižnoe izdatel'stvo.
- Johanson, L. 2013. Turkic language contacts. In: Hickey, R. (ed.) *The handbook of language contact.* Oxford: Wiley-Blackwell. 652–672.
- Khamzin, K. Z. & Makhmutov M.I. & Sayfullin G. Š. (eds.) 1965. *Arabsko-tatarsko-russkij slovar' zaimstvovaniy* [Dictionary of Arabic-Tatar-Russian loanwords] Kazan': Tatarstan Publishing House.
- Lane, E. W. 1968. *An Arabic-English lexicon, derived from the best and the most copious eastern sources; comprising a very large collection of words and significations omitted in the ḲĀMOOS with supplements to its abridged and defective explanations, ample grammatical and critical comments, and examples in prose and verse: Composed by means of the munificence of the most noble Algernon, Duke of Northumberland, K. G., and the bounty of British Government: by Edward William Lane.* Book 1, Part 4, London, 1872. Reprinted in eight parts. Part 4: Beirut & Lebanon.
- Mesaroš, D. 2000. *Pamjatniki staroj čuvašskoj very* [Memories of the Old Chuvash religion]. Čeboksary: Chuvash State Institute of Humanities.
- Moisio, A. & Saarinen, S. 2008. *Tscheremissisches Wörterbuch. Aufgezeichnet von Volmari Porkka, Arvid Genetz, Yrjö Wichmann, Martti Räsänen, T. E. Uotila und Erkki Itkonen.* (Lexica Societas Fenno-Ugrocae 32.) Helsinki: Suomalais-Ugrilainen Seura.
- Muysken P. 2000. From linguistic areas to areal linguistics. A research proposal. In: Gilbers D. & Nerbonne. J. & Schaeken, J. (eds.) *Languages in contact.* Amsterdam & Atlanta: Rodopi. 263–275.

- Osmanov, M. M. (ed.) 1966. *Tatarsko-russkij slovar'* [Tatar-Russian dictionary]. Moskva: Akademija nauk SSSR, Kazanskij institut jazyka, literatury i istorii. Izdatel'stvo Sovetskaja énciklopedija.
- Räsänen, M. 1920. *Die tschuwaschischen Lehnwörter im Tscheremissischen*. (MSFOu 48.) Helsinki: Société Finno-Ougrienne.
- Räsänen, M. 1923. *Die tatarischen Lehnwörter im Tscheremissischen*. (MSFOu 50.) Helsinki: Société Finno-Ougrienne.
- Räsänen, M. 1969. *Versuch eines etymologischen Wörterbuch der Türksprachen*. Helsinki: Suomalais-Ugrilainen Seura.
- Róna-Tas, A. 1982a. The periodization and sources of Chuvash linguistic history. In: Róna-Tas, A. (ed.) *Chuvash studies*. Budapest: Akadémiai Kiadó. 113–169.
- Róna-Tas, A. 1982b. *Studies in Chuvash etymology* 1. (Studia Uralo-Altaica 17.) Szeged. 66–134.
- Róna-Tas, A. 1988. Turkic influence on the Uralic languages. In: Sinor, D. (ed.) *The Uralic languages*. Leiden & New York: Brill. 742–780.
- Róna-Tas, A. 1998. Western Old Turkic. In: Johanson, L. et al. (eds.) *The Mainz meeting. Proceedings of the Seventh International Conference on Turkish Linguistics, August 3–6, 1994*. (Turcologica 32.) Wiesbaden. 619–626.
- Róna-Tas, A. & Berta Á. 2011. *West Old Turkic. Turkic loanwords in Hungarian*. (Turcologica 84.) Wiesbaden: Harrassowitz.
- Salmin, A. K. 2007. *Sistema religij čuvašej*. Sankt-Peterburg: Nauka.
- Scherner, B. 1977. *Arabische und neupersische Lehnwörter im Tschuwaschischen. Versuch einer Chronologie ihrer Lautveränderungen*. (Akademie der Wissenschaften und der Literatur. Veröffentlichungen der orientalischen Kommission 29.) Wiesbaden: Franz Steiner.
- Serebrennikov, B. A. 1972. O nekotoryx otličitel'nyx priznakax volgokamskogo jazykovogo sojuza [On some distinctive features of Volg-Kama Sprachbund in Bashkiria]. In: Garipov, T. M. & Čeremisina, N. V. (eds.) *Jazykovye kontakty v Baškirii* [Linguistic contacts in Bashkiria]. (Učenyje zapiski 50., Serija filologičeskix nauk.) Ufa. 7–16.
- Sergeev, L. P. 2007. *Dialektnaja sistema čuvaškogo jazyka*. Čeboksary: Chuvash State Pedagogical University.
- Thomason, S. G. & Kaufman T. 1988. *Language contact, creolization, and genetic linguistics*. Los Angeles & London: University of California Press.
- TRSl=Osmanov 1966.
- Vásáry, I. 2003². *A régi Belső Ázsia története* [The history of old Inner Asia]. Budapest: Balassi.
- Zimonyi, I. 1990. *The origins of the Volga Bulgars*. (Studia Uralo-Altaica 32.) Szeged: Department of Altaic Studies, University of Szeged.

Two questionable candidates for subordinatorship: *-mİşlİK* and *-mAzlİK* in Turkish

Annette Herkenrath & Birsel Karakoç

Herkenrath, Annette & Karakoç, Birsel 2017. Two questionable candidates for subordinatorship: *-mİşlİK* and *-mAzlİK* in Turkish. *Turkic Languages* 21, 46–78.

This paper investigates morphosyntactic, semantic and functional qualities of the complex verbal forms *-mİşlİK* and *-mAzlİK* in modern literary Turkish. It discusses their potential to serve as subordinators and explores the transitional zone between abstract nominalisation and “clausiness”, by using corpus-linguistic methodology. The results show that while these rarely used forms do have the capacity to expand into clause-like structures, they also reveal some categorially contradictory patterns. Morphosyntactically, the study attempts to rank the findings on a scale. It also looks at patterns of combinability of clausal with nominal categories. Semantically, the data reveal a tendency for these constructions to be employed in the expression of (passive) negative states and to occur with matrix predicates that express emotional experience, nonverbal communication or actions and existing states, rather than explicit verbal or cognitive processing or evaluation.

Annette Herkenrath, Justus Liebig University, Giessen, Germany, E-mail: annette.herkenrath@germanistik.uni-giessen.de

Birsel Karakoç, Department of Linguistics and Philology, Uppsala University, Box 635, SE-75126 Uppsala, Sweden. E-mail: birsel.karakoc@lingfil.uu.se

1. Introduction

This paper explores various morphosyntactic, semantic and functional qualities of the complex verbal forms *-mİşlİK* and *-mAzlİK* in modern literary Turkish and discusses the clausal potential of the constructions in which they occur. *-mİşlİK* and *-mAzlİK* are composed of the following elements: *-mİş* is a participial suffix with an aspectual, postterminal function, whereas *-mAz* is a negated intraterminal participial. *-lİK* is widely used as a highly productive nominal derivational suffix, whose usual behaviour is to attach to nouns or adjectives.¹ We look at a corpus in

1 Morphologically similar forms can be found in other Turkic languages as well, e.g. *-A(:)ndİK* and *-yĀndİK* in Turkmen, *-GANlİK* in Nohay and Kazakh, *-GAndİK* in Kirghiz, *-GānliK* and *-(ā)yātgānliK* in Uzbek, *-GANlik*, *-Adiyanlik* and *-Ivatqanlik* in Uyghur. Our assumption is that despite some morphological and compositional parallelism, their developmental paths and functional qualities in individual languages are not the same. Corresponding forms in the other Turkic languages can largely be traced back to

which *-mİşLIK* and *-mAzLIK* occasionally occur in constructions reminiscent of complement clauses and, to a lesser extent, of adverbial clauses. These marginal usages have so far largely been ignored in the literature. Examples (1) and (2) illustrate the two forms in question:

- (1) *Bilinçli, biçimsel bir kesik-lik değil bu. Cefa*
 conscious formal one cut off-DER.N not DEI suffering
çek-mişlik-ten ileri gel-en fiziksel, ruhsal bir
 draw-MİŞLIK-ABL forward come-PAR physical psychological one
kesik-lik de değil. Özel bir yapı
 cut off-DER.N also not special one structure
yeniliğ-i ol-sa ne âlâ!
 renewal-POSS.3SG be-CD.3SG what excellent

'This is not a conscious, formal state of being cut off. Neither is it a physical, psychological state of being cut off that comes from a state of having suffered. How excellent if it were a special renewal of the structure!' (Necatigil 1993: 179)

- (2) *Belki de berber-in kendin-e sığ-mazlığ-ı*
 perhaps also barber-GEN self.POSS.3SG-DAT fit into-MAZLIK-POSS.3SG
var-di orada; ...
 existent-COP.PST.3SG there

'And perhaps it was the case that the barber was in internal turmoil there; ...' (lit.: 'his not fitting into himself existed') (Toptaş 1995: 11)

Constructions based on *-mİşLIK* and *-mAzLIK* exhibit characteristics that place them on a continuum ranging from lexicalisation via abstract nominalisation to clausal constructions. The present paper is interested in the following questions: (1) To what extent and by which morphosyntactic criteria can these constructions be assigned clausal status? (2) In what areas or through what morphosyntactic contradictions does their status remain ambiguous? (3) What is semantically special about the contexts in which *-mİşLIK* and *-mAzLIK* are used? (4) What specific illocutionary quality of *-mİşLIK* and *-mAzLIK* makes them preferable to any of the canonical subordinators in the given contexts?

We draw on two theoretical strands. The first is the morphosyntactic discussion of clausal subordination in frameworks of Turkish and typological linguistics. Since subordination in Turkish makes use of nominalising linguistic means, one central issue has always been the nominal versus clausal status of these forms. We also cast a cursory glance at our data in terms of functional-semantic discussions of illocutionary status. We operationalise this in terms of morphosyntactic but above all semantic categories relating to the superordinate structure. These two larger strands

seems to be more plausible. The present paper leaves aside the larger Turkic perspective, concentrating on Turkish alone.

intersect at several points. Both depend on a close reading of the textual passages in which specific findings occur. Methodologically, the investigation follows an empirical, concordancing, approach, qualitatively exploring some distinguishing categories in several data-driven takes, to be quantitatively synthesised at the end.

Section 2 presents the data. Section 3 inventories the types of constructions in which *-mİşlik* and *-mAzlik* are used. Section 4 discusses criteria for clausal status of nominalisations. Sections 5–7 present our morphosyntactic analysis, while Section 8 focuses on the semantic issues.

2. Data

The corpus consists of 4,817 pages of contemporary Turkish literary prose, from which we compiled a concordance containing 202 entries: 113 instances of *-mİşlik* and 89 of *-mAzlik*. Overall, neither *-mİşlik* nor *-mAzlik* occurs very frequently. Our average is some two occurrences per hundred pages, slightly more for *-mİşlik*, and fewer for *-mAzlik*. Our preliminary overview suggests some quantitative differences between authors, which raise questions of style and narrative function.² See Table 1.

- 2 This study is exclusively based on published literary data. A literary corpus obviously cannot yield results that would be generalisable to all registers of language use. We do not aim at such a generalisation. As an anonymous reviewer pointed out, it would be desirable to consider spoken-language data as well. An automatic search in the ENDFAS/SKOBI corpus (754,633 words, 213,622 utterances, Rehbein 2009, Rehbein, Herkenrath & Karakoç 2009), which consists of elicited spontaneous conversations with Turkish children and their families, revealed 65 tokens of the lexicalised expression *yaramazlık* ‘naughtiness, mischievousness, good-for-nothingness’, more than four out of five from adult speakers, and one productive item (*kullanılmışlık* ‘state of having been used/ exploited’), from an adult speaker. None of these occurrences show any clausal characteristics. By contrast, we found these forms to be more abundant in literary prose texts, which we consider as ‘natural’ data in their own right. It might, however, be the case that ‘clausy’ employment of *-mİşlik* and *-mAzlik* is a phenomenon of intentionally planned as opposed to spontaneous language. Additional spoken-language corpora, collected from adult speakers only (under construction), might be considered for a comparative follow-up in the future. The present study can be seen as preliminary in that it explores some possibilities of the use of *-mİşlik* and *-mAzlik* in a specific, but nonetheless relevant, set of data.

Table 1: Database, quantitative overview

Author	First publication	Title	Pages		<i>-mİşLIK</i>		<i>-mAzLIK</i>	
			overall	concordanced	absolute	per 100 pages	absolute	per 100 pages
Akhanlı	2010	<i>Fasıl</i>	240	190	-	-	1	0.52
Atasü	1991	<i>Dağın öteki yüzü</i>	281	281	4	1.42	8	2.84
Buğra	1963	<i>Küçük Ağa</i>	486	486	2	0.41	15	3.08
Çelik	2011	<i>Öfkenin şenliği</i>	110	110	4	3.63	-	-
Dorsay	1986	<i>Yüzyüze: Sinemacılarla konuşmalar</i>	256	256	7	2.73	3	1.17
Erdoğan	1998	<i>Kırmızı pelerinli kent</i>	143	76	1	1.31	-	-
Hisar	1942	<i>Boğaziçi mehtapları</i>	309	309	-	-	2	
İlhan	1999	<i>Yengecin kıskacı: Dört uzun hikâye</i>	259	259	3	1.15	-	-
Karasu	1999	<i>Öteki metinler</i>	139	139	9	6.47	5	3.59
Kaygusuz	2006	<i>Yere düşen dualar</i>	333	207	16	7.72	3	1.44
Kemal	1960	<i>Orta direk</i>	392	392	7	1.78	6	1.53
Kutlu	2007	<i>Merhaba sevgi</i>	141	141	-	-	3	2.12
Levi	1999	<i>İstanbul bir masaldı</i>	741	332	37	11.14	13	3.91
Mağden	2006	<i>İki genç kızın romanı</i>	256	65	-	-	1	1.53
Margosyan	2001	<i>Tespîh taneleri</i>	512	163	-	-	-	-
Necatigil	1993	<i>Düzyazılar 1</i>	338	338	16	4.73	14	4.14
Nesin	2005	<i>Adamı zorla deli ederler</i>	192	192	2	1.04	4	2.08
Soysal	1968	<i>Tante Rosa</i>	105	105	2	1.90	3	2.85
Tanpınar	1983	<i>Hikayeler</i>	326	326	-	-	5	1.53
Toptaş	1995	<i>Gölgesizler</i>	226	226	2	0.88	1	0.44
Uzuner	1989	<i>Güneş yiyen çingene</i>	104	104	-	-	2	1.92
Özakman	1999	<i>Toplu oyunları 3. Gençlik ve çocuk oyunları</i>	120	120	1	0.83	-	-
Σ			6,009	4,817	113	2.34	89	1.84

3. An inventory of constructional types

-mİşlİK and *-mAzlİK* occur in different types of lexicalised expressions and fixed grammaticalised constructions. These comprise, at the lexical level: *anlaşmazlık* ‘disagreement’, *geri kalmışlık* ‘retardment’, *kendini beğenmişlik* ‘conceit(edness), arrogance’, *umursamazlık* ‘indifference’, *vurdumduymazlık* ‘insensitivity, thick-skinnedness, block-headedness’, and *yaramazlık* ‘good-for-nothingness, naughtiness, bad behaviour’.³ At the syntactic level, they comprise a series of fixed constructions: *-mAzlİKtAn gel-*, *-mAzlİğA gel-*, *-mAzlİğA vur-*, *-mİşlİğA vur-*, *-mAzlİK et-/yap-*, and *-mİşlİK-POSS var/yok*. We exclude from further analysis 35 occurrences containing lexicalised expressions and 25 fixed constructions. This halves our stock of *-mAzlİK*-findings (from 89 to 42), and also slightly reduces it on the *-mİşlİK* side (from 113 to 100), leaving us with 142 findings amenable to testing; see Table 2.

Table 2: Overall findings, excluded lexicalised and fixed forms, remaining findings

	<i>-mİşlİK</i>	<i>-mAzlİK</i>	Σ
Overall	113	89	202
Lexicalised	<i>geri kalmışlık</i> ‘retard’ (6), <i>kendini beğenmişlik</i> ‘conceit(edness), arrogance’ (3)	<i>anlaşmazlık</i> ‘disagreement’ (18), <i>umursamazlık</i> ‘indifference, ignorance’ (6), <i>vurdumduymazlık</i> ‘insensitivity, thick-skinnedness, block-headedness’ (1), <i>yaramazlık</i> ‘good-for-nothingness, naughtiness, bad behaviour’ (1)	35
Fixed	<i>-mİşlİğA vur-</i> (1), <i>-mİşlİK-POSS var/yok</i> (3)	<i>-mAzlİKtAn gel-</i> (12), <i>-mAzlİK et-/yap-</i> (7), <i>-mAzlİğA gel-</i> (1), <i>-mAzlİğA vur-</i> (1)	25
Remaining	100	42	142

The remaining 142 findings occur in different roles within their immediate syntactic environments (external level):

- in copula clauses: as subjects (1) or predicates (4 findings, e.g. *o yer bir sağırlıktı, bir terk edilmişlikti* ‘that place was a deafness, a having-been-leftness’, Levi 2000: 315),
- in existential clauses: as subjects (5, e.g. *bizde bu değerbilmezlik varken* ‘as long as there is this unappreciativeness on our part’, Nesin 2005: 178),
- in possessive clauses: as possessors (2, e.g. *Mozes Bronstein’in bu kendine çekilmişliğinin, savunulabilir yönleri de vardı elbet* ‘this internal withdrawnness

3 In fact, this list of lexicalised items might not be exhaustive. Our data include candidates such as *görmemişlik* ‘the state of not having seen, unmannerliness’ that might be seen as also bordering on lexicalisation. We retain them for the analysis.

of Mozes Bronstein's had its defensible sides, of course', Levi 2000: 52) or possessives (3, e.g. *bin yıllık duvarcı Murat Usta'nın elbet bir geçilmezliği vardı* 'Master Murat, a mason for a millennium, had, of course, his unsurpassedness', Nesin 2005: 31),

- as genitive attributes in genitive-possessive constructions (12, e.g. *Cumhuriyet Ankarası söylencesinde yuğurulmuşluğumun etkisi* 'the influence of my having-been-formed in the myth of Republican Ankara', Karasu 2010: 99),
- as non-head nouns in nominal compounds (18, e.g. *aynı değişmezlik rüyası* 'the same dream of unchangeability', Tanpınar 2002: 206),
- as a dative complement to a governing NP (1, *vicdanındaki ağırlığın önlenemezliğine inancı* 'his/her belief in the unpreventability of the feeling of heaviness in his/her conscience', Atasü 2007: 189),
- as complements of postpositions: datives (1) or ablatives (4, e.g. *bu kadar yorgunluktan, tükenmişlikten sonra* 'after so much tiredness, exhaustedness', Kemal 1984: 261),
- in constructions with noun-based postpositions: as genitive attributes (3, e.g. *bilmezliklerinin ardına saklanmak* 'to hide behind their not-knowingnesses', Levi 2000: 261; *o kadının ulaşılmazlığı [...] karşısında* 'in view of the unattainability of that woman', Levi 2000: 138) or as nominative attributes (10, e.g. *yer yetmezliği yüzünden* 'because of shortage of space/space-lacking(-ness)', Necatigil 1993: 332),
- as complements of verbs: ablatives (6), accusatives (25, e.g. *kaçınılmazlığı kavrayan bilinci* 'his/her consciousness that comprehends the unavoidable', Atasü 2007: 117), datives (6), instrumentals (1) or nominative/caseless complements (2),
- as adverbials: ablatives (3), locatives (2), *diye*-marked phrases (1), instrumentals (9, e.g. *o buruk, mahzun yenilmişliğiyle [...] yürüyordu* 'she was walking with that bitter, sad state of hers of having-been-defeated', Erdoğan 1998: 61),
- as a genitive attribute/complement in a light-verb construction (1, *tükenmiş bitmişliğinin farkına varıyordu* 's/he noticed his/her having become burnt-out and exhausted', Kemal 1984: 276),
- as subjects in verbal clauses (22, e.g. *halk türkülerindeki yapı ve duyarlılığın vazgeçilemezliği anlaşılmış* 'the unavoidable of the structure and sensitivity in the folk songs has been understood', Necatigil 1993: 167).

4. Criteria for clausiness in Turkish nominalisations

The discussion on clausal complementation in Turkish has so far focused on four core suffixes *-DİK*, *-(y)AcAK*, *-mA*, and *-(y)İŞ*. These are nonfinite elements, which combine with possessives and case. In addition, Turkish has some complex aspectual complementisers, which combine an aspectual form with a copular verb (*-(Ø)Iyor ol-*, *-mİş ol-*, *-(V)r ol-*, *-(y)AcAk ol-*), which in turn combines with one of the aforementioned core suffixes (e.g. recently Karakoç & Herkenrath 2016). This

system has largely been considered somewhat of a closed-class phenomenon. Our present approach argues in favour of an opening of this class. *-mİşlİK* and *-mAzlİK*, the complementiser status of which still remains to be clarified, are marginal in frequency of use, in productivity, and in terms of the theoretical attention they have received. Deny (1921) makes no mention of either of the forms. Banguoğlu (1995: 271) mentions *-mAzlİK* as a suffix deriving “negative action nouns” (*olumsuz kılış adları*). Lewis (1967: 170, 173, 256, see also 289) analyses *-mAkIİK*, *-mAzlİK* and *-mAmAzlİK* as forming abstract nouns and *-mİşlİK* as forming “substantival sentences”. Lewis does not discuss subordinator status or productivity, although some of his examples do have a (quasi-)clausal character, unfolding an argument structure. The grammar of Göksel & Kerslake (2005), while discussing individual components (*-lİK*, *-mİş*, and *-mAz*) does not mention any of the combined forms. Van Schaaijk (2000: 116) analyses the verbal noun *-mİşlİK* as a “propositional compound”.

So far, none of the works that mention *-mİşlİK*/*-mAzlİK*-constructions specifically focus on the question of their potentially clausal status. For our present attempt to tackle this question, we can only draw on discussions of Turkish nominalisation in general. Generatively based works in Turkish linguistics mainly discuss a “hybrid” or “mixed”, i.e. nominal and verbal/clausal, status of Turkish nominalisation/subordination. These studies speak of “mixed extended projections” (Borsley & Kornfilt 2000), “functional nominalisation”, and a “terminus of the verbal projection in nominalisations” (Kornfilt & Whitman’s 2012a, b). This discussion is related to the question of whether Turkish has “complementisers” at all, in a generative sense.⁴ Those works focus on *-DIK*, *-(y)AcAK*, *-mA* and *-(y)İş*.

Discussing the clausal character of Turkic nonfinite constructions (nominalised propositions), Johanson (e.g. 1975, 1990a, 1990b, 1996, 2010, 2013) argues that Turkish nominalised propositions can realise aspect, diathesis, and argument structure to a larger degree than can Indo-European abstract nominalisations for example.⁵ Consequently, a large area of Turkish nominalisations has to be treated as equivalent to subordinate clauses, as found in Indo-European languages. Since phenomena of abstract nominalisation do also exist in Turkish, the exact status of the

4 In generative theory, subordinating functional categories are conceived of as governing heads of the functional categories of the CP system. They can primarily be expected to govern IP- and, ultimately, VP-related projections. Turkish subordination poses a theoretical challenge to generativism in that nominal categories intervene between the VP- and the CP-area; the IP-area is a zone of transition, where inflectional categories can be verbal or nominal.

5 Kural (1998: 405–408) draws a distinction between the nominalising derivational *-(y)Im* and the semantically comparable clausal subordinator *-(y)İş*, based on diathesis, argument realisation, negation, and adverbial modification. Kural (1993, 1998) is interested in identifying the one complementising element common to the Turkish subordinating suffixes; he identifies it as the *-K*-element.

constructions in question—as abstract nominalisations or as subordinate clauses—remains to be analytically established.

Haig (2003) proposes a continuum between derivation and inflection, with an intervening “inflection-derivational borderline”. In his model, *-İİK* functions as a “first-level derivation”, applying directly to simple or derived lexemes that have not yet crossed the borderline to inflection, whereas *-mİş* and *-V^{or}/-mAz* occupy the “inflection-derivational borderline”. The application of *-İİK*-based derivational processes to derived forms, especially those already on the borderline to inflection (*-mİşlİK*), poses a challenge to Haig’s model: a phenomenon of “backlooping”, which needs to be explained. Haig’s solution consists of analysing *-mİşlİK*-forms as having undergone a process of lexicalisation enabling recursivity in the application of derivation at this stage. It is precisely this latter assumption that we hope to be able to challenge with our data; a number of the *-mİşlİK*/*-mAzİİK*-forms in our corpus show signs of being inflected forms rather than lexicalisations.

The debate outlined above can be contextualised within a crosslinguistic discussion of categorial gradation (or “non-discrete grammar”) (Ross 1972, 2004 [1973]). Ross proposes two categorial continua (“squishes”), namely one that ranges from verbs to nouns via different types of participles, adjectives, and adjectival nouns (Ross 1972: 316), and one ranging from *that*-clauses to lexical nouns via prepositionally subordinated clauses, embedded *wh*-constructions, different types of *-ing*-constructions, action nominals, and derived nominals (Ross 2004 [1973]: 351). While constructed on the basis of highly English-specific phenomena, these two continua are of a cross-linguistic interest as they concern more or less precisely the areas under investigation in connection with Turkish *-mİşlİK* and *-mAzİİK*.

While crosslinguistic discussions related to functional-semantic and -pragmatic concepts of “illocution”—e.g. Boye’s (2010) distinction between “object of perception” and “knowledge acquired” or Rehbein’s (2007) ideas of a “compositional illocution” and of the “descriptive realisations” of an illocution—form part of the theoretical background of the present study, we cannot adequately discuss them here. The present approach, with its semantic analysis, is focused on data-driven corpus-linguistic methods.

5. Attempting to construct a “nouniness-clausiness” scale for *-mİşlİK*/*-mAzİİK*-constructions

This section considers specific morphosyntactic criteria of *-mİşlİK*/*-mAzİİK*-constructions in order to place individual findings on a continuum between an abstract nominal and a clausal status.

Step 1: An inventory of nominal and clausal categories

As a first step, we consider categories at the internal level, attempting to divide them into two types: (1) categories that we take to be indicators of abstract nominalisation, and (2) categories that we take to be indicators of a clausal structure.

On the side of abstract nominalisation, we consider

- quantifiers such as *tüm* or *birtakım* (e.g. *birtakım alışılmışlıklar yüzünden* ‘because of a number of habits, lit.: states of having become habituated’, Karasu 2010: 40),
- determiners such as *bir, bu, o, hiçbir* (e.g. *bir büyülenmişlik içinde* ‘in a state of enchantment/having-been-enchanted’, Necatigil 1993: 243),
- attributive modifiers (adjectives, relative clauses) (e.g. *beraberliğin ilk günlerindeki dokunulmamışlığı* ‘the untouchedness/the state of not having been touched on the first days of their togetherness’, Levi 2000: 305, containing the adjectiviser *-ki*; *unutulmuş ve terk edilmiş kitapların içine sinmiş okunmuşluğu* ‘the state of having been read that had sunk into books that had been forgotten and abandoned’, Kaygusuz 2006: 63, containing a clausal attribute),
- non-head nouns in compounded NPs (e.g. *yer yetmezliği* ‘shortage of space/space-lacking(-ness)’, Necatigil 1993: 332), and
- plural (e.g. *en eski efsanelerin bile çok daha eski yaşamışlıklardan doğduğunu* ‘that even the oldest myths had been born from much older life experiences/having-been-lived-s’, Kaygusuz 2006: 125f).

On the clausal side, we base our categorisation on the occurrence of

- voice, above all passive (e.g. *incinmişliğini saklamaya çalışır* ‘s/he tries to hide her hurt feelings/her having-been-hurt(-ness)’, Atasü 2007: 77),
- negation (e.g. *annesi tarafından seçilmemişliğin verdiği burukluğu* ‘the bitterness given by not having been chosen by his mother’, Kaygusuz 2006: 121);
- realisation of arguments of the verb (direct and other objects): nominal (*bir köşeye atılmışlığı* ‘its having been thrown into a corner’, Levi 2000: 143) or clausal (*ne yapacağını bilmezliğin nedeni* ‘the reason for [the state of] not knowing what s/he was going to do’, Levi 2000: 211), and
- adverbial modifiers: simple (*ayrıntı üzerinde durmazlığını gösterir* ‘it shows their general tendency of not paying much attention to detail’, Karasu 2010: 24) or clausal (*bir ottan bile güçsüzken yaşamışlığınızı* ‘your having survived even while weaker than a straw’, Kaygusuz 2006: 207).

Table 3 shows how the abovementioned categories distribute in the data. It needs to be pointed out that many individual findings display more than one category, with the effect that adding occurrences of individual categories would yield more than the sum of the findings.

Table 3: Morphosyntactic categories of the internal structure⁶

		<i>-mİşlik</i> (total: 100)	<i>-mAzlık</i> (total: 42)	Overall (total: 142)
abstract nominalisation	quantifiers	1 (1%)	1 (2%)	2 (1%)
	determiners	15 (15%)	10 (24%)	25 (18%)
	attributive modifiers	11 (11%)	4 (10%)	15 (11%)
	non-head nouns in compound NPs	1 (1%)	3 (7%)	4 (3%)
	plural	13 (13%)	6 (14%)	19 (13%)
clausal	passive	51 (51%)	23 (55%)	74 (52%)
	negation ⁷	10 (10%)	3 (7%)	13 (9%)
	arguments of the verb	6 (6%)	8 (19%)	14 (10%)
	adverbial modifiers	5 (5%)	3 (7%)	8 (6%)

One central result of this quantification that might point towards a clausal interpretation is the frequency of passive constructions; more than half of the *-mİşlik/-mAzlık*-constructions are passives. Verbal morphological negation is not a rare phenomenon either, occurring in about one tenth of the findings. Arguments of the verb are realised in about 10% of the cases, albeit three times as frequently in *-mAzlık* as in *-mİşlik*-constructions.

In terms of the type of modification, attributive modifiers (about 10% of all findings) are twice as frequent as adverbial modifiers (around 5%), a finding that points towards an abstract nominal interpretation, albeit not exclusively. Furthermore, there are quite a few instances of *-mİşlik/-mAzlık*-forms preceded by determiners (15% and 23% respectively) or marked by plural (some 13%). These are phenomena of nouniness that, taken together, seem to suggest that *-mİşlik/-mAzlık*-constructions cannot unequivocally be analysed as clausal.

Step 2: Constructing a scale

In the next methodological step, we develop a scoring procedure to attain a graded overview of the distribution of nominal and clausal categories. Starting from the observation that the findings in our—literary—corpus exhibit a high degree of individual variation and constructional creativity, we are interested in compressing this highly individual diversity onto something flat and simple: a scale. As a welcome methodological side effect of this procedure, we come across and identify (1) good candidates for a clausal status, (2) candidates for counter-examples, and (3) different types of overlaps and contradictions. On a constructed scale between abstract nomi-

6 Percentages are given in rounded figures in the entire paper, except for Table 1. Totals may thus not always correspond with the separate figures.

7 With *-mAzlık*-forms, this value refers to doubly negated forms (*-mAmAzlık*).

nal and clausal status, we assign one point to each of the following clausal categories: passive, negation, realised arguments of the verb, and adverbial modification. Conversely, we assign one negative point for each of the following nominal categories: quantifiers, determiners, attributive modifiers, non-head nouns in compound NPs, and plural. We assign a double value for complex realisations of any given category. By “complex”, we mean multi-word, internally hierarchic, and clausal elements. Table 4 presents the results obtained by this procedure. We will later return to findings that have proven particularly recalcitrant, for more discussion.

Table 4: Overall distribution on the nouniness/clausiness scale

Score	-2	-1	0	1	2	3
<i>-mİşlİK</i> (Σ: 100)	4 (4%)	10 (10%)	41 (41%)	34 (34%)	6 (6%)	5 (5%)
<i>-mAzlİK</i> (Σ: 42)	1 (2%)	10 (24%)	6 (14%)	20 (48%)	4 (10%)	1 (2%)
Overall (Σ: 142)	5 (4%)	20 (14%)	47 (33%)	54 (38%)	10 (7%)	6 (4%)

One overall result is that, when *-mİşlİK* and *-mAzlİK* are taken together, 18% of all findings obtain a result on the nouny side, as opposed to 49% on the clausy side. This is 2.7 times as many: a clear preponderance on the clausy side. A third obtain a neutral result. The results show a concentration of findings in a neutral to moderate area between nouniness and clausiness, with 85% of both *-mİşlİK* and *-mAzlİK* ending up between minus one and plus one, and a preponderance in this moderate area on the clausy side. It has to be said that constructions that unobtrusively “end up” in this area often do so as an effect of contradictory patterns neutralising each other. On the more extreme values of nouniness (minus two), we find some 4% of all findings; on the more extreme values of clausiness (plus two to plus three), we find 11% of all findings (11% of the *-mİşlİK*-constructions and 12% of the *-mAzlİK*-constructions)—pointing to a clausal preponderance in the extremer areas as well.

Before turning to the highly individual contradictory cases in Step 3, we will consider two of the extreme cases. Example (3) is one of the most strongly nominal candidates in the corpus, receiving two negative points overall: two for a complex adjectival attribute (*çok daha eski* ‘much older’), one for the plural marking, and one positive (on the clausal side) for being passive. (4) contains one of the most strongly clausal candidates in the corpus.

- (3) *Geceyarısın-a* *değın,* *dışarı-da-ki* *şenliğ-e*
 midnight-DAT until outside-LOC-ADJ party-DAT
aldır-madan *önce* *yaşam* *hikâye-sin-i* *anlat-tı,* *sonra*
 take notice-CV at first life story-POSS.3SG-ACC tell-PST.3SG then

efsane-sin-i. En eski efsane-ler-in bile çok daha eski
 myth-POSS.3SG-ACC SUP old myth-PL-GEN even much CMP old
yaşa-n-mışlık-lar-dan doğ-duğ-un-u, her efsane-nin
 live-PAS-MİŞLİK-PL-ABL be born-PAR-POSS.3SG-ACC every myth-GEN
bugün-ün metafor-u ol-duğ-un-u artık
 today-GEN metaphor-POSS.3SG be-PAR-POSS.3SG-ACC now
bil-iyor-du-m. Ama hikâye-si gerçek-ten
 know-PRS-COP.PST-1SG but story-POSS.3SG indeed
katlan-ıl-maz-dı.
 tolerate-PAS-NEG.AOR-COP.PST.3SG

‘Until midnight, without paying attention to the party outside, she had told her life story, then her myth. By this time, I knew that even the oldest myths were born from much older life experiences. But her life story was indeed intolerable.’ (Kaygusuz 2006: 125f)

- (4) *O gece, hararetili bir hikâye eşliğ-in-de*
 DEI night excited one story accompaniment-POSS.3SG-LOC
yüz-ün-e yüz-ler-ce göz oy-an
 face-POSS.3SG-DAT hundred-PL-EQU eye hollow.out-PAR
baba-sı-nın karşı-sın-da-yken, yalnızca ana
 father-POSS.3SG-GEN opposite-POSS.3SG-LOC-COP.CV only mother
baba-sı-nın değil, herkes-in; öbür çocuk-lar-in,
 father-POSS.3SG-GEN not everybody-GEN other child-PL-GEN
yöre-de-ki kadın-lar-in, erkek-ler-in, hiçbir zaman ve
 area-LOC-ADJ woman-PL-GEN man-PL-GEN any time and
hiçbir biçim-de sağ-göz-ün-e bak-ma-mış
 any form-LOC well-eye-POSS.3SG-DAT look-NEG-POST
ol-duğ-un-u, bir kez bile insan bakış-ı-yla
 be-PAR-POSS.3SG-ACC one time even man look-POSS.3SG-COM
karşılaş-ma-mışlığ-in-ı fark et-miş-ti.
 meet-NEG-MİŞLİK-POSS.3-ACC difference make-POST-COP.PST.3SG

‘That night, in front of his father, who, accompanied by an excited story, had cut out hundreds of eyes in his face, he had noticed that not only his parents, but not anybody, the other children, the women in the area, the men, had ever in any way looked into his well eye, that he had not even once met a human being’s gaze.’ (Kaygusuz 2006: 157f)

Example (4) receives a score of 3 for containing negation, an adverbial (*bir kez*), and a comitative complement (*insan bakışıyla*)—three verbal/clausal categories. As a *-mİşLIK*-based phrase, it additionally occurs in parallel construction with a *-mİş olduK*-based phrase, both clausal. The two constructions—*herkesin [...] hiçbir zaman ve hiçbir biçimde sağgözüne bakmamış olduğunu* ‘that not anybody had at any time and in any way ever looked into his well eye’ and *bir kez bile insan bakışıyla*

karşılaşmamışlığını ‘that he had not even once encountered a human gaze’—are parallel accusative-marked complement clauses, depending on the same matrix predicate *fark etmişti* ‘he had noticed’.

While providing a distributive overview of potentially clausal and nominal candidates, identifying some outlying candidates on both sides, and indicating the large area of undecidability between them, our scoring procedure reaches a serious limit in failing to show the internal variation and the contradictory character of many occurrences, as could also be seen in example (3).

Step 3: Co-occurrence of nominal and clausal categories

In the next step, we look at individual findings to see how bundles of nominal and clausal categories combine. We will look at constructions containing (1) passive, (2) negation, (3) arguments of the verb, and (4) adverbial modifiers. These are “candidates with a potential for clausiness”. Our aim is to check to what extent they still co-occur with any of the nominal categories, i.e. quantifiers, determiners, attributive modifiers, non-head nouns in compound NPs, and plurals. Table 5 summarises the results; Table 6 shows the number of those candidates that do not contain any of the nominal categories. The results in Table 6 cannot be directly derived from those in Table 5 since, again, many occurrences display more than one category.

Table 5: “Candidates with a potential for clausiness” containing nominal categories

		Q	D	A	N	P
Passive (74/142)	<i>-mİşlİK</i>	-	10	7	1	10
	<i>-mAzlİK</i>	1	4	3	-	4
Negation (13/142)	<i>-mİşlİK</i>	-	3	1	-	-
	<i>-mAzlİK</i>	-	1	1	-	-
Arguments (14/142)	<i>-mİşlİK</i>	-	1	-	-	-
	<i>-mAzlİK</i>	-	1	-	-	-
Adverbials (8/142)	<i>-mİşlİK</i>	-	-	-	-	-
	<i>-mAzlİK</i>	-	-	-	-	-

Q = Quantifier, D = determiner, A = Attributive modification, N = Non-headnouns, P = Plural

Table 6: “Candidates with a potential for clausiness” not containing any nominal categories

	<i>No nominal categories</i>
Passive (74/142)	42 (56%)
Negation (13/142)	8 (62%)
Arguments (14/142)	12 (85%)
Adverbials (8/142)	8 (100%)

(1) A look at the 74 passive forms (52% of the overall sample) reveals that they can realise all of the nominal categories, sometimes in combination. On the other side, it has to be noted that 42 of the passive constructions (56%) do not realise any nominal categories at all. Furthermore, on the clausal side, two passive *-mİşLIK*-constructions realise a dative complement of the verb, and five passive constructions are adverbially modified. This mixed result in the overall large group of passives suggests that as a phenomenon closer to the derivational domain, passivisation turns out to play a lesser role when it comes to distinguishing between abstract nominalisation and clausal subordination, and thus cannot on its own serve as a criterion for clausiness.

Rather, the question has to be approached of whether (a) the co-occurring nominal categories cancel out the clausal tendency given by the passive, resulting in an overall abstract-nominal status, or whether (b) the contradiction should be accepted as a particular characteristic of *-mİşLIK*/*-mAzLIK*-constructions, an indicator of their transitional or in-between character. (5) is a passive example with nevertheless strong nouny characteristics. Based on the passive form *bilinmezlik*, it is used in construction with a clausal attribute (*keşfedilemeyecek* ‘that will/can not be discovered’) and a determiner (*hiçbir* ‘none at all’).

- (5) *Eriş-il-me-yecek* *hiçbir hedef, ulaş-ıl-ma-yacak* *hiçbir*
 attain-PAS-NEG-FUT.PAR any goal reach-PAS-NEG-FUT.PAR any
doruk, keşfed-il-eme-yecek *hiçbir bil-in-me-zlik*
 mountain top discover-PAS-MOD-NEG-FUT.PAR any know-PAS-MAZLIK
yok-tu *kardeş-ler için...*
 nonexistent-COP.PST.3SG sibling-PL for
 ‘There was no unreachable goal, no unattainable mountain top, no undiscoverable unknown for the siblings.’ (Atasü 2007: 103)

(2) The negated 13 constructions (9% of the overall sample) realise determiners in slightly less than a third of the cases and attributive modifiers in slightly less than a sixth. Thus, negation, while more of a “truly” inflectional category, also behaves somewhat ambiguously: it still allows the realisation of nominal categories some 40% of the cases, albeit in a lesser variety than passive seems to do. Example (6) shows a negated construction containing a determiner. (7), by contrast, shows a negated *-mİşLIK*-construction that is modified by means of an adverbial:

- (6) *Kanımca* *şair-i, birçok* *“Hisarcı-lar”-dan* *daha* *taze,*
 in my opinion poet-ACC many Hisarci-PL-ABL CMP fresh
daha canlı yap-an, ön plan-a al-an *da bu*
 CMP lively make-PAR foreground-DAT take-PAR also DEI
yıpran-ma-mışlık; ne hav-ı dökül-müş, ne
 fade-NEG-MIŞLIK NEG floss-POSS.3SG fall-POST NEG
iplik-ler-i epr-i miş bu farklı kumaş,
 thread-PL-POSS.3SG become threadbare-POST DEI different material

bu zanaat ol-sa gerek.
 DEI craft be-CD.3SG necessary

'In my opinion, what makes the poet fresher, more lively than a number of "Hisarcis" and what takes him to the foreground must be this non-fadedness, this different material, which has neither lost its floss nor become threadbare, and this craft.' (Necatigil 1993: 188)

- (7) *Baba-m kardeş-i-nin apansız*
 father-POSS.1SG brother-POSS.3SG-GEN suddenly
çık-a-gel-me-sin-e sevin-eme-diğ-in-i,
 appear-CV-PV-VN-POSS.3SG-DAT be happy-NEG.MOD-PAR-POSS.3SG-ACC
büyük bir kavun parça-sın-ı iştah-la
 large one honey melon piece-POSS.3SG-ACC appetite-INS
ağz-in-a at-ar-ken söyle-miş-ti.
 mouth-POSS.3SG-DAT throw-AOR-CV say-POST-COP.PST.3SG
Mülk-lerin-e ortak çık-ma-sın-dan çok,
 property-POSS.3PL-DAT partner come out-VN-POSS.3SG-ABL much
anne-si taraf-in-dan seç-il-me-miş-liğ-in
 mother-POSS.3SG side-POSS.3SG-ABL choose-PAS-NEG-MİŞLİK-GEN
ver-diğ-i burukluğ-u bir kez daha
 give-PAR-POSS.3SG bitterness-POSS.3SG one time more
canlandır-diğ-in-dan olsa gerek, kardeş-in-i
 revive-PAR-POSS.3SG-ABL be-CD.3SG necessary brother-POSS.3SG-ACC
al-abil-diğ-in-e soğuk karşıla-mış-ti.
 take-MOD-PAR-POSS.3SG-DAT cold receive-POST-COP.PST.3SG
 'My father had said that he was not happy about his brother's suddenly showing up, while devouring a large piece of honey melon with great appetite. He had received his brother as coldly as possible; it must have been because the bitterness caused in him by not having been chosen by his mother once more was revived, rather than because he had come for his share of the property.' (Kaygusuz 2006: 121)

(3) Among the 14 instances that realise arguments of the verb (9% of the overall sample), two are preceded by a determiner;⁸ both examples are from the same author. In this group, 12 (85%) occurrences lack any of the investigated nominal categories at the internal level. Example (8) shows how a determiner may nonetheless occur in a construction that realises (even clausal) arguments of the verb.

8 Ross (2004 [1973]: 372) discusses a similar phenomenon for English constructions that, while being rather far removed from nouniness, still feature determiners. See his example *That having been followed for years must have been nerve-wracking.*

- (8) *Hayat-ım-da yer ed-eceğ-in-i çok iyi*
 life-POSS.1SG-LOC place do-FUT.PAR-POSS.3SG-ACC very well
bil-diğ-im bir insan-a biraz da hazırlık-sız
 know-PART-POSS.1SG one person-DAT a little also preparation-PRIV
yakala-n-ma-m değil-di ama sadece bu ne
 catch-PAS-VN-1SG not-COP.PST.3SG but only DEI what
yap-acağ-ın-ı bil-meziğ-in neden-i, gerçek
 do-FUT.PAR-POSS.3SG-ACC know-MAZLIK-GEN reason-POSS.3SG real
bir karşılař-ma-dan duy-duğ-um korku-ydu da
 one meet-VN-ABL feel-PART-POSS.1SG fear-COP.PST3SG also
galiba aynı zaman-da.
 probably same time-LOC
 ‘But the reason for this state of not knowing what she was going to do was not only being caught somewhat unprepared by a person whom I very well knew was going to play a role in my life; it was also the fear of a real encounter at the same time, probably.’ (Levi 2000: 211)

(4) Those constructions that contain adverbial modifiers (5% of the overall sample) do not realise any of the nominal categories. These constructions, together with those that realise arguments, while relatively small in number, might be considered much clearer candidates for clausal status. Examples (9) and (10) contain constructions with complex, i.e. clausal and/or hierarchically layered, adverbial modification.

- (9) *San-ır-ım ucuz-cu-luğ-um-la hor gör-üyor-sunuz*
 assume-AOR-1SG cheap-DER-DER-N-POSS.1SG-INS despise-PRS-2PL
beni. Oysa bir düşün-ün, ıssız bir ortam-da
 1SG.ACC but one think-IMP.POL deserted one environment-LOC
bir ot-tan bile güç-süz-ken yaşa-mışlığ-ınız-ı
 one straw-ABL even force-PRIV-CV live-MİŞLIK-POSS.2PL-ACC
kanıtla-yacak herhangi bir nesne-nin anlam-ın-ı.
 prove-FUT.PAR any kind one object-GEN meaning-POSS.3SG-ACC
 ‘I assume, you despise me because of my occupation with cheap things. But think about it; the meaning of any kind of object that proves you survived in a deserted environment while weaker than a straw.’ (Kaygusuz 2006: 207)
- (10) *Yer-in-e gene ağır, gene kahred-en,*
 place-POSS.3SG-DAT again oppressive again overwhelm-PAR
bela-lı bir korku, bir dünya-nın orta-sın-da
 plague-DER one fear one world-GEN middle-POSS.3SG-LOC
tek baş-ın-a kal-mışlığ-in boş-luğ-u,
 single head-POSS.3SG-DAT remain-MİŞLIK-GEN empty-DER.N-POSS.3SG

yalnızlıđ-ı, çaresizliđ-i gel-di otur-du.
 loneliness-POSS.3SG helplessness-POSS.3SG come-PST.3SG sit-PST.3SG

'In its place established itself a fear, oppressive again, overwhelming again, painful, an emptiness, loneliness, helplessness of having remained alone in the world.' (Kemal 1984: 291)

Taken together, the observed phenomena suggest that *-mİŖlIK/-mAzlIK*-constructions are far from being analysable as unequivocally clausal. However, a categorisation by clausal phenomena suggests that there may be specific limited areas where clausal characteristics are more pronounced than in others: in the few cases in which a construction realises an argument of the verb or is adverbially modified, it seems to have crossed a boundary towards clausiness—something that cannot as easily be claimed for passive or even negated constructions, which leave more space for transition and variation.

6. Genitive attributes or genitive subjects?

There are some interesting cases of categorial ambiguity in the area of genitive NPs that are realised inside the *-mİŖlIK/-mAzlIK*-construction. Genitive NPs can, semantically, be either genitive attributes in a nominal possessive construction or genitive subjects in a clausal construction; formally, there is much less of a difference.

In example (11), the bare *-mİŖlIK*-construction *Butyakengo'nun görmüş geçirmişliđi* can be interpreted as either clausal ('that Butyakengo has experienced life') or nominal ('Butyakengo's life experience'). However, a look at context reveals that the passage is not about the fact that the character in question experienced life, but rather is about his life experience, which, as a whole, is transferred to the next generation. The superordinate predicate (*eklenirmiş*) is a verb that describes the handling of an object, and differs from the *verba dicendi* in examples (12) and (13) below. We accordingly interpret this genitive NP as a genitive attribute in a nominal possessive construction.

- (11) *Latife KeŖal'in söyle-diđ-in-e bak-ıl-ır-sa, her*
 Latife KeŖal-GEN say-PART-POSS.3SG-DAT look-PAS-AOR-COP.CD every
Çingene-nin yüz-ün-de yedi kilit var-mış.
 gypsy-GEN face-POSS.3SG-LOC seven lock existent-COP.EVID
Anne-si ya da baba-sı öl-ünce kilit-ler-in
 mother-POSS.3SG or father-POSS.3SG die-CV lock-PL-GEN
tüm-ü aç-ıl-ır-mış. Butyakengo-nun gör-müş
 all-POSS.3SG open-PAS-AOR-COP.EVID Butyakengo-GEN see-POST

geçir-mişliğ-i *ekle-n-ir-miş* *kalp-lerin-e*
 experience-MİŞLİK-POSS.3SG add-PAS-AOR-COP.EVID heart-POSS.3PL-DAT
 ‘Considering what Latife Keşal said, there were seven locks in every Gypsy’s face.
 When his/her mother or father died, all these locks would open. Butyakengo’s life
 experience would be added to their hearts.’ (Kaygusuz 2006: 112)

Examples (12) and (13) have a more clausal character. This can at the present stage be attributed to the observation that they are subordinated to a *verbum sentiendi* (*gözünden silinmişti* ‘it had been wiped from their sight’) and a *verbum dicendi* (*övünenleri* ‘those who bragged’) respectively, expressing propositions approachable from a truth-valued perspective—a preliminary impression. We will present a more detailed analysis of the semantics of matrix predicates in Section 8.

- (12) *Ortalık tıka basa insan-la dolmuştu artık, bu yüzden*
 place cramfull people-INS fill-POST-COP.PAST now for this reason
hem ayı-nın hem de berber-in herkes-ten ayrı
 both bear-GEN and also barber-GEN everybody-ABL separate
düş-müşlüğü *cam-da-ki* *yaşlı-lar-in* *göz-ün-den*
 fall-MİŞLİK-POSS.3SG window-LOC-ADJ old people-PL-GEN eye-POSS.3SG-ABL
sil-in-miş-ti.
 wipe-PAS-POST-COP.PST.3SG
 ‘The place had filled up to the brim with people; for this reason, the fact that both the bear and the barber had separated from everybody had been wiped from the sight of the old people at the window.’ (Toptaş 1995: 223)

- (13) *Bu 'tercih-leri', her şey-in yerli yerine*
 DEI preference-POSS.3PL every thing-GEN in its right place
otur-muşluğ-u-yla *övün-en-ler-i* *elbette rahatsız*
 settle-MİŞLİK-POSS.3SG-INS boast-PART-PL-ACC of.course uneasy
et-miş. Bu rahatsız-lığ-ı anla-mak zor değil-di.
 do-POST DEI uneasy-DER-ACC understand-VN difficult not-COP.PST.3SG
Hayat-ların-ı o 'yerleş-mişlik' üzer-in-e
 live-POSS.3PL-ACC DEI settle-MİŞLİK upper.surface-POSS.3SG-DAT
kur-an-lar, kendi-lerin-e, gerçekleştire-eme-dik-leri,
 found-PART-PL self-POSS.3PL-DAT realise-NEG.MOD-PART-POSS.3PL
hiçbir zaman gerçekleştire-eme-yecek-leri düş-ler-i
 any time realise-NEG.MOD-FUT.PAR-POSS.3PL dream-PL-ACC

göster-en-ler-i kabullen-mek-te elbette zorlan-ır-dı.
 show-PART-PL-ACC accept-VN-LOC of course have difficulty-AOR-COP.PST
 ‘These ‘preferences’ of theirs had, of course, disturbed the ease of those who boasted that everything was settled in its right place. It was not difficult to understand this unease. Those who built their lives on that settledness, found it hard, of course, to accept those who showed them their never-realised, never-to-be-realised dreams.’
 (Levi 2000: 301)

In some cases, the distinction is difficult to maintain, as when the interpretation hinges on the internal versus the external perspective on the event in question. The problem in example (14) is in the last sentence: *Yolcular kayınların çürümüşlüğünden okurlardı zamanı*. The point that cannot easily be decided upon is whether the travellers read the time from the *fact* that the trees are rotten (this could be propositional) or from the quality and extent of the rottenness (this could be an indicator of a state-of-affairs quality). The clausal versus nominal character of this construction thus hinges on this textual-constellational interpretation.

- (14) *Son-un-da, hayvan-la karşılaş-tık-ları yer-e*
 end-POSS.3SG-LOC animal-INS meet-PART-POSS.3PL place-DAT
yine var-dı-lar. Geniş bir koridor-un baş-m-da
 again arrive-PST-3PL wide one corridor-GEN head-POSS.3SG-LOC
dik-il-en iki ağac-ı hemen tanı-mış-lar-dı.
 plant-PAS-PART two tree-ACC immediately recognise-POST-PL-COP.PST.3SG
Dal-ları-nın çoğ-u kuru-muş, kemirgen
 branch-POSS.3PL-GEN many-POSS.3SG dry-POST rodent
yılgın-ı iki kayın. Yolcu-lar kayın-lar-ın
 befallen-PSS.3SG two beech traveller-PL beech-PL-GEN
çürü-müşlüğü-ün-den oku-r-lar-dı zaman-ı.
 rot-MİŞLİK-POSS.3SG-ABL read-AOR-3PL-COP.PST.3SG time-ACC
 ‘Finally, they reached the place again where they had encountered the animal. They had immediately recognised the two trees that stood planted at the opening of a broad corridor. Most of their branches were rotten, two beeches befallen by rodents. It was from the rottenness of the beeches that the travellers read the time.’ (Kaygusuz 2006: 192)

The genitive NP *o kadının* ‘that woman’ in the following example (15) is more clearly a subject, possibly due to the use of a parallel clausal construction based on the canonical subordinator *-mA*: *o kadının ulaşılmazlığı [karşısında]* ‘[in view of] that woman’s unattainability/not-being-attainable’ and *[o kadının] o dostu tercih etmesi karşısında* ‘in view of her preferring that friend / the fact that she preferred that friend’:

- (15) *O anla-şıl-abil-ir kiskançlık, bu sınır-da*
 DEI understand-PAS-MOD-AOR.3SG jealousy DEI border-LOC
başlı-yor-du işte. Bir dost-un bir başka-sın-a
 begin-PRS-COP.PST.3SG PTC one friend-GEN one other-POSS.3SG-DAT
kap-tır-ıl-diğ-i yer-di sanki bura-sı.
 snatch-CAUS-PAS-PAR-POSS.3SG place-COP.PST.3SG as if here-POSS.3SG
Sorun, o kadın-ın ulaş-ıl-mazlığı, o dost-u
 problem DEI woman-GEN attain-PAS-MAZLIK-POSS.3SG DEI friend-ACC
tercih et-me-si karşı-sın-da hissed-il-en-ler-in
 preference do-VN-POSS.3SG opposite-POSS.3SG-LOC feel-PAS-PAR-PL-GEN
bu küçük yenilgi-ye getir-ebil-ecek-leri
 DEI small defeat-DAT bring-MOD-FUT.PAR-POSS.3PL
düşün-ül-düğ-ün-de, daha da iç-in-den
 think-PAS-PAR-POSS.3SG-LOC more also inside-POSS.3SG-ABL
çık-ıl-amaz bir hal-e dönüş-ebil-ir-di.
 get out-PAS-NEG.MOD.AOR one situation-DAT turn-MOD-AOR-COP.PST.3SG
 ‘That understandable jealousy started at this point. This here was, as it were, a place taken by one friend from the other. The problem could, if one thought of what the things that were felt in face of that woman’s unattainability, her preferring that friend, might add to this small defeatedness, turn into an even more hopeless situation.’ (Levi 2000: 138)

7. Postpositional constructions

-mİşlik/*-mAzlık*-constructions with overall clausal characteristics occasionally appear within a postpositional phrase (18 findings in our data), distributed among the following postpositions: *ardına* ‘behind’, *bağlı* ‘dependent on’, *içinde* ‘inside’ (2), *içinden* ‘inside’, *karşısında* ‘against’ (3), *oranında* ‘at the level of’, *öte* ‘beyond’ (2), *sonra* ‘after’ (2), *üzerine* ‘about’, *yanından* ‘from the side of’, *yüzünden* ‘because of’ (3). Some of these findings are reminiscent of complex converbs based on *-DIK* or *-mA*. Examples (16) to (18) are among those that tend towards the clausal side. The construction in example (16), based on *-mAzlık karşı-sında*, has a clausal character in that it governs a direct object: *teselli* ‘comfort’. The genitive noun *içinin* can be analysed as a subject. For the use of *-mAzlık karşı-sında*, cf. example (15) above.

- (16) *Behiye güz akşamüstleri-nin insan-ı sarmala-yan*
 Behiye autumn late afternoon-GEN man-ACC wrap-PAR
o efendi ılık-ın-da, iç-i-nin teselli
 DEI gentle mildness-POSS.3SG-LOC inner-POSS.3SG-GEN comfort
kabul et-mezliği karşı-sın-da daha
 acceptance do-MAZLIK-POSS.3SG opposite-POSS.3SG-LOC more

da kütleleş-erek, habire yürü-yor. Bir kütle işte
 also become a mass-CV continuously walk-PRS.3SG one mass PTC
o. Kara bir asfalt taş-ı. Kendi iç-in-e
 DEI black one asphalt stone-POSS.3SG self inside-POSS.3SG-DAT
otur-muş. Asfalt taş-ı. Taşş.
 sit-POST asphalt stone-POSS.3SG stone
 ‘Behiye is continuously walking, in the gentle mildness of the autumn late afternoon that wraps you up, becoming increasingly a mass in face of the fact that her soul did not accept any comfort. Well, it was a mass. A black asphalt stone. It had settled inside her. An asphalt stone. Stooooone.’ (Magden 2006: 13–14)

Example (17), built on *öte* ‘beyond’, is somewhat clausal in being adverbially modified by *zaman içinde* ‘in the course of time’.

- (17) *Bu koridor-un sol taraf-ın-da terk ed-il-miş,*
 DEI corridor-GEN left side-POSS.3SG-LOC leaving do-PAS-POST
harabe bir ev var-dir. Zaman iç-in-de
 ruined one house existent-COP.PRS.3SG time inner-POSS.3SG-LOC
yık-ıl-mış, parçalan-mışlık-tan öte bir talan
 demolish-PAS-POST fall to pieces-MİŞLİK-ABL beyond one plundering
artığ-ı gibi dur-an kapı, pencere ve
 leftover-POSS.3SG like stand-PAR door window and
cam-lar-dan mahrum bir bina, oy-ul-muş
 windowpane-PL-ABL devoid one building hollow out-PAS-POST
on-lar-ca göz-den kazık-lar yardım-ı-yla ayak-ta dur-an
 ten-PL-EQU eye-ABL pile-PL help-poss.3SG-INS foot-LOC stand-PAR
bir cesed-i çağırır-ır.
 one corpse-ACC evoke-AOR.3SG
 ‘On the left hand side of this corridor was an abandoned, derelict house. Apart from having fallen derelict [lit.: having been demolished and fallen apart] over the years, a building that stood like the leftover of a plundering and was devoid of door, windows and windowpanes, it evokes a corpse standing on its feet with the help of piles of dozens of hollowed-out eyes.’ (Çelik 2011: 72)

Example (18) is the more clausal one of two findings based on *-mİşliktAn sonra*.

- (18) *Bu-nun-la ne yap-mak iste-diğ-in-i de uzun*
 DEI-GEN-INS what do-VN want-PAR-POSS.3SG-ACC also long
süre sonra anla-yacak-tı-m. Kendi-sin-i
 duration after understand-FUT-COP.PST-1SG self-POSS.3SG-ACC
biraz daha iyi tanı-ma-ya başla-yınca. Onca
 a little CMP good know-VN-DAT begin-CV dozens

ihamet-ten, *yalnız* ***birak-il-mışlık-tan*** *sonra,*
 betrayal-ABL alone leave-PAS-MİŞLİK-ABL after
övgü-ler-i-yile *Berti-den* *küçük* *bir* *intikam*
 compliment-PL-POSS.3SG-INS Berti-ABL small one revenge
al-iyor-du *asl-ın-da,* *onu* *gizli-den*
 take-PRS-COP.PST.3SG essential-POSS.3SG-LOC DEI-ACC secret-ABL
gizli-ye, *büyük* *bir* *usta-lık-la,* *değer* *ver-diğ-i*
 secret-DAT great one master-DER-INS value give-PAR-POSS.3SG
bir *insan* *karşı-sın-da* *mahcup* *ed-iyor,*
 one person front-POSS.3SG-LOC ashamed make-PRS.3SG
küçük *düş-ür-üyor-du.*
 small fall-CAUS-PRS-COP.PST.3SG

‘And I would understand much later what she wanted to do with this. When getting to know her a little better. After dozens of betrayals, after having been left alone so many times, she was in fact taking a small revenge on Berti with her compliments. She was, secretly, with great mastery, causing him to feel small, shaming him in front of a person he valued.’ (Levi 2000: 157)

These constructions, based on *-mAzLIĞI karşısında* ‘in the face of not ... -ing’, *-mİşliktAn öte* ‘beyond having ... -ed’ and *-mİşliktAn sonra* ‘after having ... -ed’ respectively, are few and far between, and are not nearly as frequent or productive as, for example, *-mAİ karşısında*, *-mAİndan öte*, or *-DİktAn sonra*. Where they do exist, they may in some cases unfold categories of a clausal structure, while others remain more on the nominal side. (16–18) are our best candidates.

8. Semantic characteristics

8.1. The subordinate level

In aspectual terms, *-mİşLIK* expresses a posttransformative state of events, while *-mAzLIK* expresses a negated intraterminal state of events. One interesting result of our semantic analysis is that of the 142 findings, 94 (66%) are lexically based on verbs that describe negative cognitive or emotional situations or experiences: 73 (73%) and 21 (50%) respectively. We will call them “predications with negative content”; they comprise 56 types in our data (67% of our 84 types overall). What is semantically at hand here to a large extent are (1) negatively experienced postterminal states after an action or event (*-mİşLIK*) and (2) internally perceived, continuing states of actions that fail to happen (*-mAzLIK*). The remaining 49 tokens (35%), or 31 types (37%), have a more neutral meaning.⁹ Figure 1 is a full alphabetical list of the forms connoting negative situations.

9 There are four types that appear in both negative and neutral meanings, depending on context.

<i>aldatılmışlık</i> ‘state of having been duped’	<i>elden ayaktan düşmüşlük</i> ‘state of having lost one's force’	<i>kocamışlık</i> ‘state of having reached old age’ (2)
<i>aşağılanmışlık</i> ‘state of having been humiliated’	<i>eprimişlik</i> ‘threadbareness’	<i>kopmuşluk</i> ‘uprootedness’
<i>atılmışlık</i> ‘state of having been thrown away’	<i>eskimişlik</i> ‘state of having become old’ (3)	<i>küçülmüşlük</i> ‘downsizedness’
<i>ayrı düşmüşlük</i> ‘state of having fallen separate’	<i>ezilmişlik</i> ‘oppression, state of having been oppressed’ (2)	<i>metod bilmezlik</i> ‘state of not knowing any method’
<i>bırakılmışlık</i> ‘abandonedness/ state of having been abandoned’ (2)	<i>geç kalmışlık</i> ‘belatedness, delay’ (2)	<i>önlenmezlik</i> ‘unpreventability/ state of not being preventable’
<i>boşalmışlık</i> ‘emptiedness/ state of having been emptied’	<i>gecikmişlik</i> ‘belatedness, delay’	<i>parçalanmışlık</i> ‘fragmentedness’ (3)
<i>boşvermişlik</i> ‘indifference/ state of having ceased to care’ (2)	<i>görmemişlik</i> ‘unmannerliness’	<i>sarsılmazlık</i> ‘unshakability’
<i>bozulmuşluk</i> ‘state of having been destroyed’	<i>güvenmemelik</i> ‘distrust/ state of being distrustful’	<i>seçilmemişlik</i> ‘state of not having been chosen’
<i>cefa çekmişlik</i> ‘state of having suffered’	<i>içinden çıkılmazlık</i> ‘state of inability to settle something’	<i>tek başına kalmışlık</i> ‘state of having been left alone’
<i>çökmüşlük</i> ‘state of having collapsed’	<i>incinmişlik</i> ‘state of having been hurt’	<i>terk edilmişlik</i> ‘state of having been left’ (13)
<i>çürümüşlük</i> ‘state of rottenness’	<i>itilmişlik</i> ‘state of having been repressed’	<i>tükenmişlik</i> ‘exhaustion’
<i>dayanılmazlık</i> ‘quality of being unbearable’	<i>kabul edilmezlik</i> ‘unacceptability/ quality of not being acceptable’ (2)	<i>tükenmiş bitmişlik</i> ‘exhaustion’
<i>değerbilmezlik</i> ‘unappreciativeness’	<i>kaçınılmazlık</i> ‘unavoidability’ (7)	<i>ulaşılmazlık</i> ‘unattainability/ quality of not being attainable’
<i>dışlanmışlık</i> ‘experience of having been excluded’ (4)	<i>kafayı bulmuşluk</i> ‘state of drunkenness’	<i>yalnız bırakılmışlık</i> ‘state of having been left alone’
<i>doğmamışlık</i> ‘quality of not having been born’ (3)	<i>kapatılmışlık</i> ‘closedness’	<i>yarım kalmışlık</i> ‘state of having remained half’ (2)
<i>dokunulmazlık</i> ‘untouchability/ quality of being untouchable’ (2)	<i>kaybolmuşluk</i> ‘lostness/ state of being lost’	<i>yaşanmamışlık</i> ‘state of not having been experienced or lived’
<i>doymamışlık</i> ‘dissatisfaction’	<i>kendine çekilmişlik</i> ‘state of having become internally withdrawn’	<i>yenilmişlik</i> ‘defeatedness/ state of having been defeated’
<i>doymazlık</i> ‘insatiability, state of not being saturated’	<i>kirlenmişlik</i> ‘state of having been dirtied’	<i>yetmezlik</i> ‘insufficiency/ quality of not being sufficient’
<i>düşmüşlük</i> ‘state of having fallen’	<i>kısıtılmışlık</i> ‘state of having been cornered’	<i>yıkılmışlık</i> ‘state of having been torn-down’
		<i>yıpranmışlık</i> ‘state of having thinned/ faded’

Figure 1: List of *-mişlik/-mAzlık*-predications expressing negative situations

Examples (19) to (21) illustrate how some expressions of negative states are embedded by means of *-mişlik* in contexts that in addition provide a negatively connoted semantic surrounding.

- (19) *Öyle, gene de gıpta et-tiğ-im bir ölüm-dür*
 that way again also envy do-PAR-POSS.1SG one death-COP.PRS
Tahir-in ölüm-ü. Hiç değil-se temiz ölüm.
 Tahir-GEN death-POSS.3SG at all not-COP.CD clean death
Yatak-lar-a bağla-n-mış, ağır ve şifa-sız bir
 bed-PL-DAT bind-PAS-POST heavy and cure-PRIV one
hastalık-la, koca-mışlığ-in, el-den ayak-tan düş-müşlüğü-ün
 illness-INS get old-MİŞLIK-GEN hand-ABL foot-ABL fall-MİŞLIK-GEN
sefalet-in-de, uzun süre, madde ve ruh azap-lar-ı
 misery-POSS.3SG-LOC long duration material and soul torment-PL-ACC
çek-erek, acınma-lar-a konu, ne-ler-den sonra
 draw-CV compassion-PL-DAT subject what-PL-ABL after
öl-se-ydi daha mı iyi-ydi?
 die-CD-COP.PST.3SG CMP Q good-COP.PST.3SG
 ‘That way, it nevertheless was a death I envied, Tahir’s death. A clean death at least. If he had died bedridden, with a heavy and incurable illness, in the misery of old age and having lost all his force, suffering material and spiritual torment, subject to compassion, and after what-all, would it have been better?’ (Necatigil 1993: 138)

- (20) *Birtakım şey-ler yap-ıl-a-gel-ir ama yap-ıl-ma-malı,*
 some thing-PL do-PAS-CV-PV-AOR.3SG but do-PAS-NEG-MOD.3SG
birtakım yer-ler-e gid-il-ebil-ir ama
 some place-PL-DAT go-PAS-MOD-AOR.3SG but
gid-il-me-meli-dir. Bun-lar-ı yap-mak-tan, ora-lar-a
 go-PAS-NEG-MOD-COP.PRS.3SG DEI-PL-ACC do-VN-ABL there-PL-DAT
git-mek-ten kendi-lerin-i al-ama-dık-ları hal-de,
 go-VN-ABL self-POSS.3PL-ACC take-NEG-MOD-PAR-POSS.3PL situation-LOC
ard-ın-dan –ya da daha “iç-in-de-yken”–
 back-POSS.3SG-ABL or also still inside-POSS.3SG-LOC-COP.CV
utanç benzer-i bir sıkıntı duy-mağ-a
 shame similar-POSS.3SG one distress feel-VN-DAT
başla-r, bir düş-müşlük, bir küçül-müşlük
 begin-AOR.3SG one fall-MİŞLIK one become small-MİŞLIK
iğrenti-sin-e kap-ıl-ır bu kimse-ler.
 disgust-POSS.3SG-DAT snatch-PAS-AOR.3SG DEI person-PL
 ‘A number of things have always been done but must not be done; a number of places can be gone to but must not be gone to. Although they cannot prevent themselves from doing those things, from going there, afterwards—or while still “inside it”—they start feeling a distress similar to shame, get caught in a state of having fallen, a disgust of having become downsized, these people.’ (Karasu 2010: 81)

- (21) *Değişim ben-im! Nefise-nin ölüm-ün-den elli yıl sonra,*
 change 1SG-1SG Neise-GEN death-POSS.3SG-ABL fifty year after
parçalan-mışlık-lar-ım fiziksel yok-ol-uş
 fall to pieces-MİŞLİK-PL-POSS.1SG physical nonexistent-become-VN
yer-in-e, yeni bir tümlenme-ye yönel-ebil-iyor-sa,
 place-POSS.3SG-DAT new one completion-DAT direct-MOD-PRS-COP.CD.3SG
sav-ım-da haklı-yım!
 claim-POSS.1SG-LOC right-1SG
 ‘I am the change! If fifty years after Nefise’s death, my fragmentations, instead of being a physical annihilation, can be directed towards a new completion, I will be right in my claim!’ (Atasü 2007: 280)

Some of the items in Figure 1 do not necessarily indicate negative states; rather, a look at context provides more clarification. Example (22) shows how the form *doğmamışlık* ‘quality of not having been born’ acquires a negative reading in context.

- (22) *Deniz-den devşir-diğ-i bellek-le, yer-e bas-ma-yan*
 sea-ABL gather-PAR-POSS.3SG memory-INS ground-DAT step-NEG-PAR
bir kişilik-te bütün bir ömür idare
 one personality-LOC entire one lifetime management
ed-ebil-ir-di. San-ıyor-um bu-nun olanaksızlığı-ın-ı
 do-MOD-AOR-COP.PST.3SG assume-PRS-1SG DEI-GEN impossibility-POSS.3SG-ACC
bil-diğ-in-den öldür-eme-di ahtapot-u.
 know-PAR-POSS.3SG-ABL kill-NEG-MOD-PST.3SG octopus-ACC
Kol-lar-ın-ı vantuzla-yan yaratı-ğ-a bak-inca, onu
 arm-PL-POSS.3SG-ACC suck-PAR creature-DAT look-CV DEI.ACC
doğ-ma-mışlık duygu-sun-dan kurtar-acak gerçeküstü
 be born-NEG-MİŞLİK feeling-POSS.3SG-ABL save-FUT.PAR surreal
hiçbir dayanak bul-ama-mış-tı.
 any support find-NEG-MOD-POST-COP.PST.3SG
 ‘With the memory gathered from the sea, given a personality that did not step on the ground, s/he could manage a whole lifetime. I assume that because she knew of this impossibility, s/he could not kill the octopus. When looking at the creature, which was sucking its arms, s/he had not been able to find any surreal support to save him/her from the feeling of not having been born.’ (Kaygusuz 2006: 111)

Another interesting result, as can be seen in the list in Figure 1 above, is that of the 94 ‘verbs with negative content’, 52 are passives (56%); 15 of the *-mAzLIK*-constructions with negative content (73%) and 37 *-mİŞLİK*-constructions with negative content (51%). Taken together, these two separate results indicate a large amount of ‘experiences of passive negative states’. See Table 7 and example (23) (also reconsider the forms *küçülmüşlük* ‘the state of having been downsized’ and *par-*

çalanmışlıklarım ‘my states of having been fragmented’ in examples 20 and 21 above).

Table 7: “Verbs with negative content” and “passive verbs with negative content”

	Σ	neutral content	negative content	
			Σ	passives with negative content
<i>-mİşlik</i>	100	27 (27%)	73 (73%)	37 (51%)
<i>-mAzlık</i>	42	21 (50%)	21 (50%)	15 (72%)
Σ	142	48 (34%)	94 (66%)	52 (57%)

- (23) *Ne yap-acağ-ın-ı kestir-eme-yen bekçi,*
 what do-FUT.PAR-POSS.3SG-ACC predict-NEG.MOD-PAR watchman
kıs-tır-ıl-mışlık duygu-su-nun öfke-yle
 squeeze-CAUS-PAS-MİŞLİK feeling-POSS.3SG-GEN anger-COM
buluş-tuğ-u nokta-dan ona bak-ıyor-du hâlâ.
 meet-PAR-POSS.3SG point-ABL DEL.DAT look-PRS-COP.PST.3SG still
 ‘The watchman, who could not predict what he would do, was still looking at him from a point at which a feeling of having been cornered met with anger.’ (Toptaş 1995: 108)

To sum up, our analysis shows that the verbal predication followed by *-mİşlik* and *-mAzlık* often depict negative situations. In the case of *-mİşlik*, the depicted situation corresponds to a postterminal state resulting from a negative action (often passively experienced). In the case *-mAzlık*, it is a situation depicting an ongoing negative state.

8.2. The superordinate level

In what follows, we concentrate on the semantics of superordinate predicates: 48 verb tokens (40 types) and 22 noun tokens (11 types), amounting to 70 tokens overall or 49% of the 142 constructions considered. As the type-token ratio shows, the lexemes occur at a highly individual rate, with the notable exception of *duygu* ‘feeling’, which occurs eleven times (8% of all the 142 findings). We are first and foremost interested in lexemes (verbs and nouns) that semantically express various forms of mental and communicative processing of knowledge and emotional content. Drawing on the methodological idea of the “semantic map” approach (e.g. Malchukov, Haspelmath & Comrie 2010), we derive from our data the following semantic groups: SHOW, SAY, EVALUATE, PERCEIVE/UNDERSTAND, FEEL, and ACT/EXIST. Considering the semantic context of the individual findings, we tentatively ascribe the category SHOW to any items that refer to a nonverbal communication and SAY to any verbal communication. EVALUATE refers to any emotional-cognitive evaluation, PERCEIVE/UNDERSTAND to any cognitive-mental processing, and FEEL to any emotional processing. A further category, ACT/EXIST, accommo-

dates verbs that designate actions or ways of existing. Figure 2 shows the distribution of these semantic groups among the 70 examined findings.

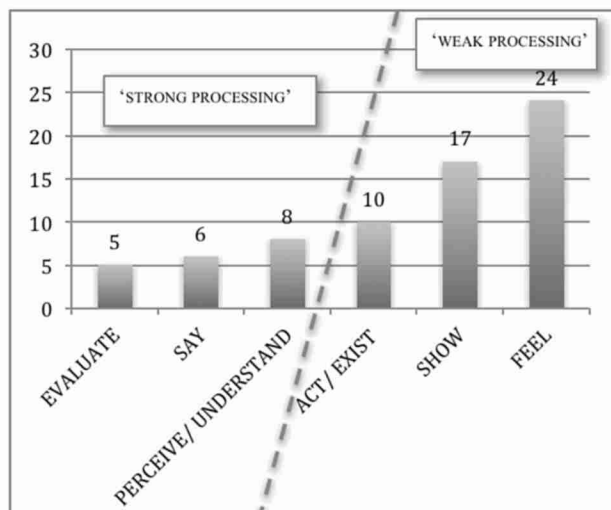


Figure 2: Semantic distribution of superordinate predicates, absolute numbers

As can be seen in this table, matrix predicates from the category FEEL make up the largest category (34%) among the 70 considered findings. Another large group consists of predicates semantically related to concepts of SHOWing (24%). Together with the groups of predicates that simply denote ACTs or ways of EXISTing (14%), they make up a large semantic field of a reality (largely consisting of negative feelings, see Section 8.1) that exists, is felt and is shown, but that is not—at the level of the matrix predications—being linguistically or cognitively processed: 73%. This processing—taking the form of speaking and analysing—happens in those contexts in which *-mİşIİK/-mAzIİK*-forms are constructed with superordinate predicates of the groups PERCEIVE/UNDERSTAND (11%), SAY (9%), and EVALUATE (7%): 27%. In accordance with their degree of active mental or linguistic processing, we term these two larger, overarching, groups “strong processing” matrix predicates versus “weak processing” matrix predicates. The weak processing matrix predicates are 2.66 times as frequent as the strong processing predicates in our data. Table 8 below lists the verbal and nominal predicates of the different semantic groups in a hierarchy representing their clausy versus nouny score.

Table 8: Superordinate verbs and nouns semantically related to mental and communicative processing of knowledge and emotional content (according to their scoring)

Score	Verb/noun	EVALUATE	SAY	PERCEIVE/ UNDER STAND	ACT/ EXIST	SHOW	FEEL
3	V <i>fark etmek</i> 'notice'			1			
	N <i>boşluk</i> '[feeling of] emptiness', <i>etki</i> 'influence'						2
2	V <i>göstermek</i> 'show', <i>kanıtlamak</i> 'prove', <i>vermek</i> 'give away, betray'					3	
	<i>övünmek</i> 'boast'	1					
	<i>tatmak</i> 'taste'						1
1	N <i>duygu</i> 'feeling'						1
	V <i>kalmak</i> 'remain', <i>taşımak</i> 'carry' (3), <i>uygulamak</i> 'apply'				5		
1	V <i>çağrıştırmak</i> 'evoke', <i>durmak</i> 'stand', <i>saklamak</i> 'hide', <i>vermek</i> 'give away, betray', <i>yazmak</i> 'write'					5	
	V <i>anlatılmak</i> 'be told', <i>anlatmak</i> 'tell', <i>nitelendirmek</i> 'characterise', <i>yüze vurmak</i> 'confront with, lit.: throw in one's face'		4				
	V <i>inandırmak</i> 'make believe', <i>inanmak</i> 'believe'	2					
	V <i>anlamak</i> 'understand', <i>kavramak</i> 'grasp' (2), <i>neden</i> 'reason'			4			
	V <i>hissettirmek</i> 'make feel', <i>yaşamak</i> 'experience, live'						2
	N <i>kanıt</i> 'proof'					1	
0	V <i>inanç</i> 'belief'	1					
	V <i>duygu</i> 'feeling' (7), <i>kaygı</i> 'worry'						8
0	V <i>eriştirmek</i> 'make attain', <i>ileri gelmek</i> 'come about', <i>kuşatmak</i> 'embrace'				3		
	V <i>barındırmak</i> 'contain', <i>göstermek</i> 'show', <i>imlemek</i> 'symbolise', <i>kalbe eklemek</i> 'add to the heart', <i>oyynamak</i> 'play, interpret'					5	
	V <i>anlatmak</i> 'tell' (2)		2				
	V <i>borçlu olmak</i> 'owe', <i>utanmak</i> 'feel ashamed'						2
	V <i>farkına varmak</i> 'realise', <i>okumak</i> 'read'			2			
	V <i>istemek</i> 'want'						1
	N <i>belge</i> 'document'					1	
0	V <i>duygu</i> 'feeling' (3), <i>gölge</i> '[feeling of] shadow', <i>iğrenti</i> 'disgust' (2)						6

-1	V	<i>uygun olmak</i> ‘fit’				1		
		<i>gösterilmek</i> ‘be shown’					1	
		<i>açılmak</i> ‘[one’s mind] open up’	1					
		<i>keşfetmek</i> ‘discover’			1			
	N	<i>utanmak</i> ‘feel ashamed’						1
		<i>rüya</i> ‘dream’					1	
-2	V	<i>doğmak</i> ‘be born, come about’				1		
			5	6	8	10	17	24

If we consider the data by zones on the clausiness/nouniness scale, it turns out that the score obtained does not seem to correlate with the degree to which matrix predicates fall into the two categories; see Table 9.

Table 9: “Weak processing predicates” and “strong processing predicates” according to scale value

Zone	‘strong processing matrix predicates’	‘weak processing matrix predicates’
high clausiness (2–3 points), total: 9	2 (22%)	7 (78%)
moderate clausiness (1 point), total: 32	11 (34%)	21 (68%)
neutral zone (0 points), total: 22	4 (18%)	18 (82%)
nouny zone (1–2 minus points), total: 7	2 (29%)	5 (71%)

A few words need to be said about predicates of the semantic category SAY, i.e. *verba dicendi*. These verbs make up only 9% of our sample of matrix lexemes. They do not occur in any of the constructions that received a high score on the clausiness side of the scale; rather, they occur in moderately clausal (1 point) or neutral (0 points on the scale) constructions. Furthermore, prototypical *verba dicendi* such as *söylemek* ‘say’, *bildirmek* ‘let know’ or *sormak* ‘ask’ are completely absent from our sample.

9. Conclusion

As our analysis reveals, *-mİşlİK* and *-mAzlık*, which do not occur more often than about twice in a hundred pages in our literary corpus, are shaky candidates for having subordinator status in several respects. Morphosyntactically, they are capable of combining with linguistic elements that can be defining criteria both for abstract nominalisation and for clausiness. Depending on the criteria applied, we find a more or less tangible tendency towards clausiness. Going by the scores on the nouniness/clausiness scale, we obtain 16 good candidates with a distinct clausal score of two to three points, or 70 (about half of our sample) if we also consider those that display a more moderate clausal score (one point). If we consider the occurrence of clausal categories that most consistently exclude the coexistence of any nominal categories, namely adverbial modification and the realisation of arguments of the verb, we obtain 20 good candidates. This leaves us with an approximate rate of between 14%

and 49% of good clausy candidates, depending on the criteria applied. There is a non-negligible counterweight on the nouny side, there are many neutral cases, and, most interestingly, there are many internal contradictions at the level of the individual construction. Classificational issues of internally contradictory constructions have granted us a close-up view of a transitional zone between nouniness and clausiness in our investigated area. This transitionality seems to extend into semantic areas, as a look at genitive NPs suggests, where the distinction between syntactic subjects of clauses and genitive attributes to nouns largely seems to be a matter of semantics.

Tentatively considering questions of productivity, we also discussed a few instances of postpositional constructions that we found reminiscent of constructions based on “complex converbs”, combining postpositions with the canonical subordinators (e.g. *-mİşlİktAn sonra* versus *-DİktAn sonra*).

At the semantic (illocutionary) level, the most interesting and unexpected result is that two thirds of the predications containing *-mİşlİK*/*-mAzlİK* show a clear tendency towards conveying emotionally negative states. More than half of these are passive forms, resulting in expressions of experiences of passive negative states. In the case of *-mİşlİK*, the depicted situation corresponds to a postterminal state resulting from an (often passively experienced) negative action or event. By contrast, *-mAzlİK* depicts an ongoing negative state. As for the superordinate predications, the analysis reveals a comparative lack of lexemes that express explicit verbal or cognitive processing or evaluation. Instead, there is a more pronounced tendency to use predicates of emotional experience, nonverbal forms of communication, or simply actions or existing states.

What this paper has not yet been able to systematically treat are functional-semantic, illocutionary comparisons between clausy candidates of *-mİşlİK*/*-mAzlİK* and the canonical complementisers, i.e. *-mA*, *-DİK*, *-mİş olduK*, and *-mİş olma*. Furthermore, a historical and synchronic look at similar constructions in other Turkic languages, taking a broader Turcological perspective, would provide us with further insights.

The combinational quality of *-mİşlİK* and *-mAzlİK* might raise more specific questions. The suffix *-lİK* realises a nominalisation only, whereas any additional information—aspectual—is hosted by the preceding suffix, *-mİş* or *-mAz*. This morphological situation brings about a higher degree of morpho-functional transparency than occurs with canonical subordinators (such as *-DİK* or *-(y)AcAK*), where categories are fused to a higher degree. This issue may need to be explored more systematically for the entire system.

One might further speculate about Haig’s (2003: 64–66) idea of “backlooping”, i.e. the morphosyntactic phenomenon whereby inflected forms undergo a (nominalising) derivational process at a higher level. In our data, this “backlooping” applies to aspectually inflected (*-mİş* or *-mAz*) constructions with internally clausy characteristics. The categorial tension that results from these highly individual constructions cannot be explained away in terms of lexicalisation, we think. Rather, the

“backlooping” of an already clausal construction into the area of abstract derivation (-*LIK*) brings about an illocutionary attenuation in otherwise specific verbalisations of emotionally negative content. This issue needs to be further explored.

Abbreviations

1	first person	IMP	imperative
2	second person	LOC	locative
3	third person	MOD	modality
ABL	ablative	NEG	negation
ACC	accusative	N	noun
ADJ	adjective	PAR	participle
CAUS	causative	PAS	passive
CD	conditional	PL	plural
CMP	comparative	POL	politeness
COM	comitative	POSS	possessive
COP	copula	POST	postterminal
DAT	dative	PRIV	privative
DEI	deixis	PRS	present tense
DER	derivational	PST	past tense
EQU	equative	PV	postverb
EVID	evidential	SUP	superlative
FUT	future	VB	verb
GEN	genitive	VN	verbal noun

Acknowledgements

This paper was first presented at a workshop on “Propositions vs. States-of-Affairs” organised by Kasper Boye and Marie-Louise Lind Sørensen at the 49th meeting of the Societas Linguistica Europaea (SLE), Naples, August 31–September 3, 2016. This workshop provided a very stimulating theoretical framework and we wish to thank the organisers as well as the participants. We feel that the largely empirical study that we present here cannot do full justice to this framework. We also wish to thank two anonymous reviewers as well as Everett Thiele for proofreading our English.

Primary sources/ data

- Akhanlı, Doğan 2010. *Fasıl*. İstanbul: Telos Yayıncılık.
 Atasü, Erendiz 2007 [1991]. *Dağın öteki yüzü*. İstanbul: Belge Yayınları.
 Buğra, Tarık 1996 [1963]. *Küçük ağa*. İstanbul: Ötügen Yayınları.
 Çelik, Jaklin 2011. *Öfkenin şenliği*. İstanbul: İletişim Yayınları.
 Dorsay, Atilla 1986. *Yüzyüze. Sinemacılarla konuşmalar*. İstanbul: Çağdaş Yayınları.
 Erdoğan, Aslı 2006 [1998]. *Kırmızı pelerinli kent*. İstanbul: Everest Yayınları.

- Hisar, Abdülhak Şinasi 1955 [1942]. *Boğaziçi mehtapları*. İstanbul: Hilmi Kitabevi.
- İlhan, Atilla 1999. *Yengecin kaskacı. Dört uzun hikâye*. Ankara: Bilgi Yayınevi.
- Karasu, Bilge 2010 [1999]. *Öteki metinler*. İstanbul: Metis Yayıncılık.
- Kaygusuz, Sema 2006. *Yere düşen dualar*. İstanbul: Doğan Kitapçılık.
- Kemal, Yaşar 1984 [1960]. *Orta direk*. İstanbul: Toros Yayınları.
- Kutlu, Ayla 2007. *Merhaba sevgi*. Ankara: Bilgi Yayınevi.
- Levi, Mario 2000 [1999]. *İstanbul bir masaldı*. İstanbul: Remzi Kitabevi.
- Mağden, Perihan 2006. *İki genç kızın romanı*. İstanbul: Merkez Kitapçılık.
- Margosyan, Mıgırdiç 2001. *Tespîh taneleri*. İstanbul: Aras Yayıncılık.
- Necatigil, Behçet 1993. *Düzyazular 1*. İstanbul: Yapı Kredi Yayınları.
- Nesin, Aziz 2005. *Adamı zorla deli ederler*. İstanbul: Adam Yayınları.
- Özakman, Turgut 1999. *Toplu oyunları 3. Gençlik ve çocuk oyunları*. İstanbul: Mitos Boyut Yayınları.
- Soysal, Sevgi 2002 [1968]. *Tante Rosa*. İstanbul: İletişim Yayınları.
- Tanpınar, Ahmet Hamdi 2002 [1983]. *Hikâyeler*. Ankara: Dergâh Yayınları.
- Toptaş, Hasan Ali 1995. *Gölgesizler*. İstanbul: Doğan Kitap.
- Uzuner, Buket 1990 [1989]. *Güneş yiyen çingene*. İstanbul: Gür Yayınları.

References

- Banguoğlu, Tahsin 1995 [1959]. *Türkçenin grameri*. Ankara: Türk Dil Kurumu.
- Borsley, Robert D. & Kornfilt, Jaklin 2000. Mixed extended projections. In: Borsley, Robert D. (ed.) *The nature and function of syntactic categories*. (Syntax and Semantics 32.) San Diego: Academic Press. 101–132.
- Boye, Kasper 2010. Reference and clausal perception-verb complements. *Linguistics* 48: 2, 391–430.
- Deny, Jean 1921. *Grammaire de la langue turque* (Dialecte osmanli). Paris: Imprimerie Nationale.
- Göksel, Aslı & Kerslake, Celia 2005. *Turkish: A comprehensive grammar*. London: Routledge.
- Haig, Geoffrey 2003. From lexical class to syntactic function: A sketch of Turkish word structure. In: Özsoy, Sumru A. & Akar, Didar & Nakipoğlu-Demiralp, Mine & Erguvanlı-Taylan, E. Eser & Aksu-Koç, Ayhan (eds.) *Studies in Turkish linguistics. Proceedings of the Tenth International Conference on Turkish Linguistics, August 16–18, 2000, Boğaziçi University, İstanbul*. İstanbul: Boğaziçi University Press. 59–68.
- Johanson, Lars 1975. Some remarks on Turkic “hypotaxis”. In: Johanson, Lars 1975. *Linguistische Beiträge zur Gesamt-turkologie*. Budapest: Akadémiai Kiadó. 210–224.
- Johanson, Lars 1990a. Subjektlose Sätze im Türkischen. In: Brendemoen, Bernt (ed.) *Altaica Osloensica*. Oslo: Universitetsforlaget. 199–218.
- Johanson, Lars 1990b. Die “problematischen” türkischen Nebensätze. *Türk Kültürü Araştırmaları* 28: 201–209.
- Johanson, Lars 1996. Kopierte Satzjunktoren im Türkischen. *Sprachtypologie und Universalienforschung (STUF)* 49: 1, 39–49.
- Johanson, Lars 2006 [1998]. The structure of Turkic. In: Johanson, Lars & Csató, Éva Á. (eds.) *The Turkic languages*. London: Routledge. 30–66.

- Johanson, Lars 2010. Three kinds of clause junctors. In: Ziegelmeyer, Georg & Cyffer, Norbert (eds.) *Aspects of co- and subordination. Case studies from African, Slavonic and Turkic languages*. Köln: Rüdiger Köppe. 9–14.
- Johanson, Lars 2013. Selection of subjunctors in Turkic non-finite complement clauses. *Bilig* 67, 73–90.
- Karakoç, Birsal & Herkenrath, Annette 2016. Clausal complementation in Turkish and Nohghay in a semantic perspective. In: Boye, Kasper & Kehayov, Petar (eds.) *Complementizer semantics in European languages*. (Empirical Approaches to Language Typology 57.) Berlin: de Gruyter. 619–664.
- Kornfilt, Jaklin & Whitman, John 2012a. Introduction: Nominalizations in syntactic theory. In: Kornfilt, Jaklin & Whitman, John (eds.) *Nominalizations in linguistic theory. Special issue of Lingua* 121: 7, 1160–1163.
- Kornfilt, Jaklin & Whitman, John 2012b. Afterword: Nominalizations in syntactic theory. In: Kornfilt, Jaklin & Whitman, John (eds.) *Nominalizations in linguistic theory. Special issue of Lingua* 121: 7, 1297–1313.
- Kural, Murat 1993. V-to(-I-to)-C in Turkish. *UCLA Occasional Papers in Linguistics* 11.
- Kural, Murat 1998. Subordinate Infs and Comp in Turkish. In: Johanson, Lars et al. (eds.) *The Mainz meeting: Proceedings of the Seventh International Conference on Turkish Linguistics*. Wiesbaden: Harrassowitz. 404–421.
- Lewis, Geoffrey L. 1967. *Turkish grammar*. Oxford: Oxford University Press.
- Malchukov, Andrej & Haspelmath, Martin & Comrie, Bernhard 2010. Ditransitive constructions: A typological overview. In: Malchukov, Andrej & Haspelmath, Martin & Comrie, Bernhard (eds.) *Studies in ditransitive constructions: A comparative handbook*. Berlin: De Gruyter. 1–64.
- Rehbein, Jochen 2007. Matrix constructions. In: Rehbein, Jochen & Hohenstein, Christiane & Pietsch, Lukas (eds.) *Connectivity in grammar and discourse*. (Hamburg Studies on Multilingualism 5.) Amsterdam: Benjamins. 419–447.
- Rehbein, Jochen 2009. *Rehbein-ENDFAS. Die Entwicklung narrativer Diskursfähigkeiten im Deutschen und Türkischen in Familie und Schule*. Universität Hamburg, EXMARaLDA: SFB 538-Korpora: http://www.exmaralda.org/sfb_e5.html.
- Rehbein, Jochen & Herkenrath, Annette & Karakoç, Birsal 2009. *Rehbein-SKOBI. Sprachliche Konnektivität bei bilingual türkisch-deutsch aufwachsenden Kindern und Jugendlichen*. Universität Hamburg, EXMARaLDA: SFB 538-Korpora: E5. Download: http://www.exmaralda.org/sfb_e5.html.
- Ross, John Robert 1972. The category squish: Endstation Hauptwort. In: Peranteau, Paul M. & Levi, Judith N. & Phares, Gloria C. et al. (eds.) *Proceedings of the eighth regional meeting of the Chicago Linguistic Society*. Chicago: Chicago Linguistic Society, University of Chicago. 316–338.
- Ross, John Robert 2004 [1973]. Nouniness. In: Aarts, Bas & Denison, David & Keizer, Evelien & Popova, Gergana (eds.) *Fuzzy grammar: A reader*. Oxford: OUP. 351–422 [First published in 1973 in: Fujimura, Osamu (ed.) *Three dimensions of linguistic theory*. Tokyo: The TEC Corporation. 137–257].
- van Schaaijk, Gerjan 2000. Higher order compounds in Turkish: Some observations. In: Göksel, Asli & Kerslake, Celia (eds.) *Studies on Turkish and Turkic languages*. (Turcologica 46.) Wiesbaden: Harrassowitz. 113–122.

Place nouns heading relative clauses with focal subjects

Gerjan van Schaaik

Schaaik, Gerjan van 2017. Place nouns heading relative clauses with focal subjects. *Turkic Languages* 21, 79–106.

Many grammars of Turkish give little attention to a type of construction which has puzzled quite a number of linguists trying to formalize the distribution of the so-called subject participle (SP) and the object participle (OP), the latter also known as the non-subject participle. A subject participle is used in three cases: when its subject is (i) the head of the RC; (ii) a constituent in a possessive relation with the head; and (iii) a non-referential noun phrase. In all other cases the object participle applies.

The present contribution provides a pragmatic analysis of the latter type of construction and proposes the name *Focus-Locus Construction*, showing that the subject is a non-referential noun phrase that is always placed in preverbal (focus) position and that the head noun of the relative clause can without exception be interpreted as a noun denoting location (locus). Such structures are presentative constructions providing new information, and they are related to existential constructions because they express “places where things happen”.

Gerjan van Schaaik, Leiden University. E-mail: g.j.van.schaaik@hum.leidenuniv.nl

1. Introduction

As is widely recognized, modern Turkish has several participles.¹ For the sake of convenience we shall confine ourselves to the participle forms that are said to have non-future and non-remote past time reference: the subject participle in *-(y)En* (subject form, henceforth: SF), and the object participle in *-TIK* (object form, henceforth, OF). They form the core of the equivalents of relative clauses (RCs), as can be exemplified by (1).

- (1) a. *fabrika-da çalış-an kardeş-im*
factory-LOC work-SF brother-POSS1S
'my brother who works at the factory'

1 The distribution of participles in relative clauses has been studied intensively over the past decades and attracts a great deal of interest to this day. Some general examples are: Underhill (1972); Hovdhaugen (1975); Hankamer & Knecht (1976); Dede (1978), Knecht (1979); Erdal (1981); Csató (1985, 1996); Zimmer (1987, 1996); Barker, Hankamer & Moore (1990); Özsoy (1994); Erkman-Akerson & Özil (1996); Kornfilt (2000); Öztürk (2008); Özçelik (2014).

- b. *kardeş-im-in* *çalış-tığ-ı* *fabrika*
 brother-POSS1S-GEN work-of-POSS3S factory
 ‘the factory where my brother works’

In (1a) the RC *fabrika-da çalış-an* ‘who works in a factory’ contains the SP² but no subject and “modifies” the head *kardeş-im* ‘my brother’, which is understood to be the subject of *çalış-* ‘work’. In (1b), on the other hand, the RC *kardeş-im-in çalış-tığ-ı* contains the OP and a subject, and modifies the head *fabrika* ‘factory’, which is not the subject of *çalış-* ‘work’. This subject is *kardeş-im* ‘my brother’, the subject-hood of which is signaled by the genitive case marker. The OF is followed by a possessive marker which agrees in person and number with the subject of the embedded verb. This opposition between SP and OP has been regarded as a strict divide with only a few exceptions.

Firstly, in a *başızozuk*-construction³, a head noun is modified by a “minimal clause” consisting of at least two constituents: a *subject* and a *predicate*. The subject is a noun plus possessive suffix third person singular (the so-called *anticipatory possessive*) in agreement with the head noun, and the predicate is a noun phrase (*topal* ‘lame’ in (3a)) or an SP (*tüt-en* ‘smoking, steaming’ in (3b)).

- (3) a. *Ayağ-ı* *topal* *bir* *sandalye-ye* *otur-du-m.*
 foot-POSS3S lame a chair-DAT sit.down-PAST1-1S
 ‘I sat down on a wobbly chair.’
- b. *Henüz duman-ı* *tüt-en* *sıcak* *bir ekmek-le* *geri* *dön-dü.*
 still smoke-POSS3S steam-SF hot a bread-INS back return-PAST1
 ‘She returned with a warm loaf of bread [its “vapor”] still steaming.’

For (3a) an alternative can be formed by the auxiliary *ol-an*, and its negative counterpart *ol-ma-yan* is obligatory in case of a negation. Compare:

- (4) a. *Ayağ-ı* *topal* *ol-an* *bir sandalye-ye* *otur-du-m.*
 foot-POSS3S lame be-SF a chair-DAT sit.down-PAST-1S
 ‘I sat down on a chair which has a wobbly [‘lame’] leg.’

- 2 The subject participle for the remote past takes the form *-miş*, for the future *-(y)Ecek*, and there are a few fossilized forms in *-(I/E)r* (also known as aorist participles). These will not be discussed here, since they are not relevant for the gist of the argument to be made in section 4.
- 3 This name derives from: *baş-ı bozuk adam* [head-his deranged man] ‘the man whose head is deranged’ and is coined by Lewis (1967: 259).

- b. *Ayağ-ı topal ol-ma-yan bir sandalye-ye otur-du-m.*
 foot-POSS3S lame be-NEG-SF a chair-DAT sit.down-PAST1-1S
 ‘I sat down on a chair which has no lame leg.’

The “minimal” subjects *ayağ-ı* ‘its leg’ in (3a) and *duman-ı* ‘its steam’ in (3b) both contain an anticipatory possessive, but this in itself is not the reason why the SP occurs. When the constituent containing the anticipatory possessive is in the nominative, it qualifies for subjecthood (5a), otherwise it must be some object of the participle; in (5b) the OP in *-TIK* shows by means of its possessive marker (possessive first person singular) who the subject really is: *ben* ‘I’.

- (5) a. *Karı-sı fabrika-da çalış-an kardeş-im ...*
 wife-POSS3S factory-LOC work-SF brother-POSS1S
 ‘My brother, whose wife works at the factory ...’
- b. *Karı-sın-ı fabrika-da gör-düğ-üm kardeş-im ...*
 wife-POSS3S-ACC factory-LOC see-OF-POSS1S brother-POSS1S
 ‘My brother, whose wife I saw at the factory ...’

Secondly, there is a type of construction which to the best of my knowledge has no “official” appellation, but which we will here term the *bülbül*-type of construction and for which I will suggest the name the *Focus-Locus Construction* in section 4. This construction too is in fact a relative clause; the minimal inventory is again a nominal head preceded by a subject plus SP. Its simplest form can be exemplified by *bülbül öten yer*,⁴ ‘the place where (a) nightingale(s) sing(s)’, a specimen much quoted in the literature on this type of Turkish relative clauses. Such constructions occur in several degrees of complexity. Here are a few other simple examples:

- (6) a. *Meşe bit-en toprak-ta, hemen hemen hiç başka ağaç*
 oak grow-SF land-LOC almost no other tree
gözük-me-z.
 to.be.seen-NEG-PRES2
 ‘On the land where oaks are standing, there are almost no other trees visible.’
- b. *Yıldırım düş-en deliğ-e bak-ar-ken ikisi de konuş-ma-dı.*
 lightning fall-SF hole-DAT look-PRES2-CONV two too speak-NEG-PAST1
 ‘Both said nothing when looking at the hole where the lightning struck.’

4 This construction is built up as follows: *bülbül öt-en yer* [nightingale sing-SF place] ‘the place where (a) nightingale(s) sing(s)’.

- c. *Bu bakteri de ışık ol-ma-yan yer-ler-de*
 this bacteria too light be-NEG-SF place-PLUR-LOC
kemosentez yap-ıyor.
 chemosynthesis do-PRES1
 ‘This bacteria too does chemosynthesis in places where there is no light.’
- d. *Güneş gir-me-yen ev-e doktor gir-er.*
 sun enter-NEG-SF house-DAT doctor enter-PRES2
 ‘In a house where no sunlight comes in, the doctor will enter.’

In all these examples (6 a-d) the subject stands bare before the verb. However, subjects can occur in their plural form (7a), can be preceded by adjectives (7b) and adverbial expressions (7c), and may pop up as two coordinated nouns (7d).

- (7) a. *Su-lar ak-an, kuş-lar öt-en toprak ...*
 water-PLUR flow-SF bird-PLUR sing-SF land
 ‘A piece of land where water runs everywhere and birds sing ...’
- b. *Sıcak su bulun-ma-yan otel ...*
 hot water be.found-NEG-SF hotel
 ‘The hotel where there is no hot water ...’
- c. *Bu yerleşme alan-lar-ı, bugün insan yaşa-yan bölge-ler-le aynı-dır.*
 this settlement field-PLUR-CM today man live-SF area-PLUR-INS same-EMPH
 ‘These fields of settlement are the same as the areas where people live nowadays.’
- d. *Elektrik ve su bul=un-ma-yan kent-te ...*
 electricity and water be.found-NEG-SF city-LOC
 ‘In the city where there is no electricity ...’

This type of construction can be based on passive⁵ and impersonal passive verbs as well. In both variants the subject of the SP can be compared to the direct object of their active counterparts. Examples are:

- (8) a. *Demirel, pasta üret=il-en mutfak-ta öğrenci-ler-le sohbet et-ti.*
 Demirel pastry produce-PAS-SF kitchen-LOC pupil-PLUR-INS chat-PAST1
 ‘Demirel had a chat with pupils in the kitchen where pastries are made.’

5 The passive suffix is separated from the stem by the equals sign (=).

- b. *Kâğıt oyna=n-an, sigara iç=il-en oda-nın açık*
 card play-PAS-SF cigarette smoke-PAS-SF room-GEN open
kapı-sın-a git!
 door-POSS3S-DAT go
 ‘Go to the open door of the room where one is smoking and playing cards.’

Quite frequently constructions are found that open with a constituent containing an anticipatory possessive plus a case marker. These constituents are in fact adverbial phrases, the cores of which are linked to the head of the RC by the possessive third person. Here are some typical examples:

- (9) a. *İç-in-de mantar yetiş-en orman ...*
 inside-POSS3S-LOC mushroom grow-SF wood
 ‘A wood where mushrooms grow ...’
- b. *İç-in-e atık su karış-an kanal ...*
 inside-POSS3S-DAT sewage water flow-SF canal
 ‘The canal that sewage water flows into ...’
- c. *Arka-sın-dan / ard-ın-dan güneş doğ-an bir dağ ...*
 backside-POSS3S-ABL sun rise-SF a mountain
 ‘A mountain behind which the sun comes up ...’
- d. *Üzer-lerin-den buz-lar sark-an ağaç-lar-la kaplı orman ...*
 surface-POSS3P-ABL icicle-PLUR hang-SF tree-PLUR-INS covered wood
 ‘A forest covered with trees from which icicles were hanging ...’
- e. *Hüseyin ben-i orta-sın-da çeşme ol-an bir küçük*
 Hüseyin I-ACC middle-POSS3S-LOC well ol-SF a small
meşdan-a getir-di.
 square-DAT take-PAST1
 ‘Hüseyin took me to a small square with a well in the middle.’

2. The main problem(s) in previous work

The main problem in the plethora of linguistic writings on Turkish relative clauses is undoubtedly how to account for the exceptions to the apparently well-established distinction between the subject participle and the non-subject participle. Rather than going into all possible approaches and solutions brought up since the first time the phenomenon of the “deviant” occurrence of the subject participle was addressed, I would prefer to pick out three works which clearly mark certain insights relevant for the analysis to be presented in section 4.

2.1. Ciopiński (1969: 59) mentions that the *bülbül*-type of construction received some attention in the larger grammars of Turkish, or in his words “*les grammaires monumentales*” [comprehensive grammars] (referring to Deny 1921 and Kononov 1956), but also that they are mostly neglected in textbooks and manuals (referring to Jansky (1943) and Lewis (1965)), a fact which he attributes to their low frequency of use, stating: “*probablement à cause de leur emploi, relativement rare*” [probably because of their relatively low usage].

Ciopiński sketches the structure of this construction in terms of lexical categories. A noun in the nominative is followed by a verb (intransitive or passive) in the form of a present participle (= SP), and this group (*groupe déterminatif* [modifying word group]) modifies a third element N. Furthermore, he states that the first noun is the logical subject of the participle and that there are no other formal indications about the syntactic relations between the nouns. Interestingly, a practical problem for Ciopiński is how to formulate a strategy that facilitates easy learning of this grammatical structure. In other words, when can a noun in the nominative plus an SP be used to modify a second noun? His solution is both bafflingly easy and elegant: it can be done by deriving logical inferences and offering them as paraphrases. Hence, *bülbül öten yer* is remodeled into *Bu yerde bülbül ötüyor* ‘At this place (here) the nightingale sings/nightingales sing’ and *Bülbül bu yerde ötüyor* ‘The nightingale sings at this place (here)’. We will return to this matter in section 4. Ciopiński (1969: 63), however, devotes only a few words on how the *bülbül*-type relates to constructions based on an OP. He states that this type of construction, with a genitive marked subject and a participle containing a possessive suffix, gives a more exact meaning, while the first type of construction, giving a vaguer meaning, are much lighter (“*Ces constructions donnent un sens plus exact, tandis que les premières, en donnant un sens plus vague, sont beaucoup légères.*”). The choice between the two types of participle, then, is made accordingly: the SP is preferred when the logical subject lacks individuality (“*Les constructions du type bülbül öten yer sont employées de préférence quand le sujet logique du déterminant n’a pas d’individualité*”).

2.2. Erdal (1981) adopts an analysis based on the “status” of the head in relation to the participle. He uses notions such as place, destination, and origin. These are usually associated with semantic roles or functions that are expressed as case markers or in terms of syntactic positions.⁶ His examples include:

- (10) a. *Yangın çık-an bir Amerikan uçak gemi-sin-de ...*
 fire break.out-SF a American airplane ship-CM-LOC
 ‘On an American aircraft carrier where a fire broke out ...’

6 In Erdal’s analysis, clause initial adverbial phrases, as in (9 a-e), are termed “possessor of X”, and X is the presumed semantic role of the head. Thus, example (9a) opens with a “possessor of place” (*içinde*) and (9b) with a “possessor of destination” (*içine*).

- b. *Duman kaç-an göz-ün-ü ovuştur-arak ...*
 smoke get.into-SF eye-POSS3S-ACC rub-CONV
 ‘While rubbing his eye(s), into which smoke had gone ...’
- c. *Kan sız-an yer-ler-e toprak ekele-me-ye başla-dı.*
 blood leak-SF place-PLUR-DAT earth sprinkle-INF-DAT begin-PAST I
 ‘He began to sprinkle earth on the places that blood was leaking from.’

In Erdal’s view, the relation in example (10a) between the participle *çık-an* ‘broke out’ and the head of the RC *uçak gemisi* ‘aircraft carrier’ is one of location, in example (10b) the relation between the participle *kaç-an* ‘getting into’ and the head of the RC *göz-ü* ‘his eye’ is one of destination, and finally, what underlies the relation between *sız-an* ‘leaking’ and the head of the RC *yer-ler* ‘places’ in (10c) is origin. Indeed, following Ciopiński, one could say things along the lines of *Bir Amerikan uçak gemi-sin-de yangın çık-tı* ‘A fire broke out on an American aircraft carrier’, *Göz-ün-e duman kaç-tı* ‘Smoke has gone into his eye(s)’, and *Bu yer-ler-den kan sız-ıyor* ‘Blood was/is leaking from these places’. In my opinion these “roles” are only relevant for the description of the paraphrases in terms of what type of adverbial constituents can be expected. They are not a part of the verbal semantics; adverbial constituents are satellites and not arguments. Furthermore, certain verbs may be accompanied by adverbial phrases with different case markings, possibly in the same clause at the same time. It was assumed that *kan sız-an yer-ler* correlates with *Bu yer-ler-den kan sız-ıyor*, but since *sız-* with an ablative complement means ‘to trickle out of, to leak from’, and with a dative complement ‘to trickle into, to leak into’, the intended meaning of, for instance, *su sız-an duvar-lar* ‘walls onto/from which water leaks’ can only be inferred from the context or situation and I think a detail such as ‘onto’ or ‘from’ is in many a case not even relevant. What is more important is the general image generated of such a wall: damp, wet or soaked with water. What the reader or hearer makes of such structures may also depend on available case recovery strategies (see Haig 1998b).

The above observations lead to the conclusion that the relation between head and verb is irrelevant for the form of the participle; all heads of the RCs in (6)–(10) are not the subject of the participle, although it is the SP that occurs. This is in contrast with Erdal’s line of thought. It is understandable because in many cases (but probably not all) two types of constructions can be contrasted which are apparently built up with the same lexical material but differ in grammatical elements. In this way, (11a) can be opposed to (11b):

- (11) a. *Yangın çık-an Amerikan uçak gemi-sin-de ...*
 fire break.out-SF American airplane ship-CM-LOC
 ‘On the American aircraft carrier where a fire / fires broke out ...’

- b. *Yangın-in çık-tığ-ı Amerikan uçak gemi-sin-de ...*
 fire-GEN break.out-OF-POSS3S American airplane ship-CM-LOC
 ‘On the American aircraft carrier where the fire broke out ...’

Quite often the choice (and hence, its form) of the participle is thought to be related to the type of subject involved, and the problem is reformulated in terms of what type of subjects might be excluded from the *bülbül*-type of construction. Semantic features such as “human being” and “animate being”, as well as pragmatic notions such as definiteness, specificity, and individuation are presented as possible factors and even a link with oppositions such as *bir kadın şapka-sı* ‘a lady’s hat’ and *bir kadının şapka-sı* ‘the hat of a lady’ is mentioned.

In order to get a more solid grip on the construction under scrutiny, Erdal (1981: 29–45) directs his attention to a great variety of authors (from 1895 to 1975) who have all commented on this type of construction. The following points in his evaluation are significant. First, demonstratives, pronouns and proper names are excluded as subjects in the *bülbül*-type of construction. Second, the subjects are all non-referential. Third, this type of construction is productive, and not a mere remnant of an archaic type of construction that lives on in the form of proverbs (cf. (6d)), as for instance advanced by Underhill 1972 (quoted by Erdal (1981: 36)). Some literary examples contra Underhill are:

- (12) a. *Et gir-me-yen yemek-te tat ol-ur mu?*
 meat enter-NEG-SF food-LOC taste occur-PRES2 Q
 ‘Is there [any] taste in food where no meat went in?’
- b. *Rüzgar gir-en bir pencere aralığ-in-ı kağıt-la tıka-di.*
 wind enter-SF a window crack-CM-ACC paper-INS stuff-PAST1
 ‘He stuffed with paper a window crack where wind came in.’

Fourth, there are several allusions in Erdal’s article to the fact that the construction with the SP (the participle ending in *-(y)En*, that is) is ancient, and that the OP (participle in *-TIK*), according to Deny (1921), is a “relatively recent Ottoman innovation”. We will return to this matter in section 5.

2.3. Haig (1998a), in his monumental study on Turkish relative clauses, ventures to develop a systematic and exhaustive account of the circumstances under which subject and object clauses are formed. As for the *bülbül*-type of construction, he claims that the structure of the noun phrase representing the *bülbül*-part (the subject) cannot sufficiently be explained in terms of definiteness, specificity, or referentiality alone, and tries to show that an adequate account should be based on three interacting parameters: control, individuation, and the conditions determining the occurrence of the anticipatory possessive. The work is amply illustrated by examples that were also published by Erdal (1981), and by further examples drawn from a corpus. The main prob-

lem is the nature of the subject in this type of construction, and Haig's account dispenses with the notion of subject incorporation, though on other grounds than does Cagri (2009), who in turn rejects the claims by Öztürk (2009) which are in favor of incorporation.

Haig (1998a: 184) gives the following summary of his findings:

1. Subjects⁷ are overwhelmingly the subjects of low control combinations, i.e. the non-human subjects of low-activity intransitives.
2. Precisely quantified subjects, e.g. with *bir*, never occur as subjects, unless the relative clause contains an anticipatory possessive.
3. Personal pronouns, nouns with normal possessive morphology, demonstratives, nouns modified by adjectives created from the adjectival suffix *-ki* are never subjects.
4. Subjects occur with only two transitive predicates: *sok-* 'sting', and *bas-* 'overgrow'. Even here, their occurrence is bound to extremely specific conditions: They are minimally individuated, the relative clause contains an accusative-marked anticipatory possessive, and the subject is [-human].

Before I present some comments on these four points, I would like to give some more data relevant for the evaluation in section 4.

3. More Data

Recall Ciopiński's remark about the frequency of use: "relatively low". Now, take into consideration that at the time of writing, in order to find and copy a suitable exemplar of this type of construction he probably had no other means to tap literary sources than the manual perusal of huge piles of books and newspapers. Modern computer applications for searching through massive corpora reveal however that the *bülbül*-type of construction is very productive and that it occurs much more frequently than would be expected on the basis of such a statement. Also, frequency alone shouldn't be the standard for judging whether a construction should be included in a manual or textbook. The guiding principle is usefulness. But apart from that, the construction has never suffered from a lack of interest in linguistic circles, particularly because it seems extremely hard to point out a sufficient number of factors to exhaustively account for its make-up. Perhaps this justifies just another modest attempt.

A crucial clue for an alternative characterization of the *bülbül*-type of construction was introduced in Ciopiński (1969: 60). What he did was to remodel a sequence of N1+SP+N2 into a logical inference that forms, as it were, an appropriate answer to the question 'What is going on at N2?'. In this way, *bülbül öten yer* can be paraphrased as *Bu yerde bülbül ötüyor* 'At this place (here) the nightingale sings / nightingales sing' and as *Bülbül bu yerde ötüyor* 'The nightingale sings at this place

7 In the original text Haig prefers the term "semi-subject" over "genitiveless subject" for subjects of the *bülbül*-type of construction.

(here)’. Now, for educational purposes, one could say that when either of the derived propositions is true, the *bülbül*-type of construction may be employed. This seemingly simple trick for determining if the construction matches the concept to be expressed is solely based on one particular feature of the head of the construction: its position in relation to the event described.

The answer is that the head of the construction denotes a place, a location.⁸ This is indeed the case with *toprak* ‘land’ in (6a, 7a), *delik* ‘hole’ in (6b), *yer* ‘place’ in (6c, 10c), *ev* ‘house’ in (6d), *otel* ‘hotel’ in (7b), *alan* ‘field’ in (7c), *kent* ‘city’ in (7d), *mutfak* ‘kitchen’ in (8a), *oda* ‘room’ in (8b), *orman* ‘wood, forest’ in (9a), *kanal* ‘canal’ in (9b), *dağ* ‘mountain’ in (9c), *ağaç* ‘tree’ in (9d), *meydan* ‘square’ in (9e), *gemi* ‘ship’ in (10a), *göz* ‘eye’ in (10b). With the exception of *gemi* ‘ship’ in (10a) and *göz* ‘eye’ in (10b), all these nouns can be conceived of as denoting a *place* rather than a *thing*.⁹ This interpretation in terms of a location is typical of immovable goods (land, forest, canal, house, etc.), but words for movable goods (“things”) can also to a great extent be used to indicate a place, that is, the place where something happens.

The constructions exemplified in (6)–(9) all show a minimal inventory: they are all based on the pattern N-V-SF-N in which the head is a place noun. In case the head of the relative clause is primarily to be regarded as a noun denoting a thing, an adverbial phrase is likely to occur. Such adverbial phrases always occur in the form of a noun¹⁰ which is “linked” to the head via the anticipatory possessive (third person). This can be exemplified by:

- (13) a. *Alt-in-a* / *üst-ün-e* *toz* *düş-en masa-lar ...*
underside-POSS3S-DAT / top-POSS3S-DAT dust fall-SF desk-PLUR
‘The desks under/on top of which dust drifts down ...’
- b. *Alt-in-dan* *duman* *çık-an* *kapı ...*
underside-POSS3S-ABL smoke come.out-SF door
‘The door from under which smoke is coming out ...’

8 Erdal (1981: 33) rejects Kononov’s idea (cited as Kononov 1956: 452–453) of an “attributive phrase expressing circumstances of *time* and *place*”, because of transitive verbs and other exceptions.

9 A motivation for the distinction between place and thing readings for nouns is given in Mackenzie (1992). The relevance for Turkish is set forth in Van Schaaik (2002: 242f) and application of these concepts to an analysis of postpositions is presented in Van Schaaik (2011).

10 These nouns often denote a side, area or space and can lexically be categorized as Relative Place Nouns, e.g. *alt* ‘underside’, *üst* ‘top’, *ön* ‘front’, *arka* ‘back’, *yan* ‘side’ etc. For details, see Van Schaaik (2011).

Although at first glance the head in these examples denotes a thing (a physical object), the purport of these statements is to indicate that the tables in (13a) are the places where dust comes/came down and that the door in (13b) is the place from which smoke emerges. This is, of course, strongly supported by the introductory adverbial phrases based on a noun indicating the exact spot relative to the larger object (*in casu*, the thing referred to by the head).

Sometimes it takes a bit of imagination to see that a noun such as ‘tea’, which may primarily be associated with a hot liquid, can also be used to refer to a shrub of the botanical genus tea, a sort of tea, a business or trade (e.g. *My brother in law is in tea*) or even a region where tea is grown. This is the case in (14).

- (14) *Üzer-in-e kar yağ-an tek çay Türk çay-ı-dir.*
 top-POSS3S-DAT snow fall-SF sole tea Turk tea-CM-EMPH
 ‘The only tea [variety] on which snow falls is Turkish tea.’

In the following examples too it is apparently expedient to specify some exact spot in relation to the head as a whole. Such a specification is usually a Relative Place Noun, as in (15 a–b), but nouns denoting some part or component of the head are also not uncommon, as in (15c).

- (15) a. *Üst-ün-den tren geç-en köprü-nün alt-ın-dan geç-ti-m.*
 top-POSS3S-ABL train pass-SF bridge-GEN under-POSS3S-ABL pass-PAST1-1S
 ‘I went under a bridge on which trains pass.’
- b. *Tam orta-sın-da kapı bulun-an dümdüz,*
 right middle-POSS3S-LOC door be-SF straight
yalın bir duvar var-dı.
 clear a wall exist-PROJ1
 ‘There was a straight clear wall where there was a door right in the middle.’
- c. *Namlu-ların-dan hafifçe duman çık-an top-lar...*
 barrel-POSS3P-ABL lightly smoke come.out-SF cannon-PLUR
 ‘Cannons where some smoke was spiraling up from their barrels ...’

Shorter variants of (15a) and (15b), for instance *tren geç-en köprü* and *kapı bulun-an duvar*, are mostly judged by native speakers as ungrammatical, especially when they occur without context. A possible reason is that such chunks do not give enough information to create a mental picture of what is happening: What is what doing to what? For the shorter *köprü*-variant, there are two factors that might explain why there is friction. Firstly, the verb *geç-* is used with different case markers, depending on the shade of meaning to be expressed. In its (transitive) sense of “to cross, to span” its object requires the accusative, as exemplified in (16 a–b):

- (16) a. *Bu köprü Boğaz-ı Ortaköy-de geç-ecek.*
 this bridge Bosphorus-ACC Ortaköy-LOC cross-FUT
 ‘This bridge will cross the Bosphorus in Ortaköy.’
- b. *İstanbul Boğaz-ı’nı geç-en köprü-ler ...*
 Istanbul strait-CM-ACC cross-SF bridge-PLUR
 ‘(The) bridges crossing / spanning the Bosphorus ...’

But in the sense of ‘to pass’, *geç-* takes a so-called oblique object with typical ablative marking, thereby indicating the path along which the passing takes place. It should come as no surprise then that ‘cross the bridge’ is rendered by the Internet dictionary *tureng.com* as *köprüden geçmek* and *köprüyü geçmek*.

Secondly, we can say that the role of *köprü* ‘bridge’ is ambiguous in relation to *tren* ‘train’ (in 15a), since some bridges are built to cross a railroad so that the train passes under the bridge, while other bridges prioritize trains by having them pass on top. Now, using an adverbial phrase of “the right format” (relative place noun linked to the head through the anticipatory possessive third person plus the appropriate case marker), it is indicated beforehand that the hearer should be prepared for a non-standard interpretation of the participle.

A similar strategy of avoiding ambiguity can be shown to be at work in other constructions as well. In (17 a–b) the head (*yer* ‘place’) of the relative clause is clearly a place noun, but in (17c) the head (*baca* ‘chimney, funnel’) is not, and as a matter of consequence, “locational” specification is provided for by the adverbial phrase *deliğ-in-den* ‘from its opening’.

- (17) a. *Duman çık-an bir yer-i göster-iyor-lar-di.*
 smoke rise-SF a place-ACC show-PRES1-3P-PROJ1
 ‘They showed [him] a place from which smoke was rising.’
- b. *Ama tam orta-da duman çık-an*
 but right middle-LOC smoke rise-SF
büyük çıplak bir yer var-di.
 vast bare a place exist-PROJ1
 ‘But right in the middle there was a vast bare place from which smoke rose up.’
- c. *Deliş-in-den daha duman çık-an baca-yı gör-dü-ler.*
 hole-POSS3S-ABL still smoke come.out-SF chimney see-PAST1-3P
 ‘They saw the chimney from whose opening smoke was still coming out.’

Of course, in parsing a sentence, the occurrence of an adverbial phrase is not in itself a signal that some constituent with “irregular” or “deviant” morphology should be expected. Compare the next two examples, of which the first is a relative clause with ordering “SP-subject” and the second is of the *bülbül*-type of construction.

- (18) a. *Otoban-ın üst-ün-den geç-en köprü-de vedalaş-tı-k.*
 motorway-GEN top-POSS3S-ABL pass-SF bridge-LOC say.goodbye-PAST1-1P
 ‘We said goodbye on the bridge that crosses the motorway.’
- b. *Belki Prens-le Prenses gid-ip iç-in-den derin bir ırmak*
 maybe prince-INS princess go-CONV inside-POSS3S-ABL deep a river
geç-en ülke-ler-in bir-in-de otur-ur-lar.
 pass-SF land-PLUR-GEN one-POSS3S-LOC live-PRES2-3P
 ‘Maybe the prince and the princess have gone to live in one of those lands
 where a deep river flows.’

It seems that transitivity as such is not necessarily the sole factor which determines whether an ordering “subject-SP” is grammatical or not. With bare constructions such as *tren geç-en köprü* ‘train passing bridge’, confusion might arise about what crosses what, since *tren* ‘train’ can be taken as the direct object.

Most constructions of this type are based on an intransitive verb, and hence confusion is out of the question, because any preverbal noun will be interpreted as the subject as long as the head stands for some location or other. But, as was the case with *geç-*, transitive verbs do have one or more objects. The verb *gir-* ‘to enter’ in the examples below usually goes with a dative object (compare Erdal’s ‘destination’): *ev* ‘house’ in (19a), *bir köy* ‘a village’ in (19b), and *bir orman* ‘a forest’ in (19c).

- (19) a. *Güneş gir-me-yen ev-e doktor gir-er.*
 sun enter-NEG-SF house-DAT doctor enter-PRES2
 ‘In a house where no sunlight comes in, the doctor will enter.’
- b. *Hiç öğretmen gir-me-miş bir köy-de okuryazar*
 no teacher enter-NEG-SF a village-LOC literate
çocuk-lar bul-du-lar.
 child-PLUR find-PAST1-3P
 ‘In a village where never a teacher had entered they found literate children.’
- c. *Balta gir-me-miş bir orman-da ...*¹¹
 axe enter-NEG-SF a forest-LOC
 ‘In a virgin forest [a forest that had never seen an axe] ...’

The lack of confusion about what does what can be attributed to the fact that (19a) is a proverb (a first-time reader or hearer might need some explanation, though), but

11 Although this is a set phrase, the subject can be replaced by for instance *insan* ‘human being’, as in: *insan gir-me-miş bir orman* ‘forest where no human has gone into’.

the correct interpretation of (19 b–c) is warranted by the knowledge that teachers and axes enter villages and forests, and that the reverse is unthinkable.

Only a few literary examples are known which are based on verbs with a direct object. So, in addition to dative, locative and ablative, accusative case markings also occur for the object of the verb, on the proviso that the object contains a possessive marker. Such structures are in fact a rarity, mainly because of the ambiguities they can entail. In (20a) for instance *bakkal* should be interpreted as a place ‘grocer’s shop’ and not as a person, and furthermore, I assume that for most people there is a stronger association between *peynir* ‘cheese’ and *fare* ‘mouse’ when it comes to ‘eating’ than between mice and grocers (imagine: ‘mouse eats grocer’ or ‘grocer eats mouse’). Therefore I reckon that such structures are fully acceptable under similar conditions. For (20b) it is difficult if not impossible to maintain that *kız* ‘girl’ stands for a place and not for a person. Yet, this example is fully acceptable, possibly owing to the sequence ‘nose-mosquito-sting’.¹² The word *masa* ‘table’ in example (20c) should not be interpreted as an object but rather as the location of some *cam* ‘glass pane’. Also in (20d) and (20e) certain ‘areas’ or ‘spots’ are meant, rather than objects. Leaving out the direct objects of (20 a–e)¹³ would predictably lead to ungrammatical structures.

- (20) a. *Peynir-in-i* *fare* *yi-yen* *bakkal* ...
 cheese-POSS3S-ACC mouse eat-SF grocer
 ‘The grocer’s at which / where mice have eaten the cheese ...’
- b. *Cam-in-i* *maymun* *kır-an* *masa* ...
 glass-POSS3S-ACC monkey break-SF table
 ‘The table whose glass pane monkeys have broken ...’
- c. *Burn-un-u* *sivrisinek* *sok-an* *kız* ...
 nose-POSS3S-ACC mosquito sting-SF girl
 ‘The girl who was stung in the nose by mosquitoes ...’

12 Haig (1998: 183) points out that *sok-* in the sense of ‘to sting’ occurs only in combination with *arı* ‘bee’, *akrep* ‘scorpion’, and *yılan* ‘snake’, and for that reason (20b) might count as a type of highly conventionalized type of expression, coming close to a “fixed” way of speaking. Another such type of verb is *bas-* in combinations such as *su bas-* ‘to flood’ and *ot bas-* ‘to overgrow with weeds’. On the other hand, in the sense of ‘break out’ we find (20d) and (20e), as well as *sis bas-* ‘to fog up’ in (29b).

13 Examples (20a) and (20b) are taken from Cagri (2009: 370–371), who translates *bakkal* by ‘shop’ in (20a) and who uses *goril* ‘gorilla’ instead of *maymun* ‘monkey(s)’ in (20b). Example (20c) is an adaptation of the often quoted *bacağ-m-i arı sok-an kız* ‘the girl whose leg a bee/ some bees stung’, a muster which probably goes back as far as Underhill (1972).

- d. *Diz-ler-i titre-yip aln-in-i ter*
 knees-POSS3S tremble-CV forehead-POSS3S-ACC sweat
bas-an genç adam.
 break.out-SF young man
 ‘The young man with trembling knees, and on whose forehead sweat broke out.’
- e. *Her yan-in-i ateş bas-an kadın kaç-ma-ya çalış-tı.*
 all side-POSS3S-ACC go.hot-SF woman escape-INF-DAT try-PAST1
 ‘The woman whose face was glowing tried to escape.’

4. An alternative characterization

The relevant data have now been presented in such a way that some conclusions can be drawn. Firstly, the *bülbül*-type of construction could be given an appropriate name. What makes this type of construction special and interesting are that the subject precedes the so-called subject participle instead of following it, and objects and adverbial phrases always precede the subject. For the constructions represented in (6)–(20), I would suggest the name *Focus-Locus Construction*, because this appellation does justice to what is most characteristic for the construction. The part *Locus* reflects the fact that the head of the relative construction signifies a location, a *locus in quo* ‘place at which’, or more precisely, a *locus actus* ‘place of the act’.¹⁴ The *Focus* part reflects the special position of the subject; it is placed in the immediate preverbal position, a syntactic position associated with the pragmatic notion of Focus.¹⁵ Secondly, the *Focus-Locus Construction* can be characterized as a *presentative expression*. Such constructions are used to introduce new information (Dik 1989) to the hearer in the form of noun phrases. Simple examples in English¹⁶ are: *There are dark clouds on the horizon*, and *There’s a white bird in this blizzard*. Such statements take the form of an existential construction, but the intended illocutionary effect is not to point out that something exists but rather, to indicate the presence or availability of something at the moment of speaking. A typical feature of such constructions is a locative phrase, which is understandable if one considers the idea that “to say that something exists is to say that it is located somewhere” (cf. Dik, 1989: 177, referring to Lyons 1967). In this way, existence and location can be said to be logically interrelated. The expression (specification) of the locative part is not necessary; when a statement is solely meant to present the thing(s) at hand, then statements like *There are some beers in the fridge* and *There’s beer, if you’d like some!* are more or less equivalent from a practical point of view.

14 The term *Locus Actus* is described on the web site Oxfordreference.com as follows: “Place of the act: The place where an act occurred; the place where a contract was performed”.

15 Dik (1989: 263–287) presents a comprehensive analysis of pragmatic functions. For Focus in Turkish, see Van Schaik (1983, 1998, 2001), and Göksel & Özsoy (2003).

16 Hannay (1985) goes into the relation between presentatives and existentials in English.

It is the combination of existence and location that underlies the relative clauses modeled after *bülbül öt-en orman* ‘forest where (a) nightingale(s) sing(s)’. The head of the construction represents the location and the fragment *bülbül öt-SF* describes, concisely and to the point, what specifies its referent: “there are nightingales singing”. And this latter type of expression is the presentative line of approach mentioned above. With this observation in mind the constructions analyzed in the present paper can be understood as being about “places where things happen”.

Of course, a number of similar constructions can be put forward in which, strictly speaking, not the head is a location proper, e.g. (13a–b) and (20b), but it cannot be denied that the precise whereabouts of the things denoted by both the adverbial phrase and by the head must coincide, given the part-whole relationship as formalized by the anticipatory possessive in these constructions. Although *Ali* in the following example is clearly not a place in the strict sense, it is still obvious in (21a) where some dog or another relieved itself. And similarly, in (21b), it is not too hard to determine the position of a whole pack of dogs relative to the single wolf.

- (21) a. *Entari-sin-e köpek işe-yen ve yeğen-i tarafından*
 robe-POSS3S-DAT dog pee-SF and niece-POSS3S by
tartakla=n-an Ali ...
 push-PAS-SF Ali
 ‘Ali, against whose robe a dog peed, and who was pushed by his niece ...’
- b. *Ard-in-da yüz köpek havla-ma-yan kurt, kurt*
 rear-POSS3S-LOC 100 dog bark-NEG-SF wolf wolf
say=ıl-ma-z.
 consider-PAS-NEG-PRES2
 ‘A wolf behind whose back 100 dogs don’t bark is not considered a wolf.’

Lastly, the internal constituent order of the *Focus-Locus Construction* must be explained. The order can easily be determined on the basis of the following observations. In its simplest form this construction comprises three constituents only: (subject participle) head. The first two together tell us something about the third one. In other words, the relative clause made up of subject and participle is an attribute of the head. Thus:

- (22) *Bülbül öt-en orman ...*
 nightingale sing-SF forest
 ‘The forest where (a) nightingale(s) sing(s) ...’

Secondly, the overall complexity can be varied by adding all kinds of extra information. For instance, the subject may be plural and can be modified for quantity:

- (23) a. *Üçüncü kat-in-dan duman-lar çık-an bina ...*
 third floor-POSS3S-ABL smoke-PLUR come.out-SF building
 ‘The building where a lot of smoke came out from the third floor ...’
- b. *Su-lar-in-dan bol balık çık-an göl-ler ...*
 water-PLUR-POSS3S-ABL much fish come.out-SF lake-PLUR
 ‘Lakes where an abundance of fish come out [from their waters] ...’
- c. *Göz-ün-den bir damla yaş ak-ma-yan bu kadın ...*
 eye-POSS3S-ABL a drop tear flow-NEG-SF this woman
 ‘This woman from whose eyes there flows not a single teardrop ...’
- d. *Yüz-ü kapalı ve ağız-in-dan tek*
 face-’her’ covered and mouth-POSS3S-ABL single
söz çık-ma-yan bir genç kız ...
 word come.out-SF a young girl
 ‘A veiled young girl from whose mouth not a single word comes out ...’
- e. *İç-in-den bir baş, iki kol ve iki ayak*
 inside-POSS3S-ABL a head, two arm and two foot
çık-an uç-ma-ya hazır bir balon ...
 come.out-SF fly-INF-DAT ready a balloon
 ‘A balloon ready to fly, out of which a head, two arms and two legs are sticking ...’

Thirdly, adverbials are not unwelcome either:

- (24) a. *Bütün gece göz-ün-e uyku gir-me-yen arkadaş-ı ...*
 whole night eye-POSS3S-DAT sleep enter-NEG-SF friend-POSS3S
 ‘His friend, who could not sleep all night, ...’
- b. *Motor-un-dan birden alev çık-an kargo uçağ-ı*
 engine-POSS3S-ABL suddenly flame come.out-SF cargo plane-CM
biz-i korkut-tu.
 us-ACC frighten-PAST1
 ‘The cargo plane from whose engine flames suddenly came out frightened us.’

Fourthly, adjectives modifying the subject are rather common as well:

- (25) a. *Kolon-lar-in-dan paslı demir-ler çık-an yarım kalmış*
 column-PLUR-POSS3S-ABL rusty iron-PLUR come.out-SF half-finished

- beton yapı-lar-ı ...*
 concrete structure-PLUR-CM
 ‘Half-finished concrete structures where rusty iron sticks out of the columns ...’
- b. *Sürekli isli bir yağmur yağ-an kapkara gökyüzün-de ...*
 incessantly sooty a rain precipitate-SF pitch black sky-LOC
 ‘In a pitch black sky from which incessantly a sooty rain falls ...’
- c. *Ağz-ın-dan asla yanlış bir sözcük çık-ma-yan*
 mouth-POSS3S-ABL never wrong a word come.out-NEG-SF
gezgin bir yargıç ...
 touring a judge
 ‘A touring judge out of whose mouth never a bad word comes ...’

Fifthly, this type of construction occurs most frequently with a participle based on an intransitive verb, and hence, an object is not to be expected. Objects are only possible, or rather obligatory, with transitive verbs, as in (20a), the example with mice devouring the cheese at the grocer’s (place).

A few more remarks regarding the examples in (20 a–e) might be in place. The verbs are all transitive, given the accusative markings of the constituents containing the anticipatory possessive. Leaving these constituents out would render these constructions ungrammatical. Furthermore, the head of (20a) can be interpreted as “person” and as “place”, that of (20b) as place only, whereas all other heads are clearly ‘human’ and do not qualify for a classification as place noun. Apparently, the Focus-Locus approach does not work in these examples for locus-part of the equation. On the other hand, such constructions based on a transitive verb are extremely rare. The examples (20 a-b) were found once and the original form of (20c) is repeated over and over again without variation in the *linguistic* literature. Examples based on collocations with *bas-* (see footnote 13) occur much more frequently in texts of a *bel-letristic* nature. All in all, the number of limitations of the proposal embodied in the pragmatic analysis presented here may be rather limited in itself; setting aside (20a-c) because of possible doubts regarding their quality or frequency, one could say that in constructions with a ‘human’ head, as in (20d-e), the transitive verb is based on a collocation of *bas-*, and that the locus-part is expressed by a direct object.

Summarizing, we can say that the *Focus-Locus Construction* is in fact a relative clause which has a lot in common with a regular finite clause. It has a verbal part in the form of a participle, a subject, and other constituents, such as objects and adverbial phrases. What makes it special and interesting are that the subject precedes the so-called subject participle instead of following it, and objects and adverbial phrases always precede the subject. However, the make-up of that type of subject is conspicuous.

The type of noun phrase that qualifies as the subject of *Focus-Locus Constructions* is in fact an almost full-blown noun phrase; the only restriction is that it be

non-referential. Its form may range from *bir* + *noun* for singularity (e.g. *bir lamba yanyyor* ‘a [single] lamp is on’) to *noun* + *-lEr* for plurality (e.g. *lamba-lar yanyyor* ‘lamps are on’) and *noun* + \emptyset (e.g. *lamba yanyyor* which can be interpreted as ‘there is light’).¹⁷

Being non-referential entails that definiteness is excluded, but at the same time the core of a non-referential noun phrase may occur in the plural (as in (23a)); it may be quantified (as in (23b–e)); and it may be qualified by adjectives (as in (25a–b)). The notion of a non-referential noun phrase is completely adequate; the occurrence of *bir* however does not imply ‘indefiniteness’ in the sense of ‘intended to construe a possible referent’ (cf. Dik, 1989: 139), but rather it is only meant to help create a *general image* of some singular individual, whereas the absence of *bir* would be sufficient to create an image of a vaguer nature and an undefined quantity.

As was pointed out in section 2, it has been suggested that these constructions could be explained in terms of subject incorporation, but the mere fact that such subjects can occur in plural form and that they can be accompanied by adjectives, expressions of quantity and adverbial phrases, argues against this idea. Moreover, the structure of a non-referential noun phrase giving shape to the subject is always identical with the ‘lean’ type of noun phrase,¹⁸ which qualifies as a possible left-hand member in a nominal compound.

As for the position of the subject relative to other constituents, it is well-known that there is a relationship between the information structure¹⁹ of a sentence and the places that some constituent may occupy. Generally speaking, the determining factors can be sought with regard not only to definiteness, but also pragmatic salience. To give a simple example, with *Kitap masada duruyor* the main point is to give information about the whereabouts of some previously mentioned book, and it could be the answer to a question like ‘Where is the book?’ With *Masada kitap duruyor*, however, this message can be understood as the answer to a question about what lies on the table. The latter example is meant to draw the hearer’s attention to the quality ‘book’ and not its quantity; there may be one or more books, because the singular word *kitap* does not give decisive information.

Now, the non-referential subject in the *Focus-Locus Construction* is always placed immediately before the verb (participle) and this is the main position used to draw attention to a constituent by putting it in focus, because it is indefinite, emphasized or because it contains “new” information.²⁰ These constructions are applied to make a statement about what is going on at a certain place (locus, location)—such

17 In this respect I follow Johanson (1991: 229). The notion of “individuality” is relevant for *bir* + *noun* and *noun* + *-lEr*, but not for *noun* + \emptyset .

18 This is the type of noun phrase unspecified for definiteness and which can be used to form nominal compounds, see Van Schaaijk (1992, 2002).

19 Johanson (1991) uses the term “mitteilungsperspektiv”, which is more or less equivalent with the functional perspective from which language “chunks” are presented.

20 For constituent ordering in the sentence, see Van Schaaijk (1983, 1998, 2001).

statements are presentative—and put the emphasis (focus) on the quality of the subject of the event; the attention is drawn for instance to ‘land where WATER flows and BIRDS sing’ in (7a), to a ‘kitchen where PASTRIES are made’ in (8a), and to ‘concrete buildings where all sorts of RUSTY IRON sticks out of their columns’ in (25b).²¹

Interestingly, temporal expressions make full use of this type of construction by applying what are known as locational metaphors. A location metaphor obtains when an event noun is used to denote “location in time and space”. In this way, in (26a) *carousal* can refer to both the *place where* and *time when* certain things happen, and *school* in (26b) may stand for a place (as in: *at school*) and for a period of time (as in: *during school*) where certain events take place. Particularly in combination with impersonal passives, Turkish makes extensive use of this possibility. Some examples:

- (26) a. *Aynı mekân-da ye=nil-en, iç=il-en, müzik dinle=n-en, dans ed=il-en*
 same space-LOC eat-PAS-SF drink-PAS-SF music listen-PAS-SF dance-PAS-SF
ve kanepeler-e uzan=il-ip felsefe tartış=il-an
 and canapé-PLUR-DAT lie.down-PAS-CONV philosophy discuss-PAS-SF
son Roma sempozyum-lar-ın-dan bu yana 2000 küsur
 last Roman carousal-PLUR-ABL since 2000 odd
yıl geç-ti.
 year pass-PAST1
 ‘Since the last Roman carousel where people ate, drank, listened to music and discussed philosophy, having lied down on canapés, an odd 2000 years have passed.’
- b. *Var mı, yabancı dil-le eğitim yap=il-an*
 exist Q foreign language-INS teaching do-PAS-SF
okul-lar-ımız-da?
 school-PLUR-POSS1S-LOC
 ‘Is there [any], in our schools where teaching is done in a foreign language?’
- c. *En önemli şey-in barış ol-duğ-u belirt=il-en*
 most important thing-GEN peace be-OF-POSS3S state-PAS-SF
açıklama-da ...
 commentary-LOC
 ‘In the commentary, in which it was stated that the most important thing is peace ...’

21 Özçelik (2014) goes into the relation between prosody and focality in Turkish relative clauses. He analyses focused constituents in terms of “movement” which occurs mostly on prosodic and not on syntactic grounds. Similarly, the words in this paragraph that have strong emphasis are capitalized.

Hence it is not surprising that temporal nouns too have a strong correlation with the notion of place. Note that in English the relative connector can take different shapes: *when*, *in/on which*, and *that*. Compare:

- (27) a. *Bir hevenk muz ve dört kasa şampanya tüket=il-en gece ...*
 a bunch banana and four crate champagne consume-PAS-SF night
 ‘The night when a bunch of bananas and four crates of champagne were used up ...’
- b. *Kabile-yle yaşa=n-an gün-ler-in arasında hiç bir ayırım*
 tribe-INS live-PAS-SF day-PLUR-GEN between at all a difference
yok-tu.
 not.exist-PROJ
 ‘There was no difference between the days during which one lived with the tribe.’
- c. *Çık-ma-mız-a izin ver=il-en hafta-nın ikinci gün-ü-ydü.*
 go.out-NOM-POSS1P-DAT allow-PAS-SF week-GEN second day-POSS3S-PROJ
 ‘It was the second day of the week in which/that permission was given to go out.’

5. Evaluation

The leading question in this section is how the characterization presented above relates to the findings of Haig (1998:184). His main points are as follows.

1. The first generalization is that the verbs in his data are overwhelmingly low-activity intransitives, denoting events in which the subject has minimal control over that event. A second category of verbs is that of passivized transitive predicates, where the grammatical subject also has no control over the event (cf. Haig 1998a: 174). This he summarized by saying that: *Subjects are overwhelmingly the subjects of low control combinations, i.e. the non-human subjects of low-activity intransitives!*

From this formulation it is not quite clear to me how exactly we should understand “low CONTROL combinations” and “low-activity intransitives”. To begin with the latter qualification, intransitives are of course verbs with one argument only, and in all examples the majority of subjects in my own data are indeed formed by nouns denoting non-humans. Apart from a feature such as HUMAN (e.g. *insan* ‘human, man’ and *öğretmen* ‘teacher’), we find features such as ANIMAL (e.g. *bülbül* ‘nightingale’, *kuş* ‘bird’, *köpek* ‘dog’, *balık* ‘fish’, *fare* ‘mouse’, *sivrisinek* ‘mosquito’, *maymun* ‘monkey’); PLANT (e.g. *meşe* ‘oak’, *mantar* ‘mushroom’, *çay* ‘tea’); (natural) phenomena (e.g. *güneş* ‘sunlight’, *rüzgar* ‘wind’, *ışık* ‘light’, *ateş* ‘fire, fever’); SUBSTANCE (e.g. *su* ‘water’, *duman* ‘smoke, vapour’, *alev* ‘flame’, *yangın* ‘fire’, *kan* ‘blood’, *ter* ‘sweat’, *toz* ‘dust’, *et* ‘meat’, *buz* ‘ice’, *demir* ‘iron’, *yaş* ‘tear’, *yağmur* ‘rain’); THING (e.g. *kapı* ‘door’, *çeşme* ‘well’, *tren* ‘train’, *balta* ‘axe’, *pasta* ‘pastry, cream-cake, pie’) and ABSTRACT (e.g. *söz* ‘speech’, *sozcük* ‘word’, *uyku* ‘sleep’).

When it comes to the notion of control, not only humans can be controllers; depending on the type of verb, animals, too, to a certain extent, can have the power to determine whether or not some action will occur. Is there really a great difference in control between the singing nightingale(s) in (22) and the eating mouse/mice in (20a)? Is it this really a relevant factor? Does not the type of verb involved also play a role? When we look at the verbs that form the subject participle, the majority denote a process, in other words, a non-controlled event.²² These include: *bit-* ‘to grow’, *sac-* ‘to strew, scatter’, *ak-* ‘to flow’, *yaşa-* ‘to live’, *yetiş-* ‘to grow’, *karış-* ‘to mix’, *doğ-* ‘to be born; to rise’, *sız-* ‘to trickle, leak’; *düş-* ‘to fall’, *yağ-* ‘to precipitate’. The verbs *öt-* ‘to sing’, *gir-* ‘to go in’, *çık-* ‘to come out’, *kaç-* ‘to enter’, and *geç-* ‘to cross’ can also be included as denoting a non-controlled event, but only on the proviso that the subject is non-human. With a human subject they denote a controlled action, and often undergo a shift in meaning. Furthermore, typical “control verbs”, verbs the first argument of which is a controller, are transitive verbs. And they do exist, albeit only under certain conditions.

None of this brings us any closer to answering the question of what is meant by low-activity. I believe the main point is that we are dealing with a construction which is solely meant to bring some phenomenon or event to the attention of the hearer. And control does not add much to the description of the type of verbs and subjects. In other words, it does not have enough explanatory power for a full account of the linguistic facts.²³ Moreover, it is hard to imagine how control and low-activity would relate to all those instances where people are involved, but not explicitly expressed, as with the impersonal passives in (8b), (25 a–c) and (27 a,b).

Another factor which is not taken into account is the way the head-noun of the subjects can be interpreted and what consequences any possible differences might have for the final analysis. Take for instance *güneş* ‘sun’, which in (9a) can be assumed to mean ‘the sun’ in the sense of ‘celestial body’, but in (6d) refers to the sun’s radiation in the form of ‘sunlight’. For words such as *demir* ‘iron’ and *buz* ‘ice’ the first sense (interpretation) that springs to mind is probably some ‘matter’, some uncountable ‘stuff’. In their plural forms these words obtain a more concrete interpretation. *Demir-ler* in (24a) becomes ‘pieces of iron (as used in construction)’ or, more precisely, ‘reinforcing iron bars’, and *buz-lar* in (9d) can be understood as ‘ice chunks’ or ‘icicles’. What we see here is a kind of ‘conversion’, shifting the interpretation from ‘matter’ or ‘substance’ to ‘object’.²⁴ This can also occur the other way around. Although *balık* ‘fish’ is an animal and hence a potential controller, in (23b) it is not some ‘individual’ or ‘animal’ that is

22 I use notions such as control, process, and action in accordance with the typology of states of affairs by Dik (1989: 89–109).

23 A domain where the notion of “control” is indispensable is subordination. See for instance Haig & Slodowicz (2006) and Van Schaaik (2014).

24 A detailed analysis of the interplay between meaning and interpretation can be found in Ebeling (2005: 97–107).

the pivot of the statement, but instead *balık* ‘fish’ here means ‘what you catch’, or a ‘substance’ to be used as food or merchandise.

2. Another parameter that Haig brings forward for a characterization of subjects is individuation, a notion which refers to the extent an NP signifies a specific quantity or amount. In his view, a noun with *bir* or with a numeral counts as “precisely quantified”, a noun in the plural as “not precisely quantified” and a bare noun as “not quantified”. The values “precisely quantified” and “not quantified” correspond to the extremes “high” and “low” on his scale of individuation (cf. Haig 1998: 176). The conditions under which precisely quantified subjects occur is specified as: *Precisely quantified subjects, e.g. with bir, never occur as subjects, unless the relative clause contains an anticipatory possessive.*

There are counterexamples to this claim, for instance (25b). Moreover, sole the fact that combinations such as *bir* + N and N + *-Ier* do occur makes it clear that individuation is indeed something observable. On the other hand, this is not directly related to definiteness, but rather with “individual singularity” and “individual plurality”, as described by Johanson (1991: 226). Such combinations are not primarily intended as means for establishing reference, but also can be used to convey a general picture or image of quantity: one, more, or unspecified.

Conversely, Haig might have thought that the possibility to make reference is correlated to the intention of doing so. This is what he shows on the basis of example (28), taken from Haig (1998a: 185).²⁵

- (28) [Ağz-in-a bardak daya-n-an] çocuk şaşır-mış, kork-muş-tu.
mouth-POSS3S-DAT glass rest-PAS-SF child be.confused be.afraid-PAST2- PROJ
‘The child [with the glass resting against his mouth] was confused and afraid.’

Haig’s comment on the translation of this example is:

I translated *bardak*, even though it is a semi-subject, with *the glass*, and I see no other option: The glass concerned has been unequivocally identified, indeed it has been a co-topic of much of the preceding text. Note however that there is no further mention of the glass beyond this point in the text.

In my opinion Haig’s assertion that there is “no other option” is correct, but the reason advanced is incorrect. The choice in English is between ‘a glass’ and ‘the glass’, and choosing the former option would possibly lead to confusion. But it is not true that reference for the purpose of identification plays a role here. What is expressed in Turkish by (28) is no more than ‘That child with *some glass* against his mouth was confused and afraid’. The core of the argument here is the general image:

25 During a ceremony an entertainer takes a sip from a glass of water and then holds the glass to a young boy’s mouth, saying that he too will now drink from it. The boy’s mother, horrified at the thought of her son drinking from the same glass, protests loudly. Example (28) describes the situation at this point.

“child-with-glass-on-mouth”, and the fact that the particular glass happens to be one and the same as the one that the conjurer has been drinking from is true, but irrelevant for the creation of the image. It is most likely very relevant for the mother of the child, but that has nothing to do with the image created by *bardak dayanana* ‘glass resting’. The *purpose* of the latter fragment (in Turkish) is not one of *identification*, but rather, of creating an image. This is corroborated by the lack of further mention in the original text, as quoted above.

A possible cause of the confusion is that the relation *definiteness* and *identification* should apparently be taken as a one-to-one relationship that always holds. Obviously this is not so. In English one can say for instance ‘John plays the piano very well’, and in such a statement *the piano* denotes nothing more than a certain type of musical instrument, and is not meant to refer to some particular piano. Even if one finds an old friend performing a piece of music by Satie, one could say with unfeigned surprise: “Gosh, I never knew that Bill played the piano!”, and *the piano* would not refer to the instrument he is actually playing, but to his particular skill: (I never knew that Bill) ‘knows how to play an instrument called piano’. Similarly, in mentioning the means of transportation one utilizes, a definite article is quite often required in Dutch, whereas in English there is no article: *Gaan jullie met de trein, de bus, of met de auto?—We gaan met de fiets en te voet* ‘Will you travel by train, bus or car?—We’ll go on bicycle and on foot’. Also (9a) contains a fragment, *güneş* ‘sun’ which is duly translated by ‘the sun’, and indeed, there is no other possibility than to do so, for this word stands for a celestial body with unique reference; their referent is inherently identifiable and such words take the definite article. But to say for that reason that *güneş* ‘sun’ in (9a) is definite as well would be pedantic.

3. *Personal pronouns, nouns with normal possessive morphology, demonstratives, nouns modified by adjectives created from the adjectival suffix -ki are never subjects.*

These observations are correct and bearing in mind the idea of non-referential subjects they should not be surprising. The referents of all the categories mentioned here are intrinsically identifiable and hence definite. This makes phrases the head of which is based on such nouns and pronoun referential, and thus, they are to be excluded from subjecthood in *Focus-Locus Constructions*.

4. *Subjects occur with only two transitive predicates: sok- ‘sting’ and bas- ‘overgrow’. Even here, their occurrence is bound to extremely specific conditions: They are minimally INDIVIDUATED, the relative clause contains an accusative-marked constituent with an anticipatory possessive, and the subject is non-human.*

It is absolutely true that the majority of verbs in the *Focus-Locus Construction* are intransitive, but there are nevertheless hardly any reasons why in principle only *sok-* ‘sting’ and *bas-* ‘overgrow’ would qualify. In (20a) we find *ye-* ‘to eat’, and in (20c) *kir* ‘to break’, while *bas-* occurs in several transitive collocations: *ot bas-* ‘to overgrow with weeds’, *su bas-* ‘to flood’, *ter bas-* ‘for sweat to break out’ as in (20d), *ateş bas-* ‘to be hot all over; to glow’ in (20e), *sel bas-* ‘to flood’ in (29a), *sis bas-* ‘to fog up’ in (29b).

- (29) a. *Kurkuru'nun sel-ler bas-an toprak-lar-ın-da gece korkutucu-dur.*
 K.GEN flood-PLUR bas-SF land-PLUR-POSS3S-LOC night scary-EMPH
 'In the lands of Kurkuru where torrential floods occur, the night is really scary.'
- b. *Bahçe-ler-den bir koku yüksel-iyor, nem, yosun, ölmüş yaprak*
 garden-PLUR-ABL a smell rise-PRES1, moisture, moss, dead leave
koku-su, sonbahar başla-dığında hafif sis bas-an
 smell-CM autumn start-TEMP light mist bas-SF
akşam-lar-a özgü.
 evening-PLUR-DAT characteristic
 'From the gardens rises a smell, moisture, moss and the smell of dead leaves;
 when the autumn starts, this is characteristic of evenings with a shallow fog
 coming up.'

6. Surviving innovations

By way of a conclusion, a few words can be said about the conditions for the use of this construction. Erdal (1981: 30) quotes Deny (1921) who points out that relativization without possessive suffixes referring to the subject is the only "original" type of relativization in Turkish. Erdal also quotes Underhill (1972: 87–99), who was of the opinion that constructions with the OP (in *-DİK*) are a "relatively recent Turkish innovation" and that "we may suppose that the retention of the more archaic *-En* construction [...] must be connected with the fact that they are proverbs". Neither Erdal nor Haig agree with this point of view, and indeed, the huge number of "counterexamples" that easily can be gathered from electronic text corpora convincingly show that the N1+SP+N2 type of structure is productive and viable.

The question is why? In other words, why is this type of construction the survivor of an innovation that has led to the differentiation between SP and OP? I think that the answer is quite simple: this type of construction has survived in its specific shape because it fills a need. That is, it has specialized into a presentative statement by narrowing down the type of noun phrases involved on either side of the participle.

Whereas in ancient times possible ambiguities in transitive verbs, as in *ot yiyen at* 'grass eating horse' and *at yiyen ot* 'grass horses eat', could be resolved on the basis of commonplace knowledge ('horses eat grass' and not the other way around), because intransitives such ambiguities did not exist anyhow. A further specialization by using place nouns for N2 and non-referential noun phrases for N1, and thereby surpassing the level of a trivial construction, made it a very good candidate for survival.

Abbreviations

1S	copula, 1st person singular	PAS	passive
3P	copula, 3rd person plural	PAST1	past tense (1): <i>-TI</i>
ABL	ablative case	PAST2	past tense (2): <i>-mİş</i>
ACC	accusative case	PLUR	plural
CM	compound marker	POSS1P	possessive 1st person plural
CONV	converb	POSS1P	possessive 1st person plural
DAT	dative case	POSS1S	possessive 1st person sing
EMPH	emphatic suffix	POSS2S	possessive 2nd person sing
FUT	future tense	POSS3P	possessive 3rd person plural
GEN	genitive case	POSS3S	possessive 3rd person sing
INF	infinitive	PRES1	present tense (1)
INS	instrumental case	PRES2	present tense (2)
LOC	locative case	PROJ1	projection suffix past
NEG	negation marker	Q	question marker
OF	form of OP: <i>-TIK</i>	SF	form of SP: <i>-(y)En</i>
OP	object participle	TEMP	temporal suffix

References

- Barker, C. & Hankamer, J. & Moore, J. 1990. *Wa* and *Ga* in Turkish. In: Dziwirek, K. & Farell, P. & Mejias-Bikandi, E. (eds.) *Grammatical relations: A cross-theoretical perspective*. Stanford (California): Center for Study of Language and Information. 21–43.
- Cagri, I. M. 2009. Arguing against subject incorporation in Turkish relative clauses. *Lingua* 119, 359–373.
- Ciopiński, J. 1969. Remarques sur les constructions syntactiques du type *bülbül öten yer* et leur réalisation dans la langue Turque. *Folia Orientalia* 10, 59–63.
- Csató, É. Á. 1985. A syntactic analysis of participial constructions in modern Turkish. In: *Beşinci Milletlerarası Türkoloji Kongresi, 23–28.09.1985. Tebliğler 1: Türk Dili 1*. İstanbul: Edebiyat Fakültesi Basımevi. 39–56.
- Csató, É. Á. 1996. A typological review of relative clause constructions in some Turkic languages. In: Rona, B. (ed.) 1996. *Current issues in Turkish linguistics. Proceedings of the Fifth Annual Conference on Turkish Linguistics*. School of Oriental and African Studies, 15-17 August, 1990. Ankara: Hitit Yayinevi. 28–32.
- Dede, M. 1978. Why should Turkish relativization distinguish between subject and non-subject head nouns? In: *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* 4. 67–78.
- Deny, J. 1921. *Grammaire de la langue turque*. Paris: Ernest Leroux.
- Dik, S. C. 1989. *The theory of functional grammar 1: The structure of the clause*. Dordrecht: Foris.
- Ebeling, C. L. 2006. *Semiotaxis. Over theoretische en Nederlandse syntaxis*. [Semiotaxis. On theoretical and Dutch syntax]. Amsterdam: Amsterdam University Press.
- Erdal, M. 1981. Turkish participles and the absence of reference. In: Young, D. W. (ed.) 1981. *Studies presented to Hans Jakob Polotsky*. East Gloucester, Mass.: Pirtle & Polson. 21–49.

- Erkman-Akerson, F. & Özil, Ş. 1996. /en/ and /dıđı/ crossings—Genitival NPs and sentential clauses as relative clause subjects. In: Konrot, A. (ed.) *Modern studies in Turkish linguistics. Proceedings of the Sixth International Conference on Turkish Linguistics. Eskişehir, 12-14 August, 1992*. Eskişehir: Anadolu University. 1–12.
- Göksel, A. & Özsoy, A. 2003. *dA*: A focus/topic associated clitic in Turkish. *Lingua* 113, 1143–1167.
- Haig, G. 1998a. *Relative Constructions in Turkish*. (Turcologica 33). Wiesbaden: Harrassowitz.
- Haig, G. 1998b. On some strategies for case recovery in Turkish relativization. In: Johanson, L. (ed.) *The Mainz meeting. Proceedings of the Seventh International Conference on Turkish Linguistics, August 3–6, 1994*. Wiesbaden: Harrassowitz. 299–320.
- Haig, G. & Slodowicz, S. 2006. Control in Turkish non-finite complements. In: Yağcıođlu, S. & Deđer, A. (eds.) *Advances in Turkish linguistics. Proceedings of the 12th International Conference on Turkish Linguistics, 11–13th August 2004, Dokuz Eylül University, Izmir*. Izmir: Dokuz Eylül Yayınları. 165–177.
- Hankamer J. & Knecht, L. 1976. The role of the subject/non-subject distinction in determining the choice of relative clause particle in Turkish. In: *Papers from the Sixth Meeting of the North Eastern Linguistics Society*, 31.10–2.11.1975. [Montreal Working Papers in Linguistics 6, May 1976] McGill University. 123–135. [Also published in: Aissen, J. & Hankamer, J. (eds.) *Harvard Studies in Syntax and Semantics* 2. (1976)].
- Hannay, M. 1985. *English existentials in Functional Grammar*. Dordrecht: Foris.
- Hovdhaugen, E. 1975. Relative clauses in Turkish. In: *1. Türk Dili Bilimsel Kurultayına sunulan bildiriler 1972*. (Türk Dil Kurumu Yayınları 413). Ankara. 551–554.
- Jansky, H. 1943. *Lehrbuch der Türkischen Sprache*. Leipzig: Harrassowitz.
- Johanson, L. 1991. Bestimmtheit und Mitteilungsperspektive im türkischen Satz. In: Johanson, L. 1991. *Linguistische Beiträge zur Gesamtürkologie*. Budapest: Akadémia Kiadó. 225–242.
- Kononov, A. N. 1956. *Grammatika sovremennogo tureckogo literaturnogo jazyka*. Moskva: Izdatel'stvo Akademii Nauk SSSR.
- Kornfilt, J. 2000. Some syntactic and morphological properties of relative clauses in Turkish. In: Artemis A. et al. (eds.) *The syntax of relative clauses*. Amsterdam: Benjamins. 121–159.
- Knecht, L. 1979. The role of the genitive suffix in relative clauses: a reply to Dede. In: *Proceedings of the 5th Annual Meeting of the Berkeley Linguistics Society*. Berkeley. 180–197.
- Lewis, G. 1965. *Teach yourself Turkish*. London: English Universities Press.
- Lewis, G. 1967. *Turkish grammar*. Oxford: Clarendon Press.
- Lyons, J. 1967. A note on possessive, existential, and locative sentences. *Foundations of Language* 3, 390–396.
- Mackenzie, J. L. 1992. Places and things. In: Fortesque, M. & Harder, P. & Kristofferson, L. (eds.) 1992. *Layered structure and reference in a functional perspective*. Amsterdam: John Benjamins. 218–231.
- Özçelik, Ö. 2014. An antisymmetric analysis of Turkish relative clauses: Implications from prosody. *Turkic Languages* 18, 247–270.
- Özsoy, S. 1994. Türkçede ortaç yapısı. *Dilbilim Araştırmaları*. Ankara: Hitit Yayınevi. 21–30.

- Öztürk, B. 2008. Relativization strategies in Turkish. In: Boeckx, C. & Ulutaş, S. (eds.) *Proceedings of Workshop on Altaic Formal Linguistics 4*. (Harvard University, Cambridge, May 2007), MITWPL, MIT: Cambridge. 241–253.
- Öztürk, B. 2009. Incorporating agents. *Lingua* 119, 334–358.
- Schaaik, G. van 1983. A functional analysis of aspects of Turkish grammar. MA Thesis. General Linguistics, University of Amsterdam.
- Schaaik, G. van 1992. The treatment of Turkish nominal compounds in FG. In: Fortescue, M. et al. (eds.) 1992. *Layered structure and reference in a functional perspective*. Amsterdam: Benjamins. 231–252.
- Schaaik, G. van. 1996. *Studies in Turkish grammar*. (Turcologica 28.) Wiesbaden: Harrassowitz.
- Schaaik, G. van. 1998. ‘İşlevsel Dilbilgisi’ nedir? [What is Functional Grammar?]. *Dilbilim Araştırmaları 1998*. Ankara. 9–25.
- Schaaik, G. van 2001. What is Functional Grammar? In: Schaaik, G. van 2001. *The Bosphorus papers. Studies in Turkish grammar 1996–1999*. İstanbul: Boğaziçi University Press. 37–56.
- Schaaik, G. van 2002. *The noun in Turkish. Its argument structure and the compounding straitjacket* (Turcologica 49.) Wiesbaden: Harrassowitz.
- Schaaik, G. van 2011. Place nouns as compound heads: a short story of fake postpositions. *Turkic Languages* 14, 206–238.
- Schaaik, G. van 2014. Complications in Turkish complementation: for Eva. In: Demir, N. & Karakoç, B. & Menz, A. (eds.) 2014. *Turcology and linguistics. Éva Ágnes Csató Festschrift*. Ankara: Hacettepe Üniversitesi Yayınları. 401–416.
- Schaaik, G. van (in progress) *Turkish grammar XXL*.
- Underhill, R. 1972. Turkish participles. *Linguistic Inquiry* 3, 87–99.
- Zimmer, K. 1987. Turkish relativization revisited. In: Boeschoten, H. & Verhoeven, L. (eds.) 1987. *Studies on Modern Turkish. Proceedings of the Third International Conference on Turkish Linguistics*. Tilburg: Tilburg University Press. 57–61.
- Zimmer, K. 1996. Overlapping strategies in Turkish relativization. In: Rona, B. (ed.) 1996. *Current issues in Turkish linguistics. Proceedings of the Fifth Annual Conference on Turkish Linguistics*. School of Oriental and African Studies, 15–17 August, 1990. Ankara: Hitit Yayınevi. 159–164.

Some remarks on viewpoint operators in Turkmen

Sema Aslan Demir

Aslan Demir, Sema 2017. Some remarks on viewpoint operators in Turkmen. *Turkic Languages* 21, 107–114.

In this paper, Turkmen postterminal markers are discussed. Turkmen belongs to the Southwestern branch of Turkic languages and is mostly spoken in Turkmenistan. Although Turkmen is an Oghuz language, it also shares some common areal features with the Northwestern (Kipchak) and Southeastern (Karluk) branches of Turkic languages. This can also be observed in the inventory of Turkmen markers of postterminality, and in some respects this situation can be interpreted as a deviation from the typology of Oghuz languages. The work will first focus on the postterminal marker *-An* (< *-GAn turur*) in competition with the other postterminal marker *-(I)pdlr* (< *-Ip turur*). After this, *-An däldir*, the negative form of *-An*, will be discussed in competition with *-mAndlr*.

Sema Aslan Demir, Hacettepe University, Faculty of Letters, Department of Turkic Languages and Literatures, Ankara, Turkey. E-mail: semaaslan@hacettepe.edu.tr

0. Introduction

The present work aims to present some interesting features of viewpoint aspect markers in Turkmen. The paper does not include evaluations of all viewpoint aspect markers in Turkmen; instead some observations and questions will be presented about the postterminal markers. Turkmen belongs to the Oghuz branch of Turkic languages and is mostly spoken in Turkmenistan. It has also speakers in Iran, Afghanistan, Uzbekistan, Tajikistan, and some parts of north Caucasus. Although it is an Oghuz language, it also shares some common areal features with the Northwestern (Kipchak) branch and the Southeastern (Karluk) branch of the Turkic languages, which can also be observed in the inventory of Turkmen viewpoint aspect markers, and in some respects this can be interpreted as a deviation from the typology of the Oghuz languages. In this work, Lars Johanson's theory and terminology (1971, 1994 and 2000) will be employed when analyzing data. Examples will be given both in the present Turkmen orthography and in transcription, because some important phonetic features of Turkmen, such as vowel lengths, labial harmony, etc., are not marked by the standard orthography.

Unlike the Slavic languages, the term aspect is here not used for denoting actional contents or situation types; rather, it denotes the viewpoint perspectives, envisaging the events in various ways, in relation to their relevant limits. Viewpoint operators offer different choices for envisaging and presenting events, by opening

perspectives on them. They operate on the actional contents (Johanson 2000: 30–31). There are three main aspectual perspectives that represent different ways of envisaging the event:

The intraterminal perspective: envisaging the event within its limits.

The postterminal perspective: envisaging the event after the transgression of its relevant limit.

The adterminal perspective: envisaging the event at the very attainment of its relevant limit (Johanson 2000: 29).

The viewpoint operators can highlight a point situated within the limits of the event, a point situated after the decisive limit of the event or the decisive limit itself (Johanson 2001: 54)

1. The competition between *-An* and *-Ipdlr*

There are two postterminal markers in the Turkmen aspectual system: *-An* and *-Ipdlr*. It must be stated that while *-Ipdlr* is a widespread postterminal marker in modern Turkmen, *-An* is less commonly used. The morpheme *-An* can be seen in tales, proverbs, and some folkloric narrations, but today *-Ipdlr* is much more common both in daily speech and literature. If we disregard some modal expressions based on *-An*, it seems that the frequency rate of *-An* has been gradually decreasing.

The descriptions of *-An* provided by Baskakov et al. (1970) and Söyegow et al. (2000)—also including findings of earlier works—are the followings:

-An signals that the speaker refers to events, which happened long before, in a way as if they were seen by the speaker. This marker is mostly used in tales and rumors. (Söyegow et al. 2000: 261–262)

-An refers to the past events, which occurred in an uncertain, indefinite past. Although the speaker did not witness the event, he is sure of its realization. (Baskakov et al. 1970: 268)

Let me reformulate the statements implicit in the above descriptions in the following way:

There is an event (E)

E happened in an indefinite past

Speaker did not witness the realization of E

Speaker explains E as if he had seen it

Speaker sure of the realization of E

These statements create some problems. How can a speaker express an unwitnessed event, which happened in an uncertain past, as if he had seen it or was sure of its

realization? In other words, how can a speaker encode a past event which he actually did not see, in a way that is as if he had seen it? See the following examples.

- (1) *Bu gün onuň geyimi-gejimi päkize, başy bolsa daralan.* SE. 154
 /Bu gün onuň geymi-gejmi pä:kide başi bolθo daralan/
 ‘Today her clothes are clean and (it is observed that) her hair has been combed.’
- (2) *Han kaka, men hiç ýere gitjek däl. Okajak. Onsoňam indi giç, meniň adym bu ýerde okuwçy diýlip ýazylan.* BT. 135
 /Xa:n qa:qa: men hi:ç yere gitjek gä:l, oko:joq. Onsoňom inni gi:ç, meniň a:dım bu yerde okuwçı diyliþ yaðlan/
 ‘Khan daddy, I will not go anywhere. I will study. Henceforth, it is late, (it is seen that) my name has (already) been written here as a student.’
- (3) *Gapynyň agzynda seki, üstünde haly düşekler ýazylan.* TDG. 262
 /gapi:nıj aýdınna θeki üθθünne xa:lı düşekler yaðlan/
 ‘(It is seen that) there is a sofa in front of the door; the mattresses, which were made of carpet, have been laid down on it.’

As these examples illustrate, the *-An* marker highlights the postterminal phase of the events, namely, the postterminal situation which results from the transgression of the relevant limit of the event observed. No past phase of the event appears in the range of vision. The event has already disappeared, but some observable traces remain. These traces are adjective-like qualities or descriptive properties, resulting from the transgression of the relevant limit. In examples 1–3, the postterminal verb forms *daralan* ⟨comb-PASS-POST.3SG⟩ ‘be combed’, *ýazylan* ⟨write-PASS-POST.3SG⟩ ‘be written’ and *ýazylan* ⟨lay down-PASS-POST.3SG⟩ ‘be laid down’, encode the property like situation, which result from postterminality, i.e. from the transgression of the relevant limit. The state of the hair, which was combed, the state of the name being written, and the situation of the carpet being laid down are highlighted.

In examples 1–3, which are taken from novels written in standard Turkmen, there are high-focal postterminals, which acquire diagnostic readings. As pointed out by Johanson (2000), more focal anterior items are also inclined to be more diagnostic (Johanson 2000: 108). Let me remind the reader how the term focality is used in Johanson’s theory. Focality is a scalar notion. Intraterminals or postterminals may display higher or lower degrees of focality. High-focal postterminality means that the relevance of the event at the moment of encoding (orientation point) is more important than the event itself. They put high focus on the postterminal state. Their range of vision is narrow, restricted to what is still relevant of the event at the orientation point (Johanson 2000: 38, 108, 110; see also Csató 2000: 30–31). If we reanalyze the examples, we can conclude that

No past phase of the event appears in the range of the speaker's vision. But he can see its results or remaining adjective-like properties, which are valid at the moment of encoding.

The speaker gets information about the postterminal phase of the event usually through visual perception. (In Turkmen *-An* can mark perceptual evidentiality.)

In the examples, the diagnostic readings seem also to be connected with the verb forms, which are passives, such as *daralan* (comb-PASS-POST.3SG), *yazylan* (write-PASS-POST.3SG) and *yazylan* (lay down-PASS-POST.3SG)

I now revise Söyegow's and Baskakov's descriptions on the basis of the above findings.

There is an event (E)

E happened in an indefinite past (proposed: No past phase of the event appears in the range of vision)

Speaker explains E as if he had seen it (proposed: Speaker describes/reproduces not the event but its concrete results by perceiving them at the moment of encoding)

Speaker is sure of the realization of E. (proposed: Speaker gets information about the postterminal phase of the event usually through visual perception, and this provides a high degree of confidence in the realization of the event.)

In the meantime, especially in folk tales, *-An* also covers less diagnostic and more historic readings, which are more event-oriented (for similar functions of *-GAN* in Noghay see Karakoç 2005: 74). In such uses, the attention is not on the postterminal state of the event, rather on the event itself. In the following example, which is taken from a folk tale, *-An* denotes succeeding event chains historically. It must be stated that it is not usual to see such examples in modern Turkmen literature or colloquial Turkmen.

- (4) *Garyp piriň eliňi daňdyryp, howzuň ýanyna elten. Esgerlerine hem howzuň suwny boşatdyran. Pirden jaýyň açaryny alyp gapylary açan. Şol wagt jaýdan kyrk gyz bilen kyrk ýigit çykyp başlan. Şolar bilen patyşanyň ogly bilen gyzy hem çykan.* TM 44
 /Ġarī:p pi:riŋ elji daŋdīrip xowduŋ ya:nna elten. Eθgerlerinā:m xowduŋ suwnu boŋotturon. Pi:rden ja:yīŋ ačarnī alīp ġapīlarī ačan. Šol waġt ja:ydan qīrq ġī:δ wilen qīrq yigit čīqīp baŋla:n. Šolor bilen pa:šša:nīŋ oylu wilen ġī:ða:m čīqan./
 'Garyp bound the hands of the master and took him to the pool. He also made his soldiers drain the water of the pool. He took the house key from the master and opened the door. In that case, forty girls and forty young men began coming out of the building. Both the son and the daughter of the padishah also came out.'

2. The relation between degree of focality and source of information

Another question to be discussed in this paper is that of the relationship between the degree of focality and the source of information. It seems that if the basis of the

knowledge about the event or postterminal state of the event is “hearsay” or “reportive”, the postterminal markers are prone to be less focal and more historic. On the other hand, if the basis of the knowledge is perceptive or inferential, the postterminal markers are prone to be more diagnostic and high-focal. The plausibility of this idea needs to be studied further, but a similar tendency can also be observed in the use of the other postterminal marker *-(I)pdIr*. In example 5, the basis of the knowledge is inference, and a diagnostic reading can be obtained. In example 6, the basis of the knowledge is hearsay, and *-(I)pdIr* (the negative form is *-mAndIr*) signals low-focal postterminality in which the events are narrated according to the historical order, like $E1 > E2 > E3 > E4...$

- (5) *Men söyginiň ugrunda şeýle ile göz-gülban bolupdyryn. Emma şeýle ile gülban bolanymy özüm bilmän galypdyryn.* P 413
 /Men öýgü:nüň u:ýrunna şeýle i:le göđ-gülba:n bolupdurun. Emma: şeýle i:le gülba:n bolanıma ö:đüm bilmä:n ğalıpdırin/
 ‘My love had brought disgrace upon me to, embarrassing me in front of everyone. Yet, I wasn’t even able to comprehend my shame.’
- (6) *Eneleriniň aýtmagyna görä, Japbakyň kakasy hem şol agyr ýüke tap getirmän, olary ýaşlykda yetim galdyrypdyr. Dört çagaly dul oturan, girdeji gözünü ýitiren, bar ünsüni iýmite ataran ene çagajyklary açlyk belasyndan goramak üçin gije-gündiz ondan-oňa sümsünipdir, olary wagtly-wagtynda timarlamaga da ýagday tapmandyr. Şonuň üçin Japbakyň ata mährini görmän, ene terbiyesini alman, esasan öz peýwagtlaryna ösüpdirler.* J 13
 /enelerinij aytmaýna görä: ğapbaqlarınj qa:qaθa:m šol aýır yükö ta:p getirmä:n olorı ya:šliqda yeti:m ğa:llirıpdır. dö:rt ça:ya:lı dul oturon gi:rdeji göđnü yitiren ba: ünθünü iýmite ataran ene ça:ýajıqları açliq bel:θınnan ğo:romoq üçün gi:je-günnüđ onnon-oňo θümθünüpdür olorı waytli-waytınna ti:ma:rlama:ya da yayday tapma:nnır šonuň üçün ğapbaqlar ata mä:hriini görmä:n ene terbiyeθini alma:n eθa:θa:n ö:đ peywaytlarna öθüpdürlör/
 ‘According to their mother, the father of Japbak could not bear this heavy responsibility and left them as orphans. Their mother, who was widowed with four children, struggled to make a living; day and night she tried to find some food for her children. She could not give much attention to their moral attitude. Without their father’s love and mother’s training, they have grown up on their own.’

3. Competition between *-An dälDir* and *-mAndIr*

In Turkmen, the negative form of *-An* is *-An dälDir*. The negative form of *-An* was originally *-mAndyr* (*-Andır: -mAndIr*) (Çaryýarow 1969: 273) But in the standard language, *-mAndIr* is usually used as the negative form of *-Ipdlr*. In this way, the suppletive negative-positive morpheme pair *-Ipdlr* vs. *-mAndIr* has developed. The generally accepted assumption is that *-An dälDir* must have arisen from a need to avoid the ambiguity between the negative forms of *-AndIr* and the negative form of

-IpdIr (Clark 1999: 237). The negated form *-An dädIr* differs from *-mAndIr* in some particular respects. The first difference concerns indirectivity. *-mAndIr* expresses indirectivity in the sense that non-occurrence of the event or absence of the post-terminal state is perceived, inferred or heard by the speaker (for indirectivity, see Johanson 2000, 2003). On the other hand, *-An dädIr* removes indirective nuances by coding a higher degree of awareness and consciousness concerning the postterminal phase of the event. (For further details about the semantic values of *-AndIr* and its modal functions, such as presumption, strong assumption etc., see Aslan Demir 2014.) Secondly, unlike *-mAndIr*, what is negated by *-An dädIr* is not the occurrence of the event itself, but a negation of the postterminal state which would occur after the transgression of the relevant limit. The semantic contents of the *-mAndIr* and *-An dädIr* can be formulated as follows: *-mAndIr*: [(event + NEG)] + viewpoint aspect, +IND; *-An dädIr*: [(event + viewpoint aspect)] + NEG, -IND. In examples 7 and 8, no indirective nuance can be perceived. But in example 9, the occurrence of the event is recognized by the speaker with a sudden remembering; i.e. the speaker expresses it in an indirect way referring to a source. This can be translated into English as ‘obviously’, ‘apparently’. Examples 7 and 8 belong to literary language, and 9 comes from colloquial Turkmen.

- (7) *Bu ýüzük siziň barmagyňyza gelişer diýip aldym. Men muny ýöne ýere saýlan dädidirin.* MT 100
/bu yüdük šiđiň barmağyđa gelişer diyip allim. men munı yö:nö yere şayla:n nä:llirin/
‘I bought this ring with the thought that suits you. I have not chosen it in vain.’
- (8) Men olardan hiç wagtam *gorkan dädidirin.* GRŞ 52
/men olordon hi:č waytam gorqon nä:llirin./
‘I’ve never been scared of them.’
- (9) *Wah, diňläsim gelyän CD-mi ýanyma almandyryn.* CT
/wax, diňlä:θim gelyä:n CD-mi ya:nıma alma:nnırın/
‘Tut-tut. I have obviously forgotten the CD, which I want to listen to.’

4. Conclusion

I have presented some properties of the two Turkmen postterminal markers *-An* and *-IpdIr*. While *-IpdIr* is a widespread postterminal marker in standard Turkmen, *-An* is used in some particular discourse types. In modern Turkmen, both in colloquial and standard, *-IpdIr* is more common.

The marker *-An* highlights the postterminal phase of an event. The speaker often refers to the postterminal phase of the event based on knowledge gained through visual perception. Johanson states that “where the types *-GAn turur* and *-(V)p turur* occur together in one system, the former tends to be pure postterminal, whereas the latter has indirective meanings” (2000: 73). In Turkmen *-An* (< *-GAn turur*) covers

both historical and diagnostic readings. The diagnostic value occurs especially when *-An* is added to passive verb forms.

Another issue discussed in this paper is the competition between *-An däldir* and *-mAndlr*. The negated form *-An däldir* differs from *-mAndlr* in particular respects. The most salient difference concerns indirectivity. *-mAndlr* expresses indirectivity in the sense that the non-occurrence of the event or the absence of the postterminal state is expressed, inferred or heard by the speaker. On the other hand, *-An däldir* removes indirective nuances and encodes a higher degree of awareness and consciousness of the postterminal phase of the event.

References

- Aslan Demir, Sema 2014. *Görünüş kategorisi: Türkmençe örneği*. Ankara: Grafiker.
- Baskakov, Nikolaj Aleksandrovic et al. (eds.) 1970. *Grammatika turkmenskogo jazyka 1. Fonetika i morfoloģija*. Aškabat: Ylym.
- Clark, Larry 1998. *Turkmen reference grammar*. (Turcologica 34.) Wiesbaden: Harrassowitz.
- Csató, Éva Á. 2000. Turkish *-miş* and *imiş* items. Johanson, Lars & Utas, Bo (eds.) *Evidentials: Turkic, Iranian and neighboring languages*. Berlin & New York: Mouton de Gruyter. 29–43.
- Çaryýarow, Bäşim 1969. *Günorta-günbatar Türki dillerde işlik zamanlary*. Aşkabat: Ylym.
- Johanson, Lars 1971. *Aspekt im Türkischen. Vorstudien zu einer Beschreibung des türkei-türkischen Aspektsystems*. (Studia Turcica Upsaliensia 1.) Uppsala: Almqvist & Wiksell.
- Johanson, Lars 1994. Türkeitürkische Aspektotempora. In: Thieroff, Rolf & Ballweg, Joachim (eds.) *Tense systems in European languages*. Tübingen: Niemeyer. 247–266.
- Johanson, Lars 2000. Viewpoint operators in European languages. In: Dahl, Östen (ed.) *Tense and aspect in the languages of Europe*. Berlin & New York: Mouton de Gruyter. 27–187.
- Johanson, Lars 2001. On three dimensions of aspectual terminality. In: Bisang, Walter (ed.) *Aspects of typology and universals*. (Studia Typologica 1.) Berlin: Akademie. 53–62.
- Johanson, Lars 2003. Evidentiality in Turkic. In: Aikhenvald, Alexandra Y. & Dixon, R[obert] M. W. (eds.) *Studies in evidentiality*. (Typological Studies in Language 54.) Amsterdam & Philadelphia: John Benjamins. 273–290.
- Karakoç, Birsal 2005. *Das finite Verbalsystem im Nogaischen*. (Turcologica 58.) Wiesbaden: Harrassowitz.
- Söýegow, Myratgeldi (ed.) 2000. *Türkmen diliniň grammatikasy: Morfoloģiýa*. Aşkabat: Ruh neşriýaty.

Sources

- AH: www.azathabar.com/a/24351463.html
- BT: *Böwsülen tümlük*. Ak Welsapar. Stockholm: Gün Neşirýaty. 2004.
- GRŞ: *Göreş*. Çary Aşyr. Aşkabat: Türkmenistan Neşirýaty. 1986.
- J: *Japbaklar*. Berdi Kerbabaýew. Aşkabat: Türkmen Döwlet Neşirýat Gullugy. 2007.
- MT: *Mülli tahyryň hudaýlygy*. Ak Welsapar. Stockholm: Gün. 2006.
- P: *Perman*. Ata Gowşudow. Aşkabat: Türkmenistan Neşirýaty. 1989.
- CT: Colloquial Turkmen
- SE: *Saýlanan eserler*. Berdi Kerbabaýew. Aşkabat: Magaryf. 1992.

- TDG: *Türkmen diliniň grammatikasy: Morfologiýa*. Aşgabat: Ruh. 2000
- TM: *Türkmen masalları*. Ryoko Asano. Ankara Üniversitesi, Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi. 2010.

On the phonetic unpredictability denoted by some Old Turkic texts written in Syriac script. Or the encoding ambiguity intrinsic to the Aramaic writing

Delio Vania Proverbio

Proverbio, Delio Vania 2017. On the phonetic unpredictability denoted by some Old Turkic texts written in Syriac script. Or the encoding ambiguity intrinsic to the Aramaic writing. *Turkic Languages* 21, 115–151.

Starting from a close examination of an Old Turkic manuscript from the Tangut city of Xaraxoto (Inner Mongolia) written in Syriac script—an offshoot of the Aramaic alphabet, which exhibits a peculiarly low complexity in its graphemic set—the present contribution consists of an empirical description of a number of graphotactic “regularities” which occur in the aforementioned text.

The original goal of this article was simply to provide a rigorous, formal account—a static, model-theoretic description—of what a number of assumptions imply in terms of graphotactic constraints. However, by manipulating our primary linguistic source as a finite linear string of symbols, we finally reached the conclusion that such a task is only achievable to a very limited extent. This is due to the intrinsic phonetic unpredictability that derives from the encoding ambiguity of the Aramaic writing.

Delio V. Proverbio, Scriptor Orientalis, Senior Research Fellow for the African, Near and Middle Eastern Collections. Vatican Apostolic Library, Cortile del Belvedere, 00120 Vatican City State. E-mail: proverbio@vatlib.it

0.1. Introduction

The present contribution consists of an empirical description of a number of graphotactic “regularities” which occur in a given text. While the theoretical framework of a first, provisional data approach has implied a multi-tiered architecture (close to a classical Autosegmental Representation), its main scope turns out to be the analysis (and manipulation) of our primary linguistic source—whose writing system is an offshoot of the Aramaic alphabet, and exhibits a peculiarly low complexity in its graphemic set—as a finite linear string of symbols.

My enquiry starts from a close examination of an Old Turkic manuscript written in Syriac script.¹ This document, already published by Mutō (2008) and now again by Zieme (2015), is preserved at the Inner Mongolia Cultural Relics and Archaeology Research Institute (内蒙古自治区文物考古研究所) in Hohhot (呼和浩特), the capital of Inner Mongolia (China), and comes from Xaraxoto (Hēichéng, 黑城, the “Black City”).² I made a new transcription from the photographs published by Mutō in Yoshida, Jun’ichi & Chimeddhorji (2008: 359–363 no. 124–133), which I subsequently checked with Zieme’s text. Here is a concordance between Mutō’s and Zieme’s signatures:

Mutō (2008)	+Zieme (2015)		
124	T, fol. 1 recto	129	T, fol. 2 verso
125	T, fol. 1 verso	130	T, fol. 4 recto
126	T, fol. 3 recto	131	T, fol. 4 verso
127	T, fol. 3 verso	132	T, fol. 5 recto
128	T, fol. 2 recto	133	T, fol. 5 verso

Originally, the primary goal of this article was simply to develop (*si parva licet*) a formal description of a synchronically well-defined occurrence of a specific linguistic phenomenon, such as, the specific vowel harmony system we can detect in a certain Old-Turkic text (described as early as by Anderson, Vergnaud and Crothers in Vago 1980; to Rose & Walker 2014)—for example, but not necessarily, in terms of a set of constraints on “well-formed” representations. In other words the aim was to provide a rigorous, formal account—a static, model-theoretic description—of what a number of assumptions imply in terms of graphotactic constraints. Among these assumptions, we included a set of specific, local statements, such as the following, paradigmatic one:

- 1 I first wish to particularly thank Peter Zieme, who, as the scholar who by far is most deeply involved in Xaraxoto philological studies, was kind enough to supply me with the (then) still unpublished text of his comprehensive book, which encompasses the most relevant Old Uyghur texts from Central Asia. I extend my thanks to the Staatsbibliothek zu Berlin—Preussischer Kulturbesitz for access to and permission to reproduce images of some relevant fragments. All images are copyright from *Depositum der Berlin Brandenburgische Akademie der Wissenschaften in der Staatsbibliothek zu Berlin—Preussischer Kulturbesitz, Orientabteilung* (cf. the International Dunhuang Project website: <http://idp.bbaw.de/idp.a4d>). I also would like to give special thanks to Marcus Kracht for his insightful and in-depth remarks, and for his theoretically oriented observations. Thanks also to Adam Jardine for his contribution.
- 2 Cf. Zieme (2013: 100) “This manuscript is of great importance as it proves that a Turkic speaking community existed in Xaraxoto” during the Tangut domination (from 1037 until at least the conquest by Genghis Khan in 1226).

The Old-Turkic syntagm ⟨pyrk'yymn⟩, which actually occurs in the aforementioned Old Turkic document written in Syriac script, should have been read as $b(x)r-kāymān$ (*bergāymān*), while a fictitious syntagm such as *⟨pyrx'yymn⟩ should have been read as $b(x)rqaymān$.

In the end, we reached the conclusion that such a task turns out to be achievable to a very limited extent, due to the intrinsic phonetic unpredictability which derives from the encoding ambiguity (the set's incompleteness) of the Aramaic writing.

Thus—redirecting our in-depth purpose of strongly advocating the idea that there could be no real understanding of a given linguistic phenomenon without constructing a logical-mathematical model of it—we will eventually come to describe a deterministic Turing-machine, called OTR (Old Turkic Reader), which will be able to fulfill the task of appropriately decoding any 'syntagmatically meaningful' string of symbols that may occur in our text.

In building up our representational system of graphotactic and feature-licensing constraints, we will implement a light variety of Monadic Second Order Language (Rogers 1998; Graf 2010: 76–77; Jardine 2014: 4). Here is a list of set and individual variables which occur in Underlying and Surface Representation (our first and second level of representation, see below):

\mathfrak{B} = at the underlying representation level, a \mathfrak{B} -object is the syntagmatically meaningful string constituted by a sequence of n \mathfrak{S} -objects;

\mathfrak{S} = at the underlying representation level, a \mathfrak{S} -object is the syntagmatically meaningful string constituted by a sequence of 3 graphemes, i.e. the underlying representation of a syllable;

\mathfrak{g} = a \mathfrak{g} -object is the underlying representation of a grapheme;³

⟨x⟩ = a grapheme in Surface Representation;

σ = a syllable in Surface Representation;

$\gamma_{1(\sigma n-m)}$ = an enumerating, univocal label which occurs in Positional Representation;

followed by a list of binary place-predicates (beyond the usual Boolean operators and quantifiers):

\triangleleft = immediate domination

\triangleleft^* = domination

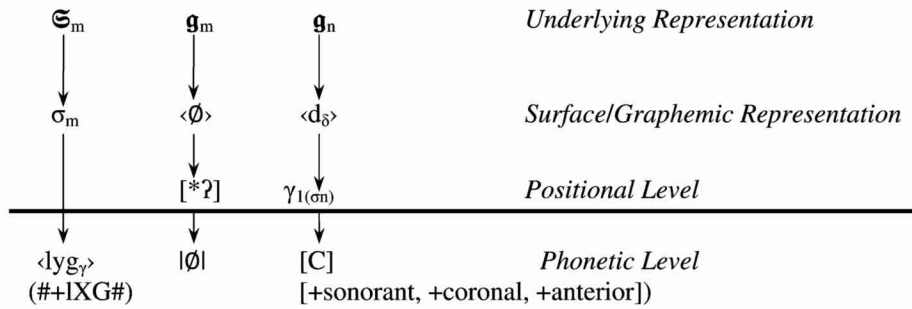
$<$ = linear precedence (left-of)

\nleftarrow = linear subsequence (not left of)

In the following, the atomic formulas $(x \triangleleft y)$, $(x \triangleleft^* z)$, “ x immediately dominates y ”, “ x (directly or indirectly) dominates (or is equal to) z ” (cf. Rogers 1998: 14–15;

3 For the sake of economy, *viz.* in order to not weigh strings down, in the following we will write $\mathfrak{g}_{(1)}$, or even \mathfrak{g}_n instead of $\mathfrak{g}_{n-m(1)}$, $\mathfrak{g}_{p(n)}$ —where the leftmost subscript, non-bracketed numerical variable expresses the ordinality of the dominating node.

Jardine 2014: 5) will invariably refer to a vertical linkage relationship, and will mean that y and (possibly) z are daughter nodes of x . These formulas encode / interpret a relationship which may be represented with the following graph:⁴



where $(\mathfrak{S}_m \triangleleft \sigma_m)$, $(\mathfrak{g}_m \triangleleft \langle \emptyset \rangle)$ and $(\mathfrak{g}_n \triangleleft^* [C])$.

The atomic formula $(x \triangleleft y)$ means that between x and y there is a relation of linear precedence; i.e. x precedes y along a linear sequence of symbols. Thus, given

$$\begin{array}{l} \mathfrak{S}_m = \langle \mathfrak{g}_m, \quad s\mathfrak{g}_m, \quad ss\mathfrak{g}_m \rangle \\ \downarrow \quad \quad \downarrow \quad \quad \downarrow \\ \sigma_m = \langle \langle x \rangle, \quad \langle y \rangle, \quad \langle z \rangle \rangle \end{array}$$

we may say that $\langle x \rangle \triangleleft \langle z \rangle$ —since, obviously, $\mathfrak{g}_m \triangleleft ss\mathfrak{g}_m$, and since $\mathfrak{g}_m \triangleleft \langle x \rangle$ and $ss\mathfrak{g}_m \triangleleft \langle z \rangle$ —while, z *does not* precedes y ($\langle z \rangle \not\triangleleft \langle y \rangle$).

0.2.

In contrast to other systems of representation, such as the three-level representation enunciated by Finlay (2009: 70), I introduced a fourth level: the Positional Representation. The reason is twofold: on the one hand, we have to manage enumerating labels such as $\gamma_{I(\sigma_m)}$, which are required to appropriately designate each syntagmatic positional location within the structure of our fixed syllabic grid (see below). On the other hand, we have to handle fictitious objects such as $[*?]$; this latter is here deemed to comply with the aforementioned fixed grid (again, see below). On the one hand, we assume that it is represented (by $\langle \emptyset \rangle$) in surface representation; in fact, it turns out to have no phonetic correspondence ($|\emptyset|$).

4 [The Anonymous Reviewer] grasped the point when insightfully observing that “The hierarchy [appears to be] constituted by abstraction. This looks a bit like stratificationalism. This is legitimate but quite unlike the ideas in [James] Rogers. I would bet that the structure of the representation will not be a tree or forest, as it is not uncommon to perform the transcription abstract-to-concrete (and back) using transducers”.

Let us now immediately come to the interpretation of a formula such as $(\mathbf{g}_m \triangleleft \langle \emptyset \rangle \triangleleft [*?] \triangleleft |\emptyset|)$. As a first step, taking as axiomatic that the unambiguous boundaries of any syntagmatically meaningful graphemic string may easily (and even algorithmically) be determined, we proceed by putting forward a syllabic model of such a (conveniently detected) syntagmatically meaningful graphemic string. Starting from the following statement, we will be able to generate a sequence of well-enumerated symbols, and therefore to syllabify the input syntagm according to the principle of maximizing the assignment to a *fixed* CVC syllable template.

Given the following, general assumptions:

$$(\forall \mathfrak{B}_p)(\forall \mathfrak{S}_{(1)}, \mathfrak{S}_{(n-m), n}) \left[\bigwedge_{i=1}^n \mathfrak{B}_p \triangleleft \mathfrak{S}_i \rightarrow \bigvee_{i < j = n} \mathfrak{S}_i = j \right] \quad (1a)$$

– which says that there can be at most n \mathfrak{S} under a \mathfrak{B} ;

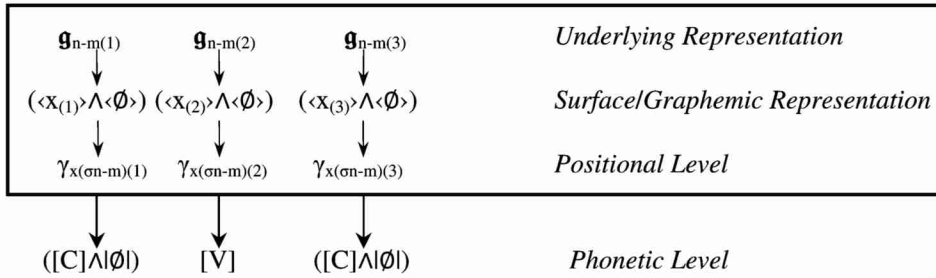
$$(\forall \mathfrak{S}_{n-m})(\forall \mathbf{g}_{(1)}, 23) \left[\left(\bigwedge_{i=1}^3 \mathfrak{S}_{n-m} \triangleleft \mathbf{g}_{(i)} \rightarrow \bigvee_{i < j < 4} \mathbf{g}_{(i) = j} \right) \wedge \left(\bigwedge_{i=1}^3 \mathfrak{S} \triangleleft \mathbf{g}_{(i)} \wedge \bigvee_{i < j < 5} \mathbf{g}_{(i) = j} \right) \right] \quad (1b)$$

which, roughly speaking, mean that every \mathfrak{S} -object is invariably and only constituted by the triplet $\langle \mathbf{g}_{(1)}, \mathbf{g}_{(2)}, \mathbf{g}_{(3)} \rangle$;⁵

Both statements (1c) and (1d) are found to be true:

$$(\forall \mathbf{g}_{n-m})(\forall x_{(1)}, 23) \left[x_{(1)} < x_{(2)} < x_{(3)} \rightarrow \left(\left(\bigwedge_{i=1}^3 \langle x_{(i)} \rangle \mathbf{g}_{n-m} \right) (\forall \mathbf{g}_{n-m} = \triangleleft \langle \emptyset \rangle) \right) \right] \quad (1c)$$

$$(\forall \mathfrak{S}_{n-m})(\forall \mathbf{g}_{(1)}, \mathbf{g}_{(2)}, \mathbf{g}_{(3)}) \left[\mathbf{g}_{(1)} < \mathbf{g}_{(2)} < \mathbf{g}_{(3)} \rightarrow \left(\left(\bigwedge_{i=1}^3 (\mathfrak{C})\mathbf{g}_{(i)} \vee \mathbf{g}_{(i)} = \triangleleft^* |\emptyset| \right) \wedge \left((\forall) \mathbf{g}_{(i)} \right) \right) \right] \quad (1d)$$



5 The two can loosely be expressed as follows: $(\forall \mathfrak{B}_p)(\mathfrak{B}_p = \langle \mathfrak{S}_1, \mathfrak{S}_{n-m}, \dots, \mathfrak{S}_n \rangle) \wedge (\forall \mathfrak{S}_{n-m} (\mathfrak{S}_{n-m} = \langle \mathbf{g}_{n-m(1)}, \mathbf{g}_{n-m(2)}, \mathbf{g}_{n-m(3)} \rangle))$.

These statements—captured by the above graph model = *mod-a*, a bi-dimensional model in which we consider the universe of Representations as coplanar with the universe of the Phonetic Utterances⁶—are axiomatic in the sense that, apart from being empirically motivated, they do not derive from other propositions.⁷

Here, we have to introduce a distinction between morphological syllabification *versus* automatic, rule-based syllabification.⁸ Let us consider two examples of syllabification which takes into account the morphological structure of Old Turkic: <pyrm lq> *bir* {-mAK} (131.1) \$CVC\$. \$CVC\$ *versus* <p lrm lq l̥ y̥> *bar* {-mA} {-Gay} (128.1), \$CVC\$. \$CV\$. \$CVC\$. As far as the last example is concerned, if we adopt the aforementioned automatic syllabification, based on the simple rule according to which all syllables should exhibit a consonant onset, we obtain the following pattern:

C	V	C	C	V	C	C	V	C
<p>	<y>	<r>	<m>	<l̥>	<ŋ>	<∅>	<l̥>	<y>

Arguing from the above grid, a vocalic-onset syllable appears to be a CVC syllable whose onset is represented by zero in graphemic/surface representation. In other words, since, according to (1b), under some empirically verifiable conditions (see immediately below, § 0.3), a certain \mathbf{g}_{n-m} -object may be represented by zero in surface representation—provided that the general statement $[(\mathbf{g}_{n-m(2\pm 1)} \triangleleft^* [C]) \rightarrow \top]$ implies that $(\exists \mathbf{g}_{n-m} | \mathbf{g}_{n-m(1)} \in \mathfrak{S}_{n-m} \wedge \mathbf{g}_{n-m(1)} \triangleleft^* [V])$ —vocalic-onset syllables are here conventionally represented as \$[*?][V][C]\$, with a fictitious glottal stop [?] in onset position, encoded by a zero-surface representation ($\mathbf{g}_m \triangleleft \langle \emptyset \rangle \triangleleft [*?]$). Furthermore, σ_n syllables exhibiting a \$[C][V]\$ phonetic structure are here conventionally represented as \$[C][V][*?], with a fictitious glottal stop [?] in coda position.

The positional function of this fictitious glottal stop, a representational segment which will be deleted at the Phonetic Level ($[*?] \triangleleft |\emptyset|$), turns out to be clear if we consider the following example:

<lybyngk l̥ z̥> (124.1), *ev* {+In} {+KA}, but also <lybyng l̥ z̥> (124.3):

6 But cf. *infra*, § 4.1, in which we will present a three-dimensional model = *mod-b*, within which the two universes are not coplanar.

7 In a personal communication (March 11, 2016), Marcus Kracht made the following remark: “That seems correct. First of all, [they are] not a theorem of MSO for the structures (which is not defined yet). But [they] need not be empirically motivated to be axiomatic. It is axiomatic if you make [them] so”.

8 Such a distinction occurs in a descriptive grammar of a Turkic language as early as in Hahn (1991: 21): “Since morpheme divisions do not necessarily correspond to the prescribed syllabic patterning, morpheme boundaries [...] come to be ignored when syllabification take place [...]”.

C	V	C	C	V	C	C	V	C
<∅>	<ly>		<∅>	<y>	<ng>	<k>	<l>	<∅>

C	V	C	C	V	C	C	V	C
<∅>	<ly>		<∅>	<y>	<ng>	<∅>	<l>	<∅>

0.3.

Regarding the above mentioned *empirically* verifiable conditions, their formalization is by no means strictly necessary for demonstrating the veridicality of statements (1c) and (1d). Nevertheless such a formalization can be implemented, though at the cost of a logical tunneling (not simply a biunivocal linkage between nodes, as in *mod-a*) between the proper plane of representation and the plane of real, concrete objects (graphic signs, phonetic gestures). See below, § 4.1. In the following, we will refer to *mod-a* exclusively.

1.0. A partial inventory of the transcriptional encoding symbols

The following partial inventory of the transcriptional encoding symbols begins with the enumeration of the elements of the dental/alveolar-encoding graphemes set, as detectable within the present text. In order to prevent any confusion with other transliteration systems, it seems appropriate to adopt the following convention: instead of making use of numerical subscripts (for example: t₁, t₂; d₁, d₂ etc.), we will adopt an arbitrary set of alphabetic symbols, taken from the Greek alphabet.

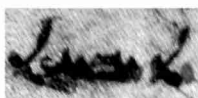
1.1.1.

<d_s> comes ultimately from Syriac <d> (<ⲃ>); <t_θ> comes from Syro-Sogdian <θ>, which in turn derives from Syriac <t> (<ⲥ>); <t_τ> comes ultimately from Syriac <τ> (<Ⲧ>). For further explanation, see *infra*, §§ 1.4.3 and 1.4.4.1.

1.1.2.

At least three allographic variants of the Syriac grapheme <ⲗ> are here merged into the encoding symbol <l>: the isolated form, the final form and the simplified final form. The grapheme <l>, when followed by a grapheme encoding for a [+cons] segment, may encode for one of the following vocalic segments: [a], [a] ([ɐ]) (cf. *infra*, § 2.1.1).

The sequences <ly> and <lw> are intrinsically ambiguous, since, as digraphs, they may encode for single vocalic segments. Cf. , for example, <lylx> *ayaq* (124.3, 4, 7), *versus* <lybynɣlθ> (124.3), *ev* {+In} {+KA}:



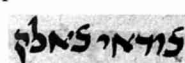
C	V	C	versus	C	V	C
⟨∅⟩	⟨l̥⟩	⟨y⟩		⟨∅⟩	⟨ly̥⟩	⟨b̥⟩

The sequence ⟨lwy⟩ is also intrinsically ambiguous; when interpreted as a trigraph, it encodes for a single vocalic segment, cf. for example ⟨lwytl̥klym̥n̥⟩ (128.2) *ōiā* {-GAy} *m(ā)n*:

C	V	C	C	V	C	C	V	C	C	V	C
⟨∅⟩	⟨lwy⟩	⟨t̥⟩	⟨∅⟩	⟨l̥⟩	⟨k̥⟩	⟨∅⟩	⟨l̥⟩	⟨y̥⟩	⟨m̥⟩	⟨∅⟩	⟨n̥⟩

1.1.3.

⟨x⟩ (𐭪):⁹ This grapheme is foreign to Syriac script, but occurs in Sogdo-Syriac writing¹⁰ (𐭪, ¹¹ 𐭪) ¹² as well as in early Neo-Persian manuscripts written in Syriac script (𐭪).¹³ It is clearly derived from Syriac ⟨k⟩ and, at any rate, has nothing to do with Avestan 𐬨 ⟨x⟩.¹⁴ In Modern Judeo-Persian manuscripts we observe a parallel phenomenon: the grapheme ⟨x⟩ (𐬨) appears to be encoded by means of a “modified” grapheme ⟨k̥⟩ (𐬨) exhibiting a superimposed diacritic mark (a superlinear oblique stroke); compare for example, a specimen from Ms. Vatican Library, Vat. pers. 61:

 خوداي خالق.¹⁵

Furthermore, in the Hebrew section of the aforementioned manuscript, the scribe made use of the same superlinear oblique stroke to distinguish between continuant

9 Zieme (2015: 21): “[...] der Text U [= Yoshida & Chimedddhorji 2008: 407–409, no. 88–91] zeigt, daß auch in türkischen Texten eine dem Sogdischen x nahestehende Form des Buchstabens verwendet wurde”.

10 Skjærvø (1996: 533–535); Yoshida (2009: 283).

11 Ms Berlin-Brandenburgische Akademie der Wissenschaften, no. 146r = T II B 31 = E26.1, cf. Sims-Williams (2012: 80).

12 Ms Berlin-Brandenburgische Akademie der Wissenschaften no. 239r (cf. *supra*).

13 Ms Berlin, Museum für Asiatische Kunst, MIK III 112 = T II B 57, cf. Dickens (2013: 364).

14 Ms Hosseini, Mashhad, 5040, f. 8r.

15 In the same way, again as a random example, in ms Paris, BnF héb. 117, we observe the same superlinear diacritic employed to distinguish between ⟨p̥⟩ and ⟨f̥⟩.

1.1.8.

⟨k⟩: cf. Yoshida (2009: 284): in Sogdo-Syriac script, the grapheme ⟨k⟩ occurs “only in *knθ*, *kθ* ‘city’ and in Syriac words”.

1.1.9.

⟨f⟩:¹⁸ another “derived” grapheme, which occurs only in Sogdo-Syriac lexemes, such as ⟨fryt’ṭ⟩: ⟨xwšnwt_θ fryt_θl_θ,¹⁹ kwyngwllpyrl̄] *kušānūt frītat köñül birlā* ‘with a heart full of joy and love’ (126.2).²⁰ The first lexeme comes from Sogdian ⟨xws’nt⟩, ⟨xwsnt⟩ *xusant*,²¹ ‘happy, joyful’ cf. Gharib (1995: 438b, no. 10708). The Turkic issue of this lexical loan appears to indicate that the actual Sogdian pronunciation was probably [xusənt].²² This may be deduced “from the tendency of the front vowel /a/ to appear as a back vowel, perhaps [o]”. Eventually, this pronunciation evolved into *[xušənit] > [xušənūt] in Turkic mouths. Cf. *infra*, § 3.1.4. Cf. also the following occurrence: ⟨k_θyšy_θnyng_θkwyngly_θnyng_θxwšn_θwt_θ[xy]l̄xynḡ] ²³ *kiši* {+nIṅ} *köñli* {+nIṅ} *kušānūt kīlgak* {+iṅA} (126.11–12, cf. § 2.1.2) ‘for the happiness of a human heart’.

The second lexeme is a phonetic calque from Sogdian ⟨fryt’ṭ⟩ *frīāt*,²⁴ ‘love’, cf. Gharib (1995: 157b, no. 3969, 3977).

Manichean		Sogdo-Syriac		Turko-Syriac	
	⟨k⟩		⟨x⟩	∅	
	⟨k⟩		⟨x⟩		⟨k⟩
	⟨p⟩		⟨f⟩		⟨p⟩
	⟨p⟩		⟨f⟩		⟨p⟩
	⟨k⟩		⟨x⟩		⟨p⟩
	⟨f⟩		⟨p⟩		⟨f⟩

18 Zieme (2015: 21): “Ein Zusatzbuchstabe wird durch einen aufgesetzten runden Haken, der fast wie ein Kreis aussieht, vom Beth gebildet und bezeichnet das spirantische”.

19 Zieme (2015: 155, fol. 3r, ll. 050–051, 160): “xwšnwt pryt’g” *xušnut birtäg*.

20 Zieme (2015: 161): “mit zufriedenem und gleichmütigem (?) Herzen”.

21 Mutō (2008: 244): “ペルシア語 *xošnud* 「adj 満ちたりた」とみなす。(Zieme)”; Zieme (2015: 164): “*xušnut* ‘zufrieden’ < np. *xošnud* ‘id.’”. The syllabic structure of the MP lexeme خوشنود (\$CVC\$+\$CVC\$) is due to a secondary development: cf. Old Persian *uθaduš*, Middle-Persian *hunsand* (γws’nt). Therefore, it cannot be invoked to support the rendering *xušnut* for the Sogdian word.


22 Sims-Williams (1981: 355, 358); Sims-Williams (1989: 181): “The phonetic range of the phoneme *a* is extremely wide, with allophone including *ə* and *ĩ* [...] as well as *o*”.

23 Zieme (2015: 156, note 528): “Vermutlich fehlt nur vom Anfang des Wortes ein Buchstabe. Am ehesten ist nach *l* ein ‘anzunehmen”.

24 Therefore, it is definitely not to be read *qutluy*, as Zieme does in Mutō (2008: 243), nor even *birtäg* (Zieme 2015: 160).

1.2. A morphological survey

1.2.1. Morpheme #+IXG#

- <yṛlyx> *yarliġ*²⁵ (125.8);
 <yṛlyxṛlṣṛ> *yarli(ġ)ka* {-sAr} (128.5); <yṛlyxlmṛmyšž> *yarli(ġ)ka* {-mA}
 {-mIš} (124.9),²⁶ <yṛlyxlywṛ>²⁷ (125.9), <yṛlyxlywṛ> (131.3) *yarli(ġ)ka*
 {-yUr};
 <syṛmlyx>  (127.5).²⁸

Let us examine the syntagmatic context of this occurrence: <snyngkwšwngṛytṛmyšš | pyrṛmṛyšyngṛpyrṛmyngw ṛṛngryk | syṛmlyxṛṛwrwr> *sāniy kūčūy yitmiš* {-čA} *bermiš* {-Iṛ} *bir meṛjū tāṛri* {+kA} *siġim* {+IXG} *turur* ‘your (best) tribute to the unique eternal God, (given) to the utmost of your capacity, is the *highness of heart*’.²⁹ *Siġim* {+IXG} is a denominal syntagm which we hypothesize to be derived from the Sogdian compound-word <sq-m’n>, *sqā-man* ‘high heart’, cf. Gharib (1995: 206b, no. 5186, 353b, no. 8774). In Turkic mouths, the actual pronunciation of this Sogdian syntagm quickly evolved from *[sqaman] ([sqamən]) to *[səcamən] > [səcamən] ≈ [siġimən], this latter issue exhibiting the expected splitting of the onset consonant cluster /sq/. The resulting syntagm would have been easily analyzed as √*siġim* {+In}.

25 Regarding this lexeme, cf. Clauson (1972: 966b(–967a)): “Although morphologically a P.N./A. in *-liġ*, it cannot be so explained etymologically, and this fact, taken with the fact that in Manichaean and Uyg. script it is habitually spelled *yrliġ*, less often *yrliġ*, strongly suggests that it is a very old l(oan)-w(ord)”. In fact, notwithstanding the present-day authoritative ‘responsum’ by Starostin, Dybo & Mudrak (2003: 972), it seems to me that Clauson was not completely wrong in his assertion, since a Proto-Turkic **jAr-* appears to be (possibly) reconstructable *only* in a comparative perspective (i.e. considering PMong. **nari-n* and PTung. **ner-/*nar-*).

26 Mutō (2008: 241): “*yarliqamaqi*”; Zieme (2015: 153, f. 1r, 09): “*y’rlyx’m’qy*” But luckily, the picture provided by Yoshida & Chimeddhorji (2008) is here unambiguous:



27 Mutō (2008: 241): *yarliqayay-m<ü>n*, afterward corrected by Zieme (2013: 101); Zieme (2015: 154, f. 1v, 021).

28 Mutō (2008: 244): *suqimliq*; Zieme (2015: 156, f. 3v, 065): *sy’ymllyx*; 160: *siximlix*; 164: “*sy’ymllyx*. Vermutlich, wenn man *siximlig* liest, eine Ableitung von *sig-* ‘to fit into’ (ED 804b) und vielleicht zu der von M. Erdal behandelten Formation mit *-(X)mlXg* (OTWF 374–376) zu stellen, also mit der Bedeutung ‘das, was dazu paßt’ > ‘das Passende’”.

29 Given his different reading, Mutō (2008: 244), translated this phrase in a different way: “あなたの力の及ぶかぎりの | 献呈は、唯一永遠の神に | 呼びかけることだと。”. Zieme (2015: 162): translated it as ‘das Passende (?)’.

⟨rħymlyx⟩³⁰ (131.7; ⟨rħym⟩ 127.11, 129.10) a denominal syntagm composed of the Syriac loan-word *rhem* ‘mercy, compassion’. Cf. §§ 1.2.4, 2.1.2.

⟨lwtlylyx⟩ *utli* +IXG (126.9) *versus*

⟨lylygγ⟩ *elig*: ⟨lylygγ⟩ (127.9), ⟨lylygγn⟩ (124.1).

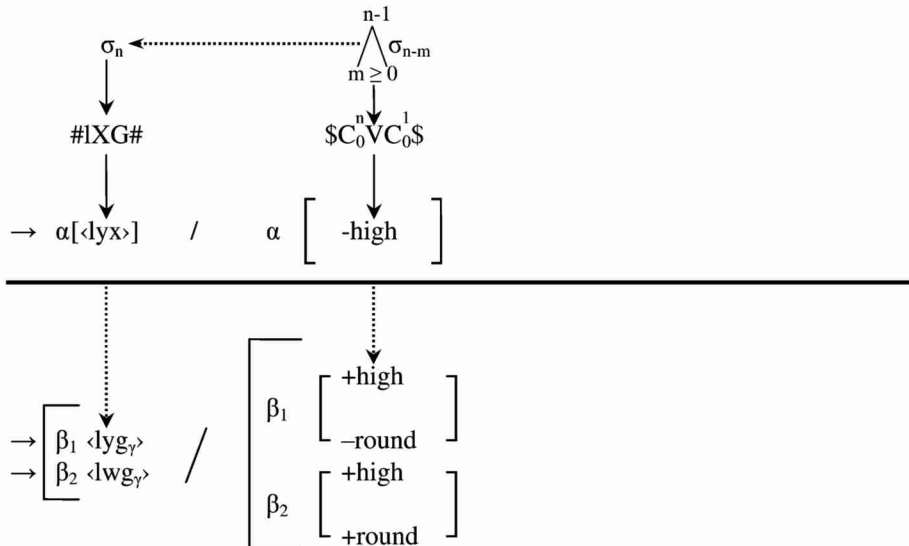
⟨trsqlygγ⟩ (125.6), cf. ⟨trsq⟩ (132.5): a denominal syntagm made up of the Sogdian loan-word ⟨trs’q⟩ *tarsāk* ‘Christian’,³¹ cf. Gharib (1995: 391a, no. 9667). According to some graphotactic evidence (cf. § 2.1.3), this loan-word appears actually to have been pronounced [tersek] by Turkic speakers.

⟨ldgwlwgγ⟩ *ädgüliig* (125.11); this denominal syntagm occurs within the following syntagmatic context: ⟨lwlldgwlwgγxylmys[y]nynglwtllyšy⟩ *ol ädgüliig kilmış* {+In} {+In} *utli* {+sI} ‘reward for his good deed’.³²

⟨kwyngwllwgγ⟩ *könüllüig* (130.1).

1.2.2.

At first glance, we would feel justified in inferring the following, seemingly obvious relationship (2a):



30 Mutō (2008: 248): *r<a>qimliq*, correctly translated as 慈悲を; Zieme (2015: 158, f. 4v, l. 091): *rħym lyx*.

31 Cf. Zieme (2013: 102): “Advice for behaviour during the ceremony follows, addressed to *t(a)rsak-a* ‘Oh Christian’”.

32 Mutō (2008: 248): *ol ädgüliig qilinč-ning* | *utli-si*; Zieme (2013: 101); Zieme (2015: 154), f. 1v, ll. 022–023: *’wl ’dgwllwg xylmysnyng* | *’wtly sy*.

We can try to better understand the above-mentioned constraint schema by putting forward the following descriptive statement: *we assume that* the morpheme #+IXG#, whose vowel nucleus appears to be marked for a certain [αhigh] feature, should agree in this respect with the preceding syllables; or, in terms of modal-logic formalism:

(2b) IDENT-I(NPUT)O(UTPUT) [high] constraint:³³

- Iff $\sigma_{n-m} [n > m \geq 0] \in N_{(\text{input})} \wedge \sigma_n \in N_{(\text{output})} \wedge \langle u, u' \rangle \in V(Nuc)$
- $\text{input} \wedge (u \wedge [\alpha\text{high}]) \rightarrow \boxed{\text{io}} \varphi (u' \wedge [\alpha\text{high}])$ ³⁴

“If a segment σ_{n-m} is an input node and a specific daughter node (its nucleus) is [αhigh], then every output correspondent of σ_{n-m} exhibits an [αhigh] nucleus”. Or: “Any correspondent of an input segment σ_{n-m} whose nucleus is specified as [αhigh], must exhibit an [αhigh] nucleus”.

Just as an example, we may recall here that a similar correlation appears to be arguable by examining, among other things, the text of the following manuscript, written in Uyghur script:

	$\langle \text{lyk} \rangle, \langle \text{lw} \rangle$: [+lik] / [+lük]	l. 11: $\langle \text{rklyk} \rangle$ <i>ärklig</i> ; 13: $\langle \text{twyk} \rangle$ <i>lyk</i> <i>tükällig</i>
British Library, Ms Or. 8218 (122) ³⁵		l. 12: $\langle \text{kwyclwk} \rangle$ <i>küčlüg</i> ; 15: $\langle \text{pwykwylwk} \rangle$ <i>bögülüg</i>
	$\langle \text{lyq} \rangle$: [+liq]	l. 10: $\langle \text{tylyq} \rangle$ <i>r_nynk</i> <i>tinliklar</i> {+nIn}

1.2.3.

In fact, (2a) and (2b) involve a considerable inference from the actually observed data, which instead fit the following, slightly different (and restrained) constraint (3), expressed in Potts & Pullum’s formalism:

- Iff $(\sigma_n | \sigma_n \triangleleft^* \# - \text{IXG} \#) \in N_{(\text{input})} \subseteq N_{(\text{output})} \wedge u \in C(\text{Coda}) \wedge u' \in V(Nuc)$
- $\text{input} \wedge (u \wedge [\alpha\text{high}]) \rightarrow \boxed{\text{io}} \varphi (u' \wedge [\alpha\text{high}])$

33 Adapted from Potts & Pullum (2002: 379), whose original formulation concerned the constraint IDENT-IO [back].

34 Potts & Pullum (2002) implemented an equivalent form of a monadic second-order language tailored to Optimality Theory.

35 Hamilton (1986: I: 27–30).

“If a segment σ_n (which corresponds to the morpheme #+IXG#) is an input node and a specific daughter node (its coda) is [a_{high}], then a specific daughter node (its nucleus) of every correspondent in the output (which, in this specific case, coincides with the same σ_n) exhibits the same [a_{high}] feature”. Roughly speaking: *we assume* that the [a_{high}] feature actually exhibited by the coda consonant of the morpheme #+IXG# (via the graphemic opposition <x> versus <g_r>) is shared by the vowel nucleus of the same morpheme.

Before proceeding further with our analysis, let us consider some syntagmatic contexts in which the morphemes #+mAK# and #+KA# occur:

1.2.4. Morpheme #-mAK#

<swyzlłšmlq> *sözláš* {-mAK} (133.5);

<pyrmlq> *bir* {-mAK} (131.1), <pyrml_qlyg_r> (127.10): *bir* {-mAK} {+IXG}, versus <pyrmlklyg_r> (131.6). Within the aforementioned syllabic context, the graphemic opposition <q> versus <k> is clearly neutralized; on the other hand, <q> and <x> never occur in the same context. We can conclude that in the frame of the present graphemic (and graphotactic) system, <q> appears to behave as a mere allograph of <k>.

<xylmlxlyx> *kilmaklık*:

<lwzwn_ywlxl | _lzw_xpw_syxly_xl_lp | _lpwš_ypyrml_qlyg_rtwrwr
_syxly | lwyzl_rhym_xymlxlyx | t_rwrwr> *uzun yol* {+KA} *azuk bu* †*çığay* {+KA} *lab buşi* [here the Sanskrit lexeme लब्ध *lābha* is immediately followed by its Chinese translation 布施 *bùshī*, which refers to the Buddhist practice of almsgiving] *bir* {-mAK} {+IXG} ‡ [†‡ = 131.6]³⁶ *tur* -Ur *çığay üzâ rāhim kıl* {-mAK} {+IXG} *tur* {-Ur} ‘[the appropriate] provision for the “long trip” is almsgiving to the poor and showing mercy upon the poor’ (127.8–12);

<y_lrm_lx̄> *yar* {-mAK} (129.[5], 6).³⁷

1.2.5 Morpheme #+KA#:³⁸

<pwylñlx> *buyan* (< Sanskrit पुण्य *punya*) {+KA}: <pwylñlx_ls_lxynyp_trsq_llyg_r_lly_tyl_ws_wn_l> *buyan* {+KA} *sakın* {-Ip} *tärsäk* {+IXG} *at* {+I} *üçün* ‘thinking (to do) a good deed for (the reason of) being a Christian [lit.: because of his Christian name]’ (125.5–7).³⁹

<şy_flyx> *çığay* {+KA} (128.1, 131.6).

36 Regarding the syntagm #*lab buşi bir*- # cf. Zieme (1979: 274); Mutō (2008: 245).

37 Cf. Clauson (1972: 969a): “[...] not easily explained semantically either as a Dev. N. in -*ma:k* fr. *yar*- or a Dev. N. in -*k* fr. *yarma*:- [...]”.

38 Erdal (2004: 171–173); Eraslan (2012: 139–145 §§ 346–352).

39 Cf. Zieme (2013: 101): ‘in the name of a Christian who thinks of reward’; Zieme (2015: 154, ll. 017–018).

<lyšwrmšk̄l̄> *içür* {-mIš} {+KA} (128.4), <lyšwrs̄lr̄> *içür* {-sAr} (125.5).
 <pyzk̄l̄> *biz* {+KA} (127.8²).⁴⁰
 <pyrk̄l̄> *bir* {+KA} (128.2, 4).
 <t̄ngryk̄l̄> *tāḡri* {+KA} (127.4) (cf. <t̄ngrȳ>, 127.1).
 <lyβyngk̄l̄> (124.1) *ev* {-Iḡ} {+KA} *versus* <lyβynḡl̄> (124.3).

1.2.6 Morpheme #-GAy#⁴¹

<p̄lrm̄l̄ſ̄l̄ȳ> *bar* {-mA} {-Gay} (128.1).
 <pyrm̄lk̄lymw> *ber* {-mA} {-GAy} mU (128.9).
 <pyrk̄lymn> *ber* {-GAy} {mAn} (128.5).

1.3.

Actually, statement (2b) fails to take into account a number of possible remarks. First of all, if we assume that syllable σ_n (in which the monosyllabic morpheme #+IXG# occurs), for example, is marked as the final target of a rightward spreading harmonic constraint, we would expect always to be able to retrieve a syllable head that triggers the progressive spreading of this specified feature. Roughly speaking: starting from the (entirely) provisional hypothesis, according to which such a case of vowel (or consonant) harmony extends to the last morphemic stratum in accordance with Kaisse's phonological-word model,⁴² we would expect to find a *stem-controlled* harmony. This is, generally speaking, certainly the case; but, definitely not from a "Turing-machine point of view"⁴³, i.e., when considering the *local* heuristic context, in which the sole possible analysis is a step-by-step decoding of a linear string of graphemes, this latter process possibly being triggered only *after having stipulated its directionality*. And that is precisely the point: if we chose a rightward direction, we would happen upon a number of cases in which the decoding process,

40 The syntagm #biz +KA# has already been regarded as a (chiefly?) Yenisean isogloss. But cf. Erdal (2004: 196 note 342, 197). Cf. <pyz ynk̄'> http://vatec2.fkidg1.uni-frankfurt.de/vatecasp/Maitrisimit_0-2.htm: r16(MaitrGeng1196), v06(MaitrGeng1756)] *versus* <pyz k̄'> [r20(MaitrGeng1800)]. At any rate, as far as the Yenisean Runic inscription Tuba II (E 36), line 2, is concerned (<http://bitig.org/?lang=e&mod=1&tid=2&oid=283&m=1>), the reading ṽṽṽṽṽ ṽṽṽṽṽ ṽṽṽṽṽ ṽṽṽṽṽ seems the only possible one according to the sole available drawing (Yıldırım & Aydin & Alimov 2013: 101, <http://img856.imageshack.us/img856/6739/tubaii.jpg>).

41 Erdal (2004: 242–244); Eraslan (2012: 320–321 §§ 584–586).

42 Cf. at least Kaisse (1986).

43 In adopting such a perspective of a Turing type automaton, we wish to stress that we are attempting to understand the focused phenomenon "in terms of a finite number of exact instructions (each instruction being expressed by means of a finite number of symbols" (Copeland 2015). Cf. *infra*, §§ 3.0 and foll.

moving forward along the syntagmatic string, would encounter *no* graphemic segment disambiguating the feature $[\pm \text{high}]$, before it reaches σ_n :

Stem	\rightarrow	σ_n [Kaisse]	stem	σ_{n-1}	σ_n [Kaisse]
$\sqrt{\langle \text{ly} \rangle}$		$\langle \text{ly}_\gamma \rangle$	$\sqrt{\langle \text{sy}\zeta \text{ly} \rangle}$		$\langle \text{x} \rangle$
$\sqrt{\langle \text{wt,ly} \rangle}$		$\langle \text{lyx} \rangle$	$\sqrt{\langle \text{ywl} \rangle}$		$\langle \text{x} \rangle$
$*\sqrt{\langle \text{rhy} \rangle}$		$\langle \text{lyx} \rangle$	$*\sqrt{\langle \text{pwyln} \rangle}$		$\langle \text{x} \rangle$
$*\sqrt{\langle \text{synym} \rangle}$		$\langle \text{lyx} \rangle$	$\sqrt{\langle \text{lyswr} \rangle}$	$\langle \text{m}\dot{\text{s}} \rangle$	$\langle \text{k} \rangle$
$\sqrt{\langle \text{ylr} \rangle}$		$\langle \text{lyx} \rangle$	$\sqrt{\langle \text{pyz} \rangle}$		$\langle \text{k} \rangle$
			$\sqrt{\langle \text{pyr} \rangle}$		$\langle \text{k} \rangle$
			$\sqrt{\langle \text{t,ngry} \rangle}$		$\langle \text{k} \rangle$

Thus, alongside the absence of any surface representation of a *graphemically marked* leftmost head trigger, we may observe that σ_n —the rightmost syllable of the PW to which an alignment constraint ALIGN[ment]-R[ightward]- $(\pm \text{high})$ is applied⁴⁴—turns out to be the only *graphemically marked* one. Obviously, we cannot by any means infer that the nucleus of syllable σ_n (or, as a general case, σ_{n-m}) will turn out to be the head trigger of a *leftward* (regressive) harmony (Spread-L[α F]: Finley 2009: 75; Rose & Walker 2014: 279).

1.3.1.

Now, when implementing a generalization of constraint (3), still (even not entirely), expressed in Potts & Pullum’s formalism, we may state that:

$$\begin{aligned} & \text{Iff } ((\forall \mathfrak{S}_x | \mathfrak{S}_x \in \mathfrak{W}_z) \wedge (\forall \sigma_{n-m} [n > m \geq 0] | \mathfrak{S}_x \triangleleft \sigma_{n-m})) \\ & - \wedge ((\forall \mathfrak{S}_y | \mathfrak{S}_y \in \mathfrak{W}_z) \wedge (\forall \sigma_{n-p} [n > p \geq 0; p \neq m] | \mathfrak{S}_y \triangleleft \sigma_{n-p})) \\ & - \wedge [(\sigma_{n-m} < \sigma_{n-p}) \vee (\sigma_{n-p} < \sigma_{n-m})] \wedge [(\sigma_{n-m} \in N_{(\text{input})}) \wedge (\sigma_{n-p} \in N_{(\text{output})})] \\ & - \wedge [(u'_{\text{input}} \in M(\text{Onset}) \vee (u'_{\text{input}} \in P(\text{Coda}))) \wedge (u''_{\text{output}} \in M(\text{Nuc}))] \\ & \text{input} \wedge (u' \wedge [\alpha \text{high}]) \rightarrow \boxed{\text{io}} \varphi (u'' \wedge [\alpha \text{high}]) \quad (4) \end{aligned}$$

This we may loosely read: “If a syllable σ_{n-m} is an input node and a specific daughter node (its onset or its coda) exhibits a certain $[\alpha \text{high}]$ feature, then each daughter node u' of every correspondent output σ_{n-p} (namely, its syllabic nucleus) exhibits the same $[\alpha \text{high}]$ feature”. Such a statement is to be considered a formal description of a $[\alpha \text{high}]$ vowel harmony, analyzed as a head trigger-free feature spreading.

44 Adapted from Rose & Walker (2014): 264: “This constraint is interpreted as satisfied when the rightmost association of a $[\text{high}]$ feature coincides with the rightmost syllable”.

As a consequence of (4), an (initially) rightward-triggered Turing's type automaton which has to examine a number of finite string \mathfrak{W}_n —where⁴⁵

$$[(\mathfrak{W}_n = \langle \mathfrak{S}_1 \dots \mathfrak{S}_n \rangle \in W \mid W \subset S \supset S_\sigma) \wedge (\forall \mathfrak{S}_{n-m} [n > m \geq 0] \in \mathfrak{W}_x \mid \mathfrak{S}_{n-m} \in S_\sigma) \cdot (\mathfrak{S}_{n-m} = \langle \mathfrak{g}_{n-m}^0, \mathfrak{sg}_{n-m}^0, \mathfrak{ssg}_{n-m}^0 \rangle)]^{46} = \text{Cond}_1$$

which we may read: “for every sequence of graphemic strings $\mathfrak{W}_n = \langle \mathfrak{S}_1 \dots \mathfrak{S}_n \rangle$ belonging to W , the set of all the “Phonological Words” (PW) detectable in our text, which is a subset of S , the set of all the syntagmatically meaningful graphemic strings which may be detectable in our text, and to which S_σ is a subset”—will detect and correctly label any \mathfrak{sg}_{n-m}^0 encoding for a vocalic segment by *bidirectionally scanning* a portion of tape equivalent to a given \mathfrak{W}_n .⁴⁷

1.4.1.

Let us observe the following alternations:

tur {-Ur}: $\langle t_0 wrwr \rangle$ (124.10; 127.5, 10, [12]), *versus* $\langle t_\tau wrwr \rangle$ (128.11, 131.7);

tānri: $\langle t_\tau ngry \rangle$ (126.3) *versus* $\langle t_0 ngry \rangle$ (127.1);

bitig: $\langle pyt_\tau yg_\gamma \mathfrak{z}t_0 \rangle$ *bitig* {+DA}: $\langle nyt_0 g_\gamma \mathfrak{z} kym \mathfrak{z} lryx \mathfrak{z} lwnglywn \mathfrak{z} nwm \mathfrak{z} \mathfrak{z} \rangle$ $\langle pyt_\tau yg_\gamma \mathfrak{z} t_0 \mathfrak{z} mšyh \mathfrak{z} mñ \mathfrak{z} ylrlyx \mathfrak{z} lmlmyš \mathfrak{z} mw \mathfrak{z} t_0 wrwr \rangle$, *netāg kim arīy ewangeljōn nom bitig* {+DA} *mšīhā māran yarli(ġ)ka* {-mA} {-mIš}⁴⁸ *mu turur*, ‘Doesn't our Lord Christ explain in the Holy Book of the Gospel how is it (possible)?’ (124.8–9);⁴⁹ *versus* $\langle pyt_0 y_\gamma g_\gamma \rangle$ (131.3).

We may provisionally argue that the opposition between $\langle t_0 \rangle$ ($\langle lnt_0 lx \rangle$ *andaḡ* [128.10]) and $\langle t_\tau \rangle$ ($\langle lrt_\tau wx \rangle$ *artuḡ* [(127.2)]) does not necessarily involve the feature opposition [-voice] *versus* [+voice], but it certainly encodes for the opposition [+tense] *versus* [-tense]: Erdal (2004: 68): “[...] *tā* is voiceless *or* strong while *dāl* and *ḍāl* are voiced *or* weak [bold *italics* are mine]”. In none of the occurrences examined above is the observed graphemic alternation phonemic.

45 Allow me to be redundantly analytic for the sake of clarity.

46 If not explicitly stated otherwise, here and in the following, the expression $\mathfrak{S}_{n-m} = \langle \mathfrak{g}_{n-m}^0, \mathfrak{sg}_{n-m}^0, \mathfrak{ssg}_{n-m}^0 \rangle$ is intended to be equivalent to $\mathfrak{S}_{n-m} = \langle \mathfrak{g}_{n-m(1)}^0, \mathfrak{sg}_{n-m(2)}^0, \mathfrak{ssg}_{n-m(3)}^0 \rangle$; cf. §§ 0.1, note 4; 0.2.

47 We will describe such a kind of automaton *infra*, on § 3.1.1. Cf., among others, Levelt (1974: 101–106); Hopcroft, Motwani, & Ullmann (2001: 316–329); Kracht (2003: 80–84); Kakde (2007: 132–138); Linz (2011: 224–237).

48 Cf. Erdal (2004: 298) (and *passim*): “In Classical Uyğur [...] *-mAmIš* appears rarely, and only in very late Uyğur sources”.

49 Zieme (2013: 101) casts light on the previous context of this phrase, which may be summarized as follows: “(How is it possible) that you will not accumulate sinful deeds, even if You behaved sinfully, with not feeding him?”.

1.4.2.

Now, what about the grapheme <d_δ>? It seems to occur only in the following contexts:

a: <ld_δ> *äd* (130.4), <ld_δyn> *adin* (128.4), <ld_δg_γwlw_gγ> *ädgülig* (125.11), <lyd_δm_lžs_lržsn> *idma* {-sAr} {-sAn} (124.5); <lyd_δm_ls_lr> *idma* {-sAr} (128.7); <ywd_δwp> *yüü* {-Xp}: here is its syntagmatic context: <swyl y_lzwxny_l ywd_δwp_lkyt_θl_rt_lš_ŷl šy_lly_lx_ll_lp_lpwšy_lpyrm_lkly_gl t_γwrwr> (131.4-5),⁵⁰ *suy* [here the chinese lexeme *zui* 罪 is glossed by its Turkic translation] *yazuk* {+nI} *yüü* {-Xp} *keđâr*⁵¹ {-DAčI} *čigay* {+KA} *lab buši bir* {-mAK} {+IXG} *tur* {-Ur}: which may be translated as ‘Almsgiving to the poor (*čigayka lab buši birmaklig*) will remove the *charging of sin* (*suy yazukni yüüüp*)’.⁵²

b: <pwld_δŷ> *bol* {-DI} (124.6), <lknd_δy> *ikindi* (133.4).

Let us merge the description of these two contexts into a single formula:

(5) Environment of <d_δ>:

Iff $Cond_I^{53} \wedge (\forall \mathbf{g}_{n-m(x)} | \mathbf{g}_{n-m(x)} \triangleleft \langle d_\delta \rangle)^{54}$
input $\rightarrow [((\mathfrak{S}_{n-m} = \mathfrak{S}_1) \wedge (\mathbf{g}_{1(1)} \triangleleft (\langle l \rangle \vee \langle y \rangle) \wedge (\mathbf{g}_{1(2)} \triangleleft (\emptyset \vee \langle y \rangle \vee \langle w \rangle) \wedge (\mathbf{g}_{1(3)} \triangleleft \langle d_\delta \rangle)) \vee (\mathbf{g}_{n-m(1)} \triangleleft \langle d_\delta \rangle) \wedge (\mathbf{g}_{n-m(3)} \triangleleft *C[+sonorant, +coronal, +anterior]))] = Env_I^{55}$

which may be read:

“given $Cond_I$ (i.e., for each graphemic string $\mathfrak{S}_{n-m} = \langle \mathbf{g}_{n-m(1)}, \mathbf{g}_{n-m(2)}, \mathbf{g}_{n-m(3)} \rangle$ which encodes for a CVC syllable σ_{n-m} ;

and for every occurrence of grapheme <d_δ> as surface representation of the underlying object $\mathbf{g}_{n-m(x)}$;

→ the grapheme <d_δ> will occur *exclusively* in syllable σ_1 , this latter being *exclusively* represented either by the strings <lØd_δ>, <lyd_δ> or <ywd_δ>—in short, any first syllable exhibiting [ð] as coda consonant should exclusively belong to the type [ʔVð], [jVð];

→ or discursively expressed: for every occurrence of grapheme <d_δ> as onset of the syllable which immediately follows σ_{n-m} , the underlying object $\mathbf{g}_{n-(m-1)(3)}$ dominates a [+sonorant, +coronal, +anterior] consonant, i.e. /l/ or /n/”.

50 Zieme (2015: 158: fol. 4recto, l. 089).

51 Starostin, Dybo & Mudrak (2003: 534).

52 Here is Zieme’s translation (Zieme [2015]: 162): ‘Sünden aufladen, und so sie entfernen kommt vom Geben von Almosen an Arme’.

53 See § 1.3.1.

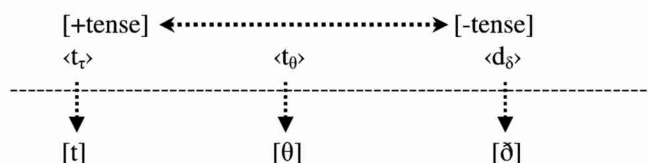
54 Here and in the following, for the sake of conciseness, every proposition of the form “ $\forall \mathbf{g}_{n-m(x)} | \mathbf{g}_{n-m(x)} \triangleleft \in N[\text{input}]$ ” is here held always to be implied by the simple statement “ $\forall \mathbf{g}_{n-m(x)}$ ”.

55 The logical structure of this formula is: if (*a* and (*b* or *c*)) then (*d* or *e*) \doteq if (*a* and *b* then *d*) or (if *a* and *c* then *e*).

Besides the aforementioned contexts, the grapheme <ḏ> occurs only in some Sogdian loanwords, such as <ḏyndḏlr> *dindar* (133.1).⁵⁶ This lexeme is calqued directly from Sogdian <δynd'ṛ> (cf. Gharib [1995]: 149b). One thing that stands out is that not only does the Aramaic grapheme <ḏ> never occur in the so-called Sogdian script “of the Ancient Letters” (Sims-Williams [1975: 136, note 23]), but also “the Sogdians do not seem to have regarded the distinction between [δ/ and θ/ as a matter of vital importance [...] nor did the Manichaeans find it necessary to differentiate /δ/ and /θ/ in their otherwise precise and unambiguous script” (Sims-Williams [1981: 349]). At any rate, in Syro-Sogdian the grapheme <ḏ> would have represented the voiced apico-dental continuant [ḏ].

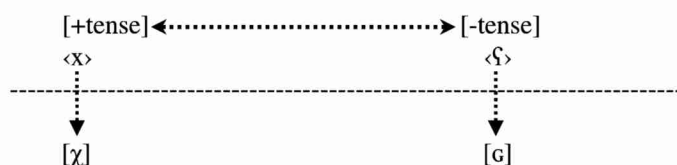
1.4.3.1.

To sum up: the Turco-Syriac graphemes <ṭ>, <ṭ̰> and <ḏ>—as they occur within the linguistic domain under scrutiny—are in fact the encodings at the surface level representation, of three allophones, whose tense articulation hierarchy may be represented as follows:



1.4.3.2.

Let us recall here paragraph 1.1.5, where we outlined in discursive form the syntagmatic environment of the grapheme <ḥ>. As already stated, the Turco-Syriac graphemes <x> and <ḥ> appear to be the encodings, at the surface level representation, of two allophones, whose tense articulation hierarchy may be represented as follows:



1.5.

Let us compare the two formulas which describe the environments of <ḏ> and <ḥ> (see §§ 1.1.5 and 1.4.2):

56 Here I must rely upon Zieme’s transcription, as I cannot draw any reliable information from the photograph provided by Yoshida & Chimeddhorji (2008: 363).

(6) Environment of $\langle d_\delta \rangle$:

Iff $Cond_1 \wedge (\forall \mathbf{g}_{n-m(x)} \mid \mathbf{g}_{n-m(x)} \triangleleft \langle d_\delta \rangle)$

input $\rightarrow [((\mathfrak{S}_{n-m} = \mathfrak{S}_1) \wedge (\mathbf{g}_{1(1)} \triangleleft \langle \cdot \rangle \vee \langle y \rangle) \wedge (\mathbf{g}_{1(2)} \triangleleft (\emptyset \vee \langle y \rangle \vee \langle w \rangle) \wedge (\mathbf{g}_{1(3)} \triangleleft \langle d_\delta \rangle)) \vee (\mathbf{g}_{n-m(1)} \triangleleft \langle d_\delta \rangle) \wedge (\mathbf{g}_{n-m(1)(3)} \triangleleft^* [C][+son, +coronal, +ant])] = Env_1$

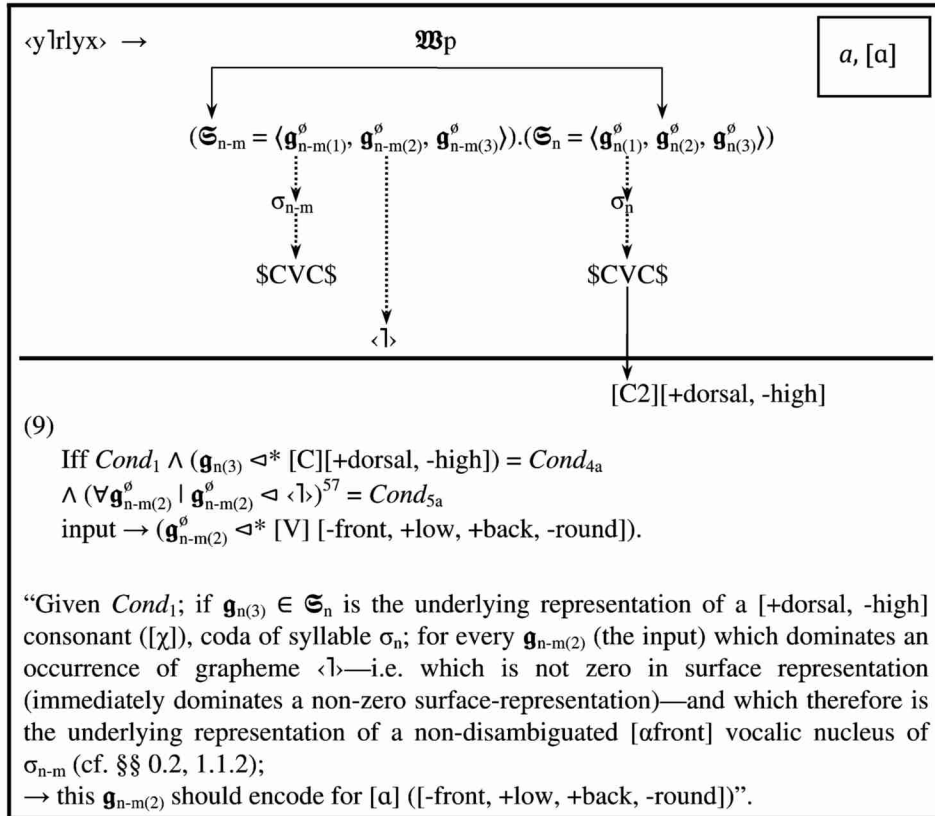
(7) Environment of $\langle \zeta \rangle$:

Iff $Cond_1 \wedge (\forall \mathbf{g}_{n-m(x)} \mid \mathbf{g}_{n-m(x)} \triangleleft \langle \zeta \rangle)$

input $\rightarrow [(\mathbf{g}_{n-m(3)} \triangleleft \langle \zeta \rangle) \wedge (\mathbf{g}_{n-m(1)(1)} \triangleleft \langle \cdot \rangle) \vee (\mathbf{g}_{n-m(1)} \triangleleft \langle \zeta \rangle) \wedge (\mathbf{g}_{n-m(1)(3)} \triangleleft^* [C][+son, +coronal, +ant])] = Env_2$

2.0. A partial vowels inventory

2.1.1.



2.1.2.

$\langle \text{rhy}^{\text{m}}\text{lyx} \rangle \text{ r}^{\text{h}}\text{i}^{\text{m}}\text{l}^{\text{i}}\text{g} \rightarrow \mathfrak{B}_p$

$\acute{a}, [\text{v}]$

$$(\mathfrak{S}_{n-m} = \langle \mathfrak{g}_{n-m(1)}^{\circ}, \mathfrak{g}_{n-m(2)}^{\circ}, \mathfrak{g}_{n-m(3)}^{\circ} \rangle) \cdot (\mathfrak{S}_{n-1}) \cdot (\mathfrak{S}_n = \langle \mathfrak{g}_{n(1)}^{\circ}, \mathfrak{g}_{n(2)}^{\circ}, \mathfrak{g}_{n(3)}^{\circ} \rangle)$$

(10)

Iff $\text{Cond}_1 \wedge \text{Cond}_{4a} \wedge (\forall \mathfrak{g}_{n-m(2)}^{\circ} \mid \mathfrak{g}_{n-m(2)}^{\circ} \triangleleft \langle \emptyset \rangle) = \text{Cond}_{5b}$
input $\rightarrow (\mathfrak{g}_{n-m(2)}^{\circ} \triangleleft^* [\text{V}] [-\text{front}, +\text{low}, +\text{back}, +\text{round}])$

Given Cond_1 and Cond_{4a} ; for every $\mathfrak{g}_{n-m(2)}$ (the input) which is zero in surface representation (which immediately dominates a zero surface-representation);
 \rightarrow this $\mathfrak{g}_{n-m(2)}$ should encode for [v] ([-front, +low, +back, +round]).
Cf. $\langle [\text{xy}] \text{I} \text{S} \emptyset \text{x} \rangle [\text{x} \text{i} \text{l} \text{g} \text{v} \text{x}]$ (126.12), $\langle \text{t}_0 \emptyset \text{b} \text{l} \text{r} \rangle \text{i} \acute{a} \text{v}$ {-Ar} (130.5); $\langle \text{lydml} \text{I} \text{s} \text{l} \text{r} \text{z} \text{s} \emptyset \text{n} \rangle \text{i} \acute{d} \text{m} \text{a}$ {-sAr} $\text{s} \acute{a} \text{n}$ (124.5).

2.1.3.

$\langle \text{swyzl} \text{I} \text{s} \text{m} \text{l} \text{q} \rangle \text{s} \acute{o} \text{z} \text{l} \acute{a} \text{s} \{-\text{mAK}\} \rightarrow \mathfrak{B}_p$

$\acute{a} [\text{a}], [\text{v}]$

$$(\mathfrak{S}_1) \cdot (\mathfrak{S}_{n-m} = \langle \mathfrak{g}_{n-m(1)}^{\circ}, \mathfrak{g}_{n-m(2)}^{\circ}, \mathfrak{g}_{n-m(3)}^{\circ} \rangle) \cdot (\mathfrak{S}_n = \langle \mathfrak{g}_{n(1)}^{\circ}, \mathfrak{g}_{n(2)}^{\circ}, \mathfrak{g}_{n(3)}^{\circ} \rangle)$$

(11)

Iff $\text{Cond}_1 \wedge (\mathfrak{g}_{n(3)}^{\circ} \triangleleft^* [\text{C}] [+ \text{dorsal}, + \text{high}, + \text{tense}]) = \text{Cond}_{4b} \wedge \text{Cond}_{5a}$
input $\rightarrow (\mathfrak{g}_{n-m(2)}^{\circ} \triangleleft^* [\text{V}] [-\text{front}, +\text{low}, -\text{back}])$

“Given Cond_1 ; if $\mathfrak{g}_{n(3)}$ is the underlying representation of a [+dorsal, +high, +tense] consonant ([k]), coda of syllable σ_n ; and given Cond_{5a} ;
 \rightarrow then $\mathfrak{g}_{n-m(2)}^{\circ}$ should encode for [a] ([-front, +low, -back])”.

2.1.4.

$\langle \text{t}_{\text{r}} \text{s} \text{q}_{\text{L}} \text{lyg}_{\text{r}} \rangle \text{t} \acute{a} \text{r} \text{s} \acute{a} \text{k} \text{l} \text{i} \text{g} \rightarrow \mathfrak{B}_p$

$\acute{a}, [\text{E}]$

$$(\mathfrak{S}_{n-m} = \langle \mathfrak{g}_{n-m(1)}^{\circ}, \mathfrak{g}_{n-m(2)}^{\circ}, \mathfrak{g}_{n-m(3)}^{\circ} \rangle) \cdot (\mathfrak{S}_{n-1}) \cdot (\mathfrak{S}_n = \langle \mathfrak{g}_{n(1)}^{\circ}, \mathfrak{g}_{n(2)}^{\circ}, \mathfrak{g}_{n(3)}^{\circ} \rangle)$$

(12)

Iff $\text{Cond}_1 \wedge (\mathfrak{g}_{n(3)}^{\circ} \triangleleft^* [\text{C}] [+ \text{dorsal}, + \text{high}, - \text{tense}]) = \text{Cond}_{4b} \wedge \text{Cond}_{5b}$
input $\rightarrow (\mathfrak{g}_{n-m(2)}^{\circ} \triangleleft^* [\text{V}] [+ \text{front}, - \text{low}, - \text{back}, - \text{tense}])$

“Given Cond_1 ; if $\mathfrak{g}_{n(3)}$ is the underlying representation of a [+dorsal, +high, -tense] consonant ([y]), coda of syllable σ_n ; and given Cond_{5b} ;
 \rightarrow then $\mathfrak{g}_{n-m(2)}^{\circ}$ should encode for [E] ([+front, -low, -back, -tense])”.

2.1.5.

$\langle \text{lylyg}_\gamma \rangle \text{elig} \rightarrow \mathfrak{B}_p$

$(\mathfrak{C}_{n-m} = (\mathfrak{g}_{n-m(1)}^\theta, \mathfrak{g}_{n-m(2)}^\theta, \mathfrak{g}_{n-m(3)}^\theta)) \cdot (\mathfrak{C}_n = (\mathfrak{g}_{n(1)}^\theta, \mathfrak{sg}_{n(2)}^\theta, \mathfrak{ssg}_{n(3)}^\theta))$

(13)

Iff $Cond_1 \wedge Cond_{4b} \wedge (\mathfrak{g}_{n-m(2)}^\theta \triangleleft (\langle y \rangle \vee \langle \text{ly} \rangle) = Cond_{5c}$

input $\rightarrow (\mathfrak{g}_{n-m(2)}^\theta \triangleleft^* [V] [+front, -low, -back, +tense])$

“Given $Cond_1$ and $Cond_{4b}$; and given $Cond_{5c}$;
 \rightarrow then $\mathfrak{g}_{n-m(2)}^\theta$ should encode for [e] ([+front, -low, -back, +tense])”

Cf. $\langle \text{lyky} \check{\text{y}} \rangle \text{eki}$ (128.2, 8), but also $\langle \text{nyg}_\gamma \text{w} \rangle$ (128.12) *negü*, $\langle \text{nyt}_\theta \text{yg}_\gamma \rangle$ (124.8) *ne-täg*.

The first four aforementioned relationships may be merged into a single statement (14):

$$\begin{aligned} &\text{Iff } Cond_1 \wedge ((Cond_{4a} \vee Cond_{4b}) \wedge (Cond_{5a} \vee Cond_{5b})) \\ &\text{input } \rightarrow (\mathfrak{g}_{n-m(2)}^\theta \triangleleft^* [V] [-round]) \vee (\mathfrak{g}_{n-m(2)}^\theta \triangleleft^* [V] [+round]) \\ &\vee (\mathfrak{g}_{n-m(2)}^\theta \triangleleft^* [V] [-front, +low]) \vee (\mathfrak{g}_{n-m(2)}^\theta \triangleleft^* [V] [+front, -low]). \end{aligned}$$

If we consider $Cond_1$ as an obviously implied condition, and if we replace $Cond_{4a}$ and $Cond_{4b}$ with the following, more general conditions:

$$\begin{aligned} &(\exists \mathfrak{g}_{n-m(2\pm 1)}^\theta \mid \mathfrak{g}_{n-m(2\pm 1)}^\theta \triangleleft^* [C] [+dorsal, -high]) = Cond_{4c} \\ &(\exists \mathfrak{g}_{n-m(2\pm 1)}^\theta \mid \mathfrak{g}_{n-m(2\pm 1)}^\theta \triangleleft^* [C] [+dorsal, +high]) = Cond_{4d} \end{aligned}$$

we obtain (15):

$$\begin{aligned} &\text{Iff } (Cond_{4c} \vee Cond_{4d}) \wedge (Cond_{5a} \vee Cond_{5b}) \\ &\rightarrow (\mathfrak{g}_{n-m(2)}^\theta \triangleleft^* [V] [-round]) \vee (\mathfrak{g}_{n-m(2)}^\theta \triangleleft^* [V] [+round]) \\ &\vee (\mathfrak{g}_{n-m(2)}^\theta \triangleleft^* [V] [-front, +low]) \vee (\mathfrak{g}_{n-m(2)}^\theta \triangleleft^* [V] [+front, -low]). \end{aligned}$$

3.0. The Old Turkic Reader

Heinz & Lai (2013: 55–61) strongly advocate the theoretic assessment according to which a phonological representation (for example a possible phonological representation of vowel harmony) may be implemented in terms of p-subsequential relations, i.e. algorithms implemented in a quasi-deterministic input transducer—i.e., very roughly speaking, in terms of a (quasi-)deterministic Turing-Machine calculus.

Now, if we define the following alphabet:

$$\begin{aligned}
\Gamma &= \Gamma_{\text{input}} \cup \Gamma_{\text{output}} = \{ \gamma_1 = \mathbf{g}_{n-m(2\pm 1)}^\sigma \in C; \\
\gamma_a &= \mathbf{g}_{n-m(2)}^\sigma \triangleleft^* [V_x] \mid (([V_x] \in V_{-\emptyset}) \wedge (V_{-\emptyset} \not\exists \langle \emptyset \rangle)); \\
\gamma_z &= \mathbf{g}_{n-m(2)}^\sigma \triangleleft \langle \uparrow \rangle; \\
\gamma_b &= \mathbf{g}_{n-m(2)}^\sigma \triangleleft^* [V_x] \mid [V_x] \in V_{[-\text{front}, +\text{low}, -\text{back}]}; \\
\gamma_c &= \mathbf{g}_{n-m(2)}^\sigma \triangleleft^* [V_x] \mid [V_x] \in V_{[+\text{front}, -\text{low}, -\text{back}, -\text{tense}]}; \\
\gamma_d &= \mathbf{g}_{n-m(2)}^\sigma \triangleleft^* [V_x] \mid [V_x] \in V_{[-\text{front}, +\text{low}, +\text{back}, -\text{round}]}; \\
\gamma_e &= \mathbf{g}_{n-m(2)}^\sigma \triangleleft^* [V_x] \mid [V_x] \in V_{[-\text{front}, +\text{low}, +\text{back}, +\text{round}]}; \\
\gamma_f &= (\mathbf{g}_{n-m(2)}^\sigma \triangleleft \langle \emptyset \rangle) \wedge (\mathbf{g}_{n-m(2)}^\sigma \triangleleft^* [V_x]); \\
\gamma_2 &= \mathbf{g}_{n-m(2\pm 1)}^\sigma \triangleleft^* [C_x] \mid [C_x] \in C_{[+\text{dorsal}, +\text{high}]}; \\
\gamma_3 &= \mathbf{g}_{n-m(2\pm 1)}^\sigma \triangleleft^* [C_x] \mid [C_x] \in C_{[+\text{dorsal}, +\text{high}]}; \\
\gamma_4 &= (\mathbf{g}_{n-m(2\pm 1)}^\sigma \triangleleft \langle \emptyset \rangle) \wedge (\mathbf{g}_{n-m(2\pm 1)}^\sigma \triangleleft^* [C_x]); \\
\sigma_{1(p=2\pm 1)} &; \\
\sigma_{n-m(p=2\pm 1)} &; \\
\sigma_{n(p=2\pm 1)} & \}
\end{aligned}$$

where C is (the set of all the \mathbf{g} -objects which immediately dominate) [= $Statement_{\mathbf{g}_1}$] graphemes encoding for [+cons] segments; $C_{[+\text{dorsal}, +\text{high}]} \subset C$ is [$Statement_{\mathbf{g}_1}$] graphemes encoding for [+cons, +dorsal] segments specified for the [+high] feature; $C_{[+\text{dorsal} -\text{high}]} \subset C$ is [$Statement_{\mathbf{g}_1}$] graphemes encoding for [+cons, +dorsal] segments specified for the [-high] feature; V is [$Statement_{\mathbf{g}_1}$] graphemes encoding for [+vocalic] segments; $V_{-\emptyset} \subset V$ is [$Statement_{\mathbf{g}_1}$] *non-zero* graphemes encoding for [+vocalic] segments; $V_{[-\text{front}, +\text{low}]} \subset V$ is [the set of all the \mathbf{g} -objects which dominate = $Statement_{\mathbf{g}_2}$] [+vocalic] segments specified for the [-front, +low] features; $V_{[+\text{front}, -\text{low}]} \subset V$ is [$Statement_{\mathbf{g}_2}$] [+vocalic] segments specified for the [+front, -low] features; $V_{[-\text{round}]} \subset V$ is [$Statement_{\mathbf{g}_2}$] [+vocalic] segments specified for the [-round] feature; $V_{[+\text{round}]} \subset V$ is [$Statement_{\mathbf{g}_2}$] [+vocalic] segments specified for the [+round] feature; $\sigma_{1(p=2\pm 1)}$, $\sigma_{n-m(p=2\pm 1)}$, $\sigma_{n(p=2\pm 1)}$ are sequencing labels which are automatically attributed to each cell: $\sigma_{1(1)}$ and $\sigma_{n(3)}$ mark respectively the beginning (the head of the nonblank portion of the tape, the leftmost nonblank cell) and the end of the syntagmatic string respectively;

and if we put:

$$\begin{aligned}
((\forall \gamma_z) (\gamma_1 \prec \gamma_z) \vee (\gamma_z \prec \gamma_1)) &= A \\
((\forall \gamma_f) (\gamma_1 \prec \gamma_f) \vee (\gamma_f \prec \gamma_1)) &= B
\end{aligned}$$

we can first translate statement (15) into the following formula (16):

$$\begin{aligned}
(\forall \gamma_1 \mid \gamma_1 = \gamma_2) \cdot [[A \rightarrow (\gamma_z = \gamma_d)] \vee [B \rightarrow (\gamma_f = \gamma_e)]] \\
\wedge (\forall \gamma_1 \mid \gamma_1 = \gamma_3) \cdot [[A \rightarrow (\gamma_z = \gamma_b)] \vee [B \rightarrow (\gamma_f = \gamma_c)]]
\end{aligned}$$

Then, we can implement a Deterministic Finite State Transducer $OTR = (Q, \Sigma, \Gamma, \delta, \#, q_0, q_f)$ defined as follows:

$Q = \{q_0, q_1, q_2, q_3, q_4, q_5, q_6, q_7, q_8, q_9, q_{10}, q_f\}$, $\Sigma = \{\gamma_b, \gamma_c, \gamma_d, \gamma_e\}$, $\Gamma = \{\gamma_1, \gamma_a, \gamma_z, \gamma_b, \gamma_c, \gamma_d, \gamma_e, \gamma_f, \gamma_2, \gamma_3, \gamma_4, \sigma_{1(p)}, \sigma_{n-m(p)}, \sigma_{n(p)}\}$.

The tape symbols γ_f and γ_4 occur in a given cell when, respectively, a [+vocalic] or a [+cons] segment is encoded by $\langle \emptyset \rangle$ in surface representation.

3.1.1.

First example: let us consider the string $\langle \text{pyrklymn} \rangle \text{ber} \{-\text{GAy}\} \{\text{mAn}\}$ (128.5). As a first step, we ‘interpret’ it according to our CVC syllabification model (statement 1):

$\langle p \rangle$	$\langle y \rangle$	$\langle r \rangle$	$\langle k \rangle$	$\langle l \rangle$	$\langle y \rangle$	$\langle m \rangle$	$\langle n \rangle$
$\langle X_1 \rangle$	$\langle X_2 \rangle$	$\langle X_3 \rangle$	$\langle X_4 \rangle$	$\langle X_5 \rangle$	$\langle X_6 \rangle$	$\langle X_7 \rangle$	$\langle X_8 \rangle$
$\mathfrak{g}_{1(1)}$	$\mathfrak{g}_{1(2)}$	$\mathfrak{g}_{1(3)}$	$\mathfrak{g}_{2(1)}$	$\mathfrak{g}_{2(2)}$	$\mathfrak{g}_{2(3)}$	$\mathfrak{g}_{3(1)}$?

Now, since $(s \langle X_7 \rangle \triangleleft [C]) \rightarrow (\mathfrak{g}_{n-m(2)} \triangleleft \langle \emptyset \rangle)$, it is the case that:

$\mathfrak{g}_{1(1)}$	$\mathfrak{g}_{1(2)}$	$\mathfrak{g}_{1(3)}$	$\mathfrak{g}_{2(1)}$	$\mathfrak{g}_{2(2)}$	$\mathfrak{g}_{2(3)}$	$\mathfrak{g}_{3(1)}$	$\mathfrak{g}_{3(2)}$	$\mathfrak{g}_{3(3)}$
↓	↓	↓	↓	↓	↓	↓	$\langle \emptyset \rangle$	↓
$[C_{1(1)}]$	$[V_{1(2)}]$	$[C_{1(3)}]$	$[C_{2(1)}]$	$[V_{2(2)}]$	$[C_{2(3)}]$	$[C_{3(1)}]$	$[V_{3(2)}]$	$[C_{3(3)}]$

We then transform it into the following TM finite tape:

#	$\gamma_1 \sigma_{1(1)}$	$\gamma_a \sigma_{1(2)}$	$\gamma_1 \sigma_{1(3)}$	$\gamma_3 \sigma_{n-m(1)}$	$\gamma_z \sigma_{n-m(2)}$	$\gamma_1 \sigma_{n-m(3)}$	$\gamma_f \sigma_{n(1)}$	$\gamma_f \sigma_{n(2)}$	$\gamma_1 \sigma_{n(3)}$	#
---	--------------------------	--------------------------	--------------------------	----------------------------	----------------------------	----------------------------	--------------------------	--------------------------	--------------------------	---

Let OTR starts:

01. $\delta(q_0, \gamma_1 \sigma_{1(1)}) = (q_1, \emptyset, +1)$
02. $\delta(q_1, \gamma_a \sigma_{1(2)}) = (q_1, \emptyset, +1)$
03. $\delta(q_1, \gamma_1 \sigma_{1(3)}) = (q_1, \emptyset, +1)$
04. $\delta(q_1, \gamma_3 \sigma_{n-m(1)}) = (q_2, \emptyset, +1)$
05. $\delta(q_2, \gamma_z \sigma_{n-m(2)}) = (q_3, \gamma_b \sigma_{n-m(2)}, +1)$
06. $\delta(q_3, \gamma_1 \sigma_{n-m(3)}) = (q_2, \emptyset, +1)$
07. $\delta(q_4, \gamma_1 \sigma_{n(1)}) = (q_2, \emptyset, +1)$
08. $\delta(q_2, \gamma_f \sigma_{n(2)}) = (q_4, \gamma_c \sigma_{n(2)}, +1)$
09. $\delta(q_4, \gamma_1 \sigma_{n(3)}) = (q_5, \emptyset, -1)$
10. $\delta(q_5, \gamma_c \sigma_{n(2)}) = (q_5, \emptyset, -1)$
11. $\delta(q_5, \gamma_1 \sigma_{n(1)}) = (q_5, \emptyset, -1)$
12. $\delta(q_5, \gamma_1 \sigma_{n-m(3)}) = (q_5, \emptyset, -1)$
13. $\delta(q_5, \gamma_b \sigma_{n-m(2)}) = (q_5, \emptyset, -1)$
14. $\delta(q_5, \gamma_3 \sigma_{n-m(1)}) = (q_5, \emptyset, -1)$
15. $\delta(q_5, \gamma_1 \sigma_{1(3)}) = (q_5, \emptyset, -1)$
16. $\delta(q_5, \gamma_a \sigma_{1(2)}) = (q_5, \emptyset, -1)$
17. $\delta(q_5, \gamma_1 \sigma_{1(1)}) = (q_f, \emptyset, 0)$

Symbols					
States	$\gamma_1 \sigma_{1(1)}$	$\gamma_a \sigma_{1(2)}$	$\gamma_1 \sigma_{1(3)}$	$\gamma_3 \sigma_{n-m(1)}$	$\gamma_a \sigma_{n-m(2)}$
q_0					
q_1	$(q_1, \emptyset, +1)$	$(q_1, \emptyset, +1)$	$(q_1, \emptyset, +1)$		
q_2				$(q_2, \emptyset, +1)$	
q_3					$(q_3, \gamma_b \sigma_{n-m(2)}, +1)$

q ₄					
q ₅		(q ₅ , Ø, -1)	(q ₅ , Ø, -1)	(q ₅ , Ø, -1)	(q ₅ , Ø, -1)
q _f	(q _f , Ø, 0)				

Symbols				
States	$\gamma_1\sigma_{n-m(3)}$	$\gamma_1\sigma_{n(1)}$	$\gamma_1\sigma_{n(2)}$	$\gamma_1\sigma_{n(3)}$
q ₀				
q ₁				
q ₂	(q ₂ , Ø, +1)	(q ₂ , Ø, +1)		
q ₃				
q ₄			(q ₄ , $\gamma_c\sigma_{n(2)}$, +1)	
q ₅	(q ₅ , Ø, -1)	(q ₅ , Ø, -1)	(q ₅ , Ø, -1)	(q ₅ , Ø, -1)
q _f				

The instructions 5 ($\delta(q_2, \gamma_2\sigma_{n-m(2)}) = (q_3, \gamma_b\sigma_{n-m(2)}, +1)$) and 8 ($\delta(q_2, \gamma_f\sigma_{n(2)}) = (q_4, \gamma_c\sigma_{n(2)}, +1)$) are implementations of, respectively, licensing constraint (11) and (12), and therefore the output is : *b(x)rkáymän* [berkəymən].

3.1.2.

Another example is: $\langle t_r l x l t_r y n g t_0 \rangle$ (127.2). The lexeme $\langle t_r l x l t_r \rangle$ being a loan-word based on Syriac *taht* ‘under, beneath’ (which eventually became [taxat] in Turkic mouths), should possibly not be translated as ‘worship’:⁵⁸ $\langle pyr_myngw_t_0 ngr\ddot{y} _snyn\grave{g}l \ t_r l x l t_r y n g t_0 _l r t_r w x _l \ n l r s f r l t_r y l l m l \ddot{s} \rangle$ *bir meñü täğri sän* {+Iñ} *takat* {+Iñ} {+DA} *artuḡ nā* (á)r {-sAr} *tilā* {-mAz} (127.1–3) ‘An eternal God does not require anything more than your *submission*’.

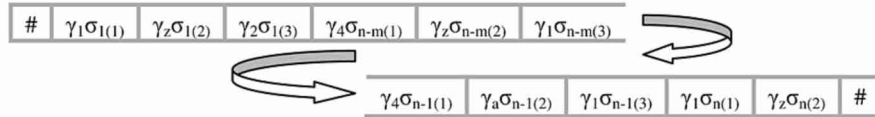
$\langle t_r \rangle$	$\langle l \rangle$	$\langle x \rangle$	$\langle l \rangle$	$\langle t_r \rangle$	$\langle y \rangle$	$\langle n g \rangle$	$\langle t_r \rangle$	$\langle l \rangle$
$\langle x_1 \rangle$	$\langle x_2 \rangle$	$\langle x_3 \rangle$	$\langle x_4 \rangle$	$\langle x_5 \rangle$	$\langle x_6 \rangle$	$\langle x_7 \rangle$	$\langle x_8 \rangle$	$\langle x_9 \rangle$

Now, since $(s\langle x_3 \rangle \triangleleft [V]) \rightarrow (g_{2(1)} \triangleleft \langle \emptyset \rangle)$, but also $(s\langle x_5 \rangle \triangleleft [V]) \rightarrow (g_{2(1)} \triangleleft \langle \emptyset \rangle)$ (see statement (1)), so it is the case that:

58 As did Mutō (2008: 244): “崇拜を”. But cf. Zieme (2015: 164): “[...] *taxat* ‘Gehorsam’, vgl. osm. *taat* (طاعت) ‘act of obedience to God; act of piety’ (Redhouse 2011, 1072a) (< ar. $\sqrt{tā}$)”. Though perfectly conceivable, and semantically convincing—although *taxat* would have appeared in conjunction with *tapay*, cf., at any rate, Zieme (2015: 160, line 104): *qurbana tapinur*—Zieme’s proposal does not seem completely unchallengeable from a phonetic point of view, since it relies on the implication of the following, hypothetical process: [$*t_0\ddot{s}æt$] > [$*tagat$] > [taxat]. On the other hand, a hypothetical lexeme **tayāt* could well have been written $\langle t_r l x l t_r \rangle$, but hardly $\langle t_r l x l t_r \rangle$ (cf. §§ 1.1.5 and 1.4.3.2).

$g_{1(1)}$	$g_{1(2)}$	$g_{1(3)}$	$g_{2(1)}$	$g_{2(2)}$	$g_{2(3)}$	$g_{3(1)}$	$g_{3(2)}$	$g_{3(3)}$	$g_{4(1)}$	$g_{4(2)}$
↓	↓	↓	$\langle \emptyset \rangle$	↓	↓	$\langle \emptyset \rangle$	↓	↓	↓	↓
$[C_{1(1)}]$	$[V_{1(2)}]$	$[C_{1(3)}]$	$[C_{2(1)}]$	$[V_{2(2)}]$	$[C_{2(3)}]$	$[C_{3(1)}]$	$[V_{3(2)}]$	$[C_{3(3)}]$	$[C_{4(1)}]$	$[V_{4(2)}]$

The resulting TM finite tape is:



Again, OTR may start:

1. $\delta(q_0, \gamma_1\sigma_{1(1)}) = (q_1, \emptyset, +1)$
2. $\delta(q_1, \gamma_2\sigma_{1(2)}) = (q_1, \emptyset, +1)$
3. $\delta(q_1, \gamma_2\sigma_{1(3)}) = (q_6, \emptyset, +1)$
4. $\delta(q_6, \gamma_4\sigma_{n-m(1)}) = (q_7, |\emptyset|, +1)$
5. $\delta(q_7, \gamma_z\sigma_{n-m(2)}) = (q_8, \gamma_d\sigma_{n-m(2)}, +1)$
6. $\delta(q_8, \gamma_1\sigma_{n-m(3)}) = (q_6, \emptyset, +1)$
7. $\delta(q_6, \gamma_4\sigma_{n-1(1)}) = (q_7, |\emptyset|, +1)$
8. $\delta(q_7, \gamma_a\sigma_{n-1(2)}) = (q_6, \emptyset, +1)$
9. $\delta(q_6, \gamma_1\sigma_{n-1(3)}) = (q_6, \emptyset, +1)$
10. $\delta(q_6, \gamma_c\sigma_{n(1)}) = (q_6, \emptyset, -1)$
11. $\delta(q_6, \gamma_z\sigma_{n(2)}) = (q_8, \gamma_d\sigma_{n-m(2)}, +1)$
12. $\delta(q_8, \gamma_1\sigma_{n(3)}) = (q_9, \emptyset, -1)$
13. $\delta(q_9, \gamma_1\sigma_{n-1(3)}) = (q_9, \emptyset, -1)$
14. $\delta(q_9, \gamma_a\sigma_{n-1(2)}) = (q_9, \emptyset, -1)$
15. $\delta(q_9, |\emptyset|\sigma_{n-1(1)}) = (q_9, \emptyset, -1)$
16. $\delta(q_9, \gamma_1\sigma_{n-m(3)}) = (q_9, \emptyset, -1)$
17. $\delta(q_9, \gamma_d\sigma_{n-m(2)}) = (q_9, \emptyset, -1)$
18. $\delta(q_9, |\emptyset|\sigma_{n-m(1)}) = (q_9, \emptyset, -1)$
19. $\delta(q_9, \gamma_2\sigma_{1(3)}) = (q_9, \emptyset, -1)$
20. $\delta(q_{19}, \gamma_z\sigma_{1(2)}) = (q_{10}, \gamma_d\sigma_{1(2)}, -1)$
21. $\delta(q_{10}, \gamma_1\sigma_{1(1)}) = (q_f, \emptyset, 0)$

The instructions 5, 11 and 20 are implementations of licensing constraint (9), and therefore the output is: *taqat(x)\eta ta*.

3.1.3

A third example: $\langle \text{lknd}_8y \rangle$ (133.4):

$\langle \text{ } \rangle$	$\langle k \rangle$	$\langle n \rangle$	$\langle d_8 \rangle$	$\langle y \rangle$
$\langle X_1 \rangle$	$\langle X_2 \rangle$	$\langle X_3 \rangle$	$\langle X_4 \rangle$	$\langle X_5 \rangle$

According to (1), we obtain the following grid:

$\langle \emptyset \rangle$	$\langle \bar{l} \rangle$	$\langle k \rangle$
$\gamma_1 \sigma_{1(1)}$	$\gamma_2 \sigma_{1(2)}$	$\gamma_1 \sigma_{1(3)}$

Then, according to (1b):

$\langle \emptyset \rangle$	$\langle n \rangle$
$\gamma_4 \sigma_{n-m(1)}$	$\gamma_1 \sigma_{n-m(x)}$

this latter, according to (1a), being immediately converted in

$\langle \emptyset \rangle$	$\langle \emptyset \rangle$	$\langle n \rangle$
$\gamma_4 \sigma_{n-m(1)}$	$\gamma_f \sigma_{n-m(2)}$	$\gamma_1 \sigma_{n-m(3)}$

Eventually,

$\langle \emptyset \rangle$	$\langle \bar{l} \rangle$	$\langle k \rangle$	$\langle \emptyset \rangle$	$\langle \emptyset \rangle$	n	$\langle d_\delta \rangle$	$\langle y \rangle$	$\langle \emptyset \rangle$
-----------------------------	---------------------------	---------------------	-----------------------------	-----------------------------	---	----------------------------	---------------------	-----------------------------

from the resulting TM finite tape is:

#	$\gamma_4 \sigma_{1(1)}$	$\gamma_2 \sigma_{1(2)}$	$\gamma_1 \sigma_{1(3)}$	$\gamma_4 \sigma_{n-m(1)}$	$\gamma_f \sigma_{n-m(2)}$	$\gamma_1 \sigma_{n-m(3)}$	$\gamma_1 \sigma_{n(1)}$	$\gamma_a \sigma_{n(2)}$	$\gamma_4 \sigma_{n(3)}$	#
---	--------------------------	--------------------------	--------------------------	----------------------------	----------------------------	----------------------------	--------------------------	--------------------------	--------------------------	---

OTR's output: *ākänd(x) [økendi]*.

3.1.4.

As a final example, let us examine the controversial instance $\langle xwšnw\tau_0 \rangle$ (126.2), cf. § 1.1.9. From the initial string

$\langle x \rangle$	$\langle w \rangle$	$\langle š \rangle$	$\langle n \rangle$	$\langle w \rangle$	$\langle \tau_0 \rangle$
$\langle x_1 \rangle$	$\langle x_2 \rangle$	$\langle x_3 \rangle$	$\langle x_4 \rangle$	$\langle x_5 \rangle$	$\langle x_6 \rangle$

we obtain the following TM tape:

$\langle x \rangle$	$\langle w \rangle$	$\langle š \rangle$	$\langle \emptyset \rangle$	$\langle \emptyset \rangle$	$\langle n \rangle$	$\langle \emptyset \rangle$	$\langle w \rangle$	$\langle \tau_0 \rangle$
$\gamma_2 \sigma_{1(1)}$	$\gamma_a \sigma_{1(2)}$	$\gamma_1 \sigma_{1(3)}$	$\gamma_4 \sigma_{n-m(1)}$	$\gamma_f \sigma_{n-m(2)}$	$\gamma_1 \sigma_{n-m(3)}$	$\gamma_4 \sigma_{n(1)}$	$\gamma_a \sigma_{n(2)}$	$\gamma_1 \sigma_{n(3)}$

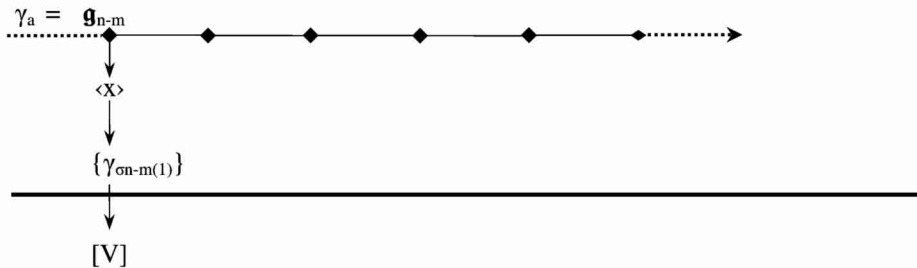
OTR's output: *q(x)šān(x)t*.

4.0. Unpredictability

Let us return to statement (16):

$$(\forall \gamma_1 | \gamma_1 = \gamma_2) \cdot [[A \rightarrow (\gamma_z = \gamma_d)] \vee [B \rightarrow (\gamma_f = \gamma_e)]] \wedge (\forall \gamma_1 | \gamma_1 = \gamma_3) \cdot [[A \rightarrow (\gamma_z = \gamma_b)] \vee [B \rightarrow (\gamma_f = \gamma_c)]]$$

Recalling what we previously stated in paragraph 0.1, we can affirm that γ_1, γ_2 , etc.,⁵⁹ may be represented in the form of one-branch trees stemming from a string of equal-ranked nodes (mother-nodes):



Furthermore, we may say that such a linear relationship between a mother-node and its daughter node is peculiar to each mother-node, i.e. is a property of it. Thus $(\forall \mathbf{g}_{n-m}) \rightarrow (\exists ! \varphi_{n-m} | \mathbf{g}_{n-m} \models \varphi_{n-m})$, i.e. there is exactly one particular linear relationship that turns out to be satisfied for each mother node (17).

This means that iff $(\varphi_{n-m} = \varphi_{n-p}) \rightarrow (\mathbf{g}_{n-m} = \mathbf{g}_{n-p})$.

According to a simplistic interpretation of statement (17), we would say that within any given finite graphemic string, each grapheme exhibits a sequential (ordinal) uniqueness.

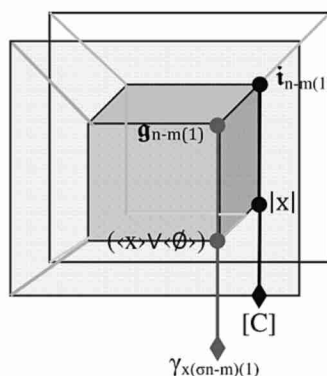
4.1. A brief excursus: *mod-b*

In the following, brief excursus, we will employ the function φ_{n-m} , the property of being linked to a specific daughter -node, to interpret the formalization of the above mentioned empirically verifiable conditions under which our initial statements (1a) and (1b), otherwise assumed as axiomatic, are thinkable as somewhat true (see § 0.2). As already anticipated in § 0.3, this model implies a (costly) logical entanglement, a logical tunneling between the bidimensional universe of representation and the universe of concrete objects.

Let us introduced a three-dimensional variant of *mod-a* (see § 0.2), by splitting the plane of representation into two parallel planes: the proper plane of representa-

⁵⁹ We remind that the symbolic string $\gamma_x \sigma_{n-m(2\pm 1)}$ is a mere positional label which encapsulates some input-information about the “nature” of the correlated \mathbf{g} -object and its exact position within a given finite string of symbols.

tion and the plane of real, concrete objects (graphic signs, phonetic gestures). Both planes exhibit a three-level structure:



mod-b

Although it is by no means strictly essential, we may regard $i_{n-m(1)}$, the “real-world” counterpart of $g_{n-m(2)}$, as the neural state correlated with a specific graphic gesture.

Thus, since we empirically observe the following:

$$(\forall |x| \ i_{n-m(1)} \triangleleft |x|) \cdot ((s|x| \triangleleft [C]) \rightarrow (g_{n-m(2)} \triangleleft \langle \emptyset \rangle))$$

“for every $i_{n-m(1)}$ -object which immediately dominates a certain graphic signs $|x|$, if the immediate right-successor of this latter encodes for a [+cons] segment, *this implies* that the immediate right-successor of the ‘correspondent’ $g_{n-m(1)}$ -object ($g_{n-m(2)}$) is represented by \emptyset in surface representation”;

$$((\forall \langle x \rangle \ i_{n-m(3)} \triangleleft |x|) \cdot (s|x| \triangleleft [V] \vee s|x| \triangleleft [C][+sonorant, +nasal, -labial]) \rightarrow (g_{n-m+1(1)} \triangleleft \langle \emptyset \rangle))$$

“for every $i_{n-m(3)}$ -object which immediately dominates a certain graphic signs $|x|$, if the immediate right-successor of this latter encodes for a vocalic segment, or for a nasal [-labial] segment, *this implies* that the immediate right-successor of the corresponding $g_{n-m(3)}$ -object (*viz.* $g_{n-m+1(1)}$, which dominates the onset of syllable σ_{n-m+1}) is represented by \emptyset in surface representation”;

$$((\forall |x| \ i_{n(2)} \triangleleft |x|) \cdot (\nexists s|x|) \rightarrow (g_{n(3)} \triangleleft \langle \emptyset \rangle))$$

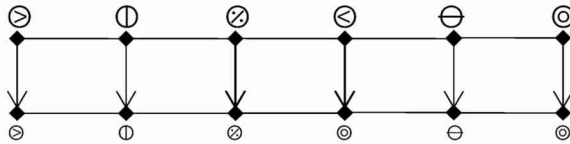
“for every $i_{n(2)}$ -object which immediately dominates a certain graphic signs $|x|$, if there does *not* exist any immediate right-successor of this latter (which is recognizable as the rightward boundary of the given “syntagmatically meaningful graphemic

string”), *this implies* that the immediate right-successor of the correspondent $\mathbf{g}_{n(2)}$ -object ($\mathbf{g}_{n(3)}$) is represented by \emptyset in surface representation”;

we can generalize by saying that, when in the world of concrete objects, a certain $\mathbf{i}_{n-m(x)}$ satisfies the property $\varphi_{n-m(x)}$ ($\mathbf{i}_{n-m(x)} \vDash \varphi_{n-m(x)}$), we infer (logical tunneling) that in our representational universe is $\mathbf{g}_{n-m(2\pm 1)} \vDash \varphi_{n-m(\emptyset)}$.

4.2

Let now consider a well-ordered set of totally arbitrary symbols, namely the alphabet $\Gamma_x = \{\ominus, \oplus, \otimes, \odot, \ominus, \otimes\}$. The following graph (18) represents a finite, well-ordered string of such symbols, each of which exhibits a specific, unambiguous mother-daughter relationship with another symbol.



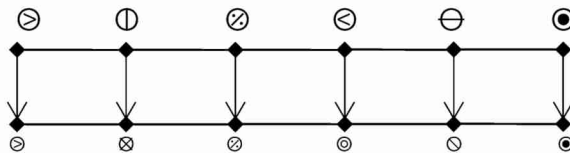
Since, according to (17), $(\forall \mathbf{g}_{n-m}) \rightarrow (\exists ! \varphi_{n-m} | \mathbf{g}_{n-m} \vDash \varphi_{n-m})$, the above mentioned graph implies the following, peculiar set of conditions:

$$\Phi_x = \{ \oplus \vDash \varphi_{\ominus}, \oplus \vDash \varphi_{\oplus}, \otimes \vDash \varphi_{\otimes}, \otimes \vDash \varphi_{\ominus}, \ominus \vDash \varphi_{\ominus}, \ominus \vDash \varphi_{\otimes} \}$$

$$\text{Thus, } (\forall \Gamma_x) \rightarrow (\Gamma_x \vDash \Phi_x)$$

At this point, we proceed to expand our primitive, discursive assumption (§ 0.1), by introducing a more general, fictitious example. Starting from a number of disparate considerations (synchronic, diachronic and comparative), we may state that, at least within the frame of our linguistic domain, a fictitious Old-Turkic syntagm such as $*\langle p'rmk \rangle$ should have been read as *bärmäk*, while a fictitious syntagm such as $*\langle p'rmx \rangle$ should have been read as *barmâq* (roughly speaking: we are looking at a simple case of front-back vowel harmony).

Here we touch on a crucial issue. While we can assume that graph (18) represents precisely the relationship $*\langle p'rmk \rangle \rightarrow bärmäk$, the relationship $*\langle p'rmx \rangle \rightarrow barmâq$ may be, in its turn, represented as follows (19):

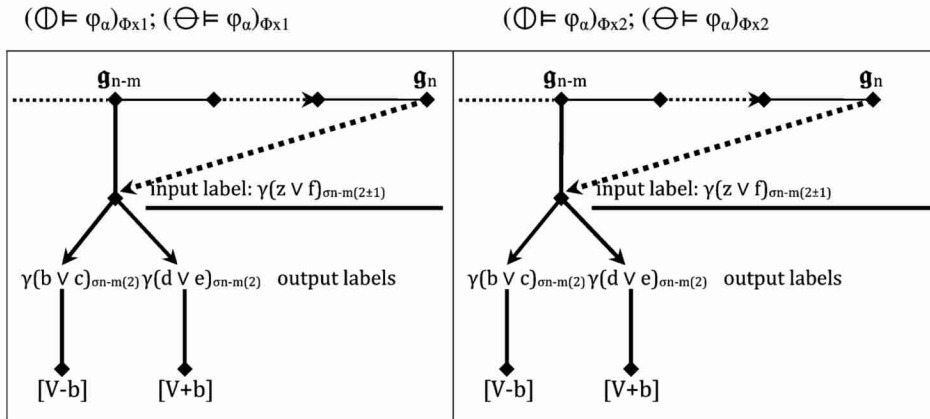


The problem arises that graph (19) disproves statement (17). We are forced to reformulate the representation enunciated in § 4.1 in the following terms:

Each g_{n-m} exhibits three distinct properties: the fact of being *preliminarily* classified as a certain kind of ‘ γ_x ’, i.e. of being associated to a certain input label (of exhibiting the property φ_γ , see § 3.0.); the fact of holding a certain position (labeled $\sigma_{n-m(2\pm 1)}$) within a finite linear string of symbols (of exhibiting the property φ_σ); and the fact of satisfying a certain “vertical”, mother-daughter relationship with an appropriate output-label (the fact of exhibiting the property φ_α). According to graphs (17) and (19):

$\Phi_{x1} = \{ \begin{array}{l} (\ominus \models \varphi_\gamma \wedge \ominus \models \varphi_\sigma \wedge \ominus \models \varphi_\alpha); \\ (\oplus \models \varphi_\gamma \wedge \oplus \models \varphi_\sigma \wedge \oplus \models \varphi_\alpha); \\ (\otimes \models \varphi_\gamma \wedge \otimes \models \varphi_\sigma \wedge \otimes \models \varphi_\alpha); \\ (\oslash \models \varphi_\gamma \wedge \oslash \models \varphi_\sigma \wedge \oslash \models \varphi_\alpha); \\ (\ominus \models \varphi_\gamma \wedge \ominus \models \varphi_\sigma \wedge \ominus \models \varphi_\alpha); \\ (\ominus \models \varphi_\gamma \wedge \ominus \models \varphi_\sigma \wedge \ominus \models \varphi_\alpha); \\ (\ominus \models \varphi_\gamma \wedge \ominus \models \varphi_\sigma \wedge \ominus \models \varphi_\alpha) \end{array} \}$	$\Phi_{x2} = \{ \begin{array}{l} (\ominus \models \varphi_\gamma \wedge \ominus \models \varphi_\sigma \wedge \ominus \models \varphi_\alpha); \\ (\oplus \models \varphi_\gamma \wedge \oplus \models \varphi_\sigma \wedge \oplus \models \varphi_\alpha); \\ (\otimes \models \varphi_\gamma \wedge \otimes \models \varphi_\sigma \wedge \otimes \models \varphi_\alpha); \\ (\oslash \models \varphi_\gamma \wedge \oslash \models \varphi_\sigma \wedge \oslash \models \varphi_\alpha); \\ (\ominus \models \varphi_\gamma \wedge \ominus \models \varphi_\sigma \wedge \ominus \models \varphi_\alpha); \\ (\ominus \models \varphi_\gamma \wedge \ominus \models \varphi_\sigma \wedge \ominus \models \varphi_\alpha); \\ (\ominus \models \varphi_\gamma \wedge \ominus \models \varphi_\sigma \wedge \ominus \models \varphi_\alpha) \end{array} \}$
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

What we *actually* observe can be illustrated by the following graph:



The following statement captures some aspects of the above reformulated representation:

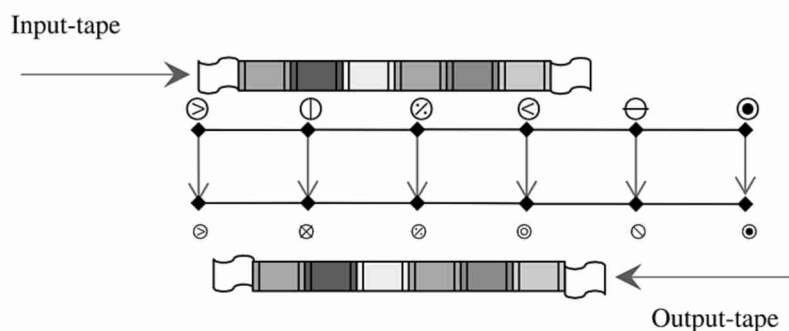
$$\begin{aligned} \text{iff } & (\exists \Phi_{x1})(|\Phi_{x1}| - |(\ominus \models \varphi_\gamma \wedge \ominus \models \varphi_\sigma)| = |\Phi_{x2}| - |(\ominus \models \varphi_\gamma \wedge \ominus \models \varphi_\sigma)|) \\ \rightarrow & [\exists \varphi_\alpha^{\Phi_{x1}} | ((\oplus \models \varphi_\alpha) \in \Phi_{x1})] \end{aligned}$$

$$\begin{aligned} &\wedge ((\ominus \models \varphi_\alpha) \in \Phi_{x1}) \cdot [(\oplus \models \varphi_\alpha^{\Phi_{x1}} \neq \oplus \models \varphi_\alpha^{\Phi_{x2}}) \\ &\wedge (\ominus \models \varphi_\alpha^{\Phi_{x1}} \neq \ominus \models \varphi_\alpha^{\Phi_{x2}})] \quad (20) \end{aligned}$$

What exactly does statement (20) mean? Let us consider the following, straightforwardly discursive observation: the sole “inner” detectable mark which allows us to discriminate between the (possibly harmonic) vowel pattern encoded by the sequence of non-adjacent grapheme ‘>’ and ‘<’, when occurring in the strings *<p’rmØk> (to be read as *bärmäk*) and *<p’rmØx> (to be read as *barmâq*), is the alternation of the graphemes <k> and <x> at the end of the sequence <p’rmØ>.

Such an otherwise insightful observation does not capture the core meaning of statement (20), which pivots around the inequalities $\oplus \models \varphi_\alpha^{\Phi_{x1}} \neq \oplus \models \varphi_\alpha^{\Phi_{x2}}$, and $\ominus \models \varphi_\alpha^{\Phi_{x1}} \neq \ominus \models \varphi_\alpha^{\Phi_{x2}}$.

So, in the end, does this mean that the transition function φ_α is *non-deterministic*? In fact, in this specific case (and only in it), it *does not*, and the Finite State Transducer described in § 3.0 is definitely deterministic (on the equivalence between non-deterministic and deterministic finite automata compare Hopcroft & Motwani & Ullmann [2001: 60-64]; Barua & Gupta [2013: 8]). But in order to capture a complete representation of the φ_α denoted by a symbol occurring in a certain cell on the virtual tape, OTR requires a piece of information which is *possibly not stored in that cell*. This information is *possibly* retrievable from another, distant cell or even the last cell of a given finite string. Our transducer turns out to be a palindrome automaton.



4.3 Encoding ambiguity

Until now, we have exclusively considered examples in which we have been able unambiguously to determine the φ_α of a given **g**-object (its univocal relationship with a phonetic surface representation), while avoiding the specter of phonetic unpredictability. This, instead, appears to be the general rule within our specimen.

As a first, possibly dubious example, let us consider the following string: *pyr_myšyng_{pyr}myngw* (127.4): *bermiš* {-In} *bir mejü*.

The following example is definitely much more insidious. Let us consider again some syntagmatic contexts in which the morpheme {+DA} occurs:

⟨ τ _lx⟨ τ _lynḡt₀⟩ (127.2), cf. § 3.1.2;

⟨sw_βs⟨lmyšynt₀⟩⟩:

⟨pyr_lšyp_{sw}β_slmyšynt₀⟨ τ _lyg_γlyllḡwz_{pyr}šlḡ_{swyyx}swβ⁶⁰l lyšwrslr⟩ *bir ač* {-Ip} *suvsā* {-mIš} {+In} {+DA} *teg yalḡuz bir čan* (< Chinese 盞) *soyik*⁶¹ *suḡ ičür* {-sAr} '[...] when he is thirsty, (he) let (him) drink only a cup of cold water [...]' (125.3–5).⁶² Cf. ⟨pyr_lšlḡ_{swyyx}lswβ_llyšwrmyš⟩ (128.3–4).

⟨py τ yg_γčt₀⟩ bitig {+DA}, cf. § 1.4.1:

stem	σ_{n-2}	σ_{n-1}	σ_n [Kaisse]
*√⟨ τ _l x⟨ τ _l ⟩		⟨ynḡ⟩	⟨ τ ₀ ⟩
√⟨swβs⟩	⟨myš⟩	⟨yn⟩	⟨ τ ₀ ⟩
√⟨py τ yg _γ ⟩			⟨ τ ₀ ⟩

In paragraph 3.1.2 we considered the string ⟨ τ _lx⟨ τ _lynḡt₀⟩ to be calculable by our Finite State Transducer OTR since we assumed that, for every occurrence of the grapheme ⟨ τ ⟩ within the aforementioned “Phonological Word” (or Prosodic Syntagm, whose boundaries are axiomatically taken as unambiguously detectable, see §§ 0.2, 1.3.1), its φ_α will be univocally determinable. But since we *do not know* whether or not, within this specific synchronic domain, the morpheme {+DA} would actually have been opaque to rightward vowel harmony spreading, our assumption turns out to be untenable.

5. Conclusions

In an overwhelming majority of cases, it seems impossible to unambiguously determine the φ_α of a given grapheme (its univocal relationship with a phonetic surface representation). The phonetic unpredictability appears to be the general rule within our specimen.

References

Copeland, B. Jack 2015. The Church-Turing thesis. In: Zalta, Edward N. (ed.) *The Stanford encyclopedia of philosophy*. URL <<http://plato.stanford.edu/archives/sum2015/entries/church-turing/>>

60 Mutō (2008: 244): [...] *bir čan . soyiq suḡ* [...]; Zieme (2013: 101): *bir čan soyix suḡ*; Zieme (2015: 163, l. 016).

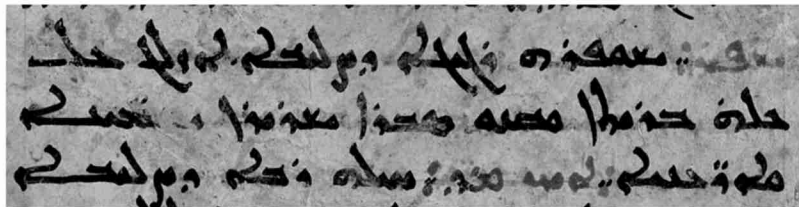
61 Starostin, Dybo & Mudrak (2003: 1336): OTurk. *soyiq*.

62 Zieme (2015: 154, ll. 15–17; 161, note 561).

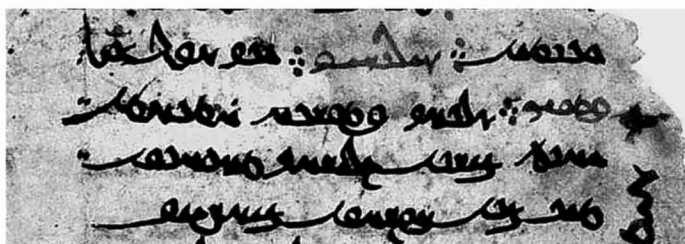
- Dickens, Mark 2013. The importance of the Psalter at Turfan. In: Li & Winkler (2013: 357–380).
- Erdal, Marcel 2004. *An old Turkic grammar*. Leiden: Brill.
- Eraslan, Kemal 2012. *Eski Uygur Türkçesi grameri*. (Atatürk Kültür, Dil ve Tarih Yüksek Kurumu Türk Dil Kurumu Yayınları 1047.) Ankara: Türk Dil Kurumu.
- Finley, Sara 2009. Directionality in vowel harmony. In: Lehnert-LeHouillier, Heike & Fine, Alex B. (eds.) *University of Rochester Working Papers in the Language Sciences* 5: 1, 69–88.
- Gharib, Badresaman 1995. *Sogdian dictionary (Sogdian-Persian-English)*. Tehran: Farhang Publications.
- Graf, Thomas 2010. Formal parameters of phonology: From Government Phonology to SPE. In: Icard, Thomas & Muskens, Reinhard (eds.) *Interfaces: Explorations in logic, language and computation*. (Lecture Notes in Artificial Intelligence 6211.) Berlin: Springer. 72–86.
- Goldsmith, John A. & Riggle, Jason & Yu, Alan C. 2014². *The handbook of phonological theory*. (Blackwell Handbooks in Linguistics) Chichester: Wiley Blackwell.
- Hahn, Reinhard F. 1991. *Spoken Uyghur*. Seattle and London: University of Washington Press.
- Hamilton, James Russel 1986. *Manuscripts ouïgours du IXe–Xe siècle de Touen-Houang. Texts établis, traduits et commentés* 1–2. Paris: Peeters France (Fondation Singer-Polignac).
- Heinz, Jeffrey & Lai, Regine 2013. Vowel harmony and subsequentiality. In: Kornai, András & Kuhlmann, Marco (eds.) *Proceedings of the 13th Meeting on the Mathematics of Language, August 9, 2013, Sofia* (MoL 13.). Madison [WI]: Omnipress. 52–63.
- Hopcroft, John E. & Motwani, Rajeev & Ullmann, Jeffrey D. 2001². *Introduction to automata theory, languages, and computations*. Reading [MA]: Addison-Wesley.
- Hunter, Erica C.D. & Dickens, Mark 2014. *Syrische Handschriften 2. Texte der Berliner Turfansammlung. Syriac texts from the Berlin Turfan Collection*. Stuttgart: Franz Steiner Verlag.
- Jardine, Adam 2014. Logic and the generative power of autosegmental phonology. *Proceedings of the Annual Meetings on Phonology*. URL <<http://journals.linguisticsociety.org/proceedings/index.php/amphonology/article/view/4/41>>.
- Kaisse, Ellen 1986. Towards a lexical phonology of Turkish. In: Brame, Michael & Contreras, Heles & Newmeyer, Frederick J. (eds.) *A Festschrift for Sol Saporta*. Seattle [WA]: Noit Amrofer Publisher Co. 231–240.
- Kracht, Marcus 2003. *The mathematics of language*. (Studies in Generative Grammar 63.) Berlin & New York: Mouton de Gruyter.
- Levelt, Willem Johannes Maria 1974. *Formal grammars in linguistics and psycholinguistics* 1. *An introduction to the theory of formal languages and automata*. The Hague & Paris: Mouton.
- Li Tang & Winkler, Dietmar W. (eds.) 2013. *From the Oxus river to the Chinese shores: Studies on East Syriac Christianity in China and Central Asia*. (Orientalia Paristica Oecumenica 5.). Zurich & Berlin: Lit Verlag.
- Morag, Shelomo 1962. *The vocalization system of Arabic, Hebrew, and Aramaic. Their phonetic and phonemic principles*. (Janua linguarum, 13). 'S-Gravenhage: Mouton & Co.
- Mutō, Shinichi 2008. シリア文字文書 [Documents in Syriac characters]. In: Yoshida, Jun'ichi & Chimeddhorji (2008: 232–252).

- Potts, Christopher & Pullum, Geoffrey K. 2002. Model theory and the content of OT constraints. *Phonology* 19, 361–393.
- Proverbio, Delio Vania 2012. A tentative (graphemically-based) reconstruction of the vowel phonology of an early 18th-century Turkish-*garšūnī* text from Edessa (present-day Şanlıurfa). *Turkic Languages* 16: 2, 200–214.
- Rogers, James 1998. *A descriptive approach to language-theoretic complexity*. Stanford [CA]: Center for the Study of Language and Information.
- Rose, Sharon & Walker, Rachel 2014. Harmony systems. In: Goldsmith, Riggle & Yu (2014: 239–290).
- Sims-Williams, Nicholas 1975. Notes on Sogdian palaeography. *Bulletin of the School of Oriental and African Studies* 38: 1, 132–139.
- Sims-Williams, Nicholas 1981. The Sogdian sound system and the origins of the Uyghur script. *Journal Asiatique* 269, 347–360.
- Sims-Williams, Nicholas 1989. Sogdian. In: Schmitt, Rüdiger (ed.) *Compendium linguarum iranicarum*. Wiesbaden: Ludwig Reichert. 173–192.
- Sims-Williams, Nicholas 2012. *Mitteliranische Handschriften 4. Iranian manuscripts in Syriac Script in the Berlin Turfan Collection*. Stuttgart: Franz Steiner Verlag.
- Skjærvø, P. Oktor 1996. Aramaic scripts for Iranian languages. In: Daniels, Peter T. & Bright, William (eds.) *The world's writing systems*. New York & Oxford: Oxford University Press.
- Starostin, Sergei & Dybo, Anna & Mudrak, Oleg 2003. *Etymological dictionary of the Altaic languages*. (Handbook of Oriental Studies 8, 8/1–3). Leiden & Boston: Brill.
- Vago, Robert M. 1980. *Issues in vowel harmony. Proceeding of the CUNY Linguistic Conference on Vowel Harmony, 14th May 1977*. Amsterdam: John Benjamins.
- Yoshida, Jun'ichi & Chimeddorji 2008. ハラホト出土モンゴル文書の研究 [Study on the Mongolian documents found at Qaraqota] / 吉田順一, チメドドルジ. 東京: 雄山閣, 2008.
- Yoshida, Yutaka 2009. Sogdian. In: Windfuhr, Gernot (ed.) *The Iranian languages*. London & New York: Routledge. 279–335.
- Zieme, Peter 1979. Uigurisch *lab* 'Spende'. *Altorientalische Forschungen* 6, 275–277.
- Zieme, Peter 2006. A cup of cold water. Folios of a Nestorian-Turkic manuscript from Kharakhoto. In: Malek, Roman & Hofrichter, Peter (eds.): *Jingjiao. The Church of the East in China and Central Asia*. St. Augustin: Institut Monumenta Serica.
- Zieme, Peter 2013. Turkic Christianity in the Black City (Xaraxoto). In: Li & Winkler (2013: 99–104).
- Zieme, Peter 2015. Altuigurische Texte der Kirche des Ostens aus Zentralasien. (Gorgias Eastern Christian Studies 41). Piscataway [NJ]: Gorgias Press.

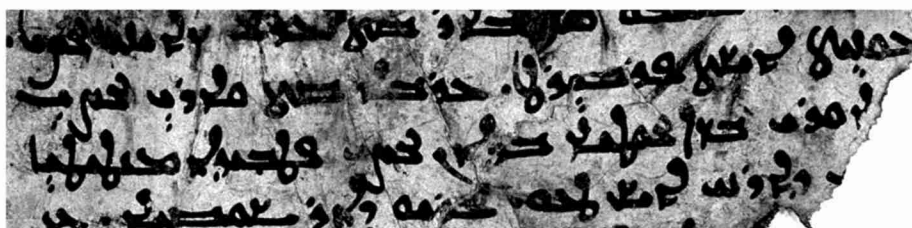
Appendix



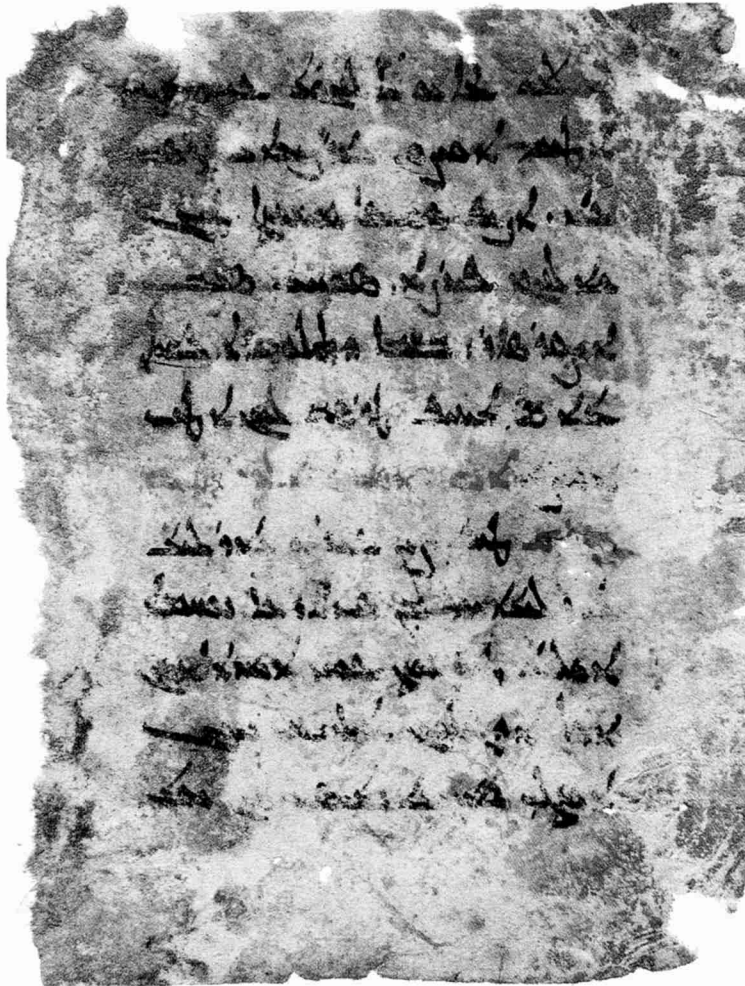
Specimen of a Syriac ms in Syriac script: Ms Berlin, Museum für Asiatische Kunst, MIK III 111v = T II B 37, cf. Hunter & Dickens (2014: 354-355, no. 394).



Specimen of a Syriac ms in Uyghur script: Ms Berlin, Museum für Asiatische Kunst, MIK III 58r, cf. Dickens (2013: 369); Hunter & Dickens (2014: 352-353, no. 391).



Specimen of a Sogdian ms in Syriac script: Ms Berlin-Brandenburgische Akademie der Wissenschaften no. 239r = T III B = E28/66 from Bulayīq; cf. Sims-Williams (2012: 152).



Specimen of an Old Turkic ms in Syriac script: Hohhot (呼和浩特),⁶³ Inner Mongolia Cultural Relics and Archaeology Research Institute (内蒙古自治区文考古研究所), Ms from Xaraxoto (Hēichéng, 黑城): Mutō (2008) no. 125 = Zieme (2015), T, fol. 1 verso.

63 The “Blue City”.

Report

Turkic Linguistics: The State of the Art. Workshop at the University of Mainz in March 2016

Éva Á. Csató

Csató, Éva Á. 2017. *Turkic Linguistics: The State of the Art*. Workshop at the University of Mainz in March 19, 2016. *Turkic Languages* 21, 152–155.

This report gives an account of an international workshop organized at the University of Mainz in March 2016 on the occasion of the incorporation of the Department of Oriental Studies (Seminar für Orientkunde) into the newly established Department of Slavistics, Turcology and Circum-Baltic Studies. The workshop also commemorated the 80th birthday of Lars Johanson, professor of Turcology at the University of Mainz. All of the more than fifty participants had some relation to the Turcology in Mainz, as former doctoral students, research fellows, or project participants.

Éva Á. Csató, Department of Linguistics and Philology, Uppsala University, Box 635, SE-751 26 Uppsala. E-mail: eva.csato@lingfil.uu.se

In January 2016, the Department of Oriental Studies (in German *Seminar für Orientkunde*), at Johannes Gutenberg University Mainz, became part of a larger administrative unit together with Slavistics and Circum-Baltic Studies. The name of the new department is the Department of Slavistics, Turcology and Circumbaltic Studies (in German: *Institut für Slavistik, Turkologie und zirkumbaltische Studien*). This provided a good occasion for Turcologists Hendrik Boeschoten and László Károly to convene a workshop with the aim of discussing the state of the art in Turkic linguistics and to stipulate promising paths for further research collaborations.

The over fifty participants were Turcologists, linguists, former doctoral students, visiting scholars at the former Department of Oriental Studies, and colleagues who for many years had collaborated with the Turcologists in Mainz in various ways.

The participants also commemorated Lars Johanson's 80th birthday. Lars Johanson was born in Sweden and studied at Uppsala University. He has been affiliated with the University of Mainz since 1972, first as a Humboldt scholar and since 1973 as a "Privatdozent". In 1981 he succeeded Professor Johannes Benzing to the chair of Turcology at the Department of Oriental Studies. During his time as professor, Mainz became an internationally known center of Turkic linguistics. Johanson supervised over ten doctoral students, most of them today themselves professors at different Turkish and West European universities. Several leading Turcologists worked as Humboldt scholars in Mainz. Johanson initiated and, during the first two

periods, led the project “Historical and Linguistic Aspects of Turco-Iranian Contacts in the South Anatolian and West Iranian Area”, at Mainz University 1997–2008.

The opening speech was held by Professor Hendrik Boeschoten, who in 2002 was appointed professor of Turcology in Mainz as the successor to Lars Johanson.

The first lecture, on *Diminutive and Honorific Language Means in Kazakh and Siberian Turkic*, was given by Irina Nevskaya of Goethe University of Frankfurt and Novokuzneck State Pedagogical Institute. The written version of this talk, co-authored by Saule Tazhibaeva, will be published in *Turkic Languages*. Nevskaya was awarded a Humboldt scholarship and worked as a research fellow in Mainz in 1998 and 1999.

Marcel Erdal, Free University of Berlin, also a former Humboldt fellow in Mainz, reported on *The Dialectological Status of the Runiform Sources*. Between 2013 and 2016, Erdal has been engaged in a research project on Old Turkic language varieties (*Alttürkische Sprachvarietäten*). The project aims to identify dialectal variation in Turkic inscriptions and manuscripts from the period 800–1300, found mainly in Mongolia, South Siberia, and Kyrgyzstan.

Uli Schamiloğlu, University of Wisconsin, gave a talk on *The Rise of Runic Turkic as the First Turkic Vernacular Literary Language*. He argued that Syriac Turkic, Volga Bolgharian, and the Middle Turkic literary language of the Golden Horde all ceased to exist as a result of the Black Death and were later replaced by epigraphic and literary languages closer to the vernacular. A written version of this paper will be published in *Turkic Languages*.

Martine Robbeets, Max Planck Institut für Menschheitsgeschichte, Jena, earlier Humboldt fellow and now affiliated to Mainz University as “Privatdozent”, has had a long and inspiring collaboration with Lars Johanson. The two scholars introduced and consolidated the term “Transeurasian languages” for a grouping consisting of Japonic, Koreanic, Tungusic, Mongolic, and Turkic. They have contributed several edited volumes to the theoretical discussion of genealogical and areal relations among these languages. At the symposium, Robbeets presented her concept of core structures of Transeurasian as a follow-up to Lars Johanson’s paper *So Close and Yet So Distant... On Turkic Core Structures, Genealogical and Typological Grouping of Varieties, and Mutual Intelligibility* published in *Ankara Papers in Turkish and Turkic Linguistics* (Zeyrek, D. et al. eds. 2015: 583–592). The written version of Robbeets’ paper is published in this issue of *Turkic Languages*.

A. Sumru Özsoy, Bogaziçi University, has been working on Turkish Sign Language since 2004. She gave an interesting talk on comparatives in Turkish Sign Language with a number of videos illustrating the strategies used by Turkish signers to express different types of comparisons. A written version co-authored with Hünar Kaşıkara will be published in a forthcoming volume, “Comparative Constructions” (edited by Wojtylak, K. & Treis, Y. & de Vries L., Free University of Amsterdam).

Carol Pfaff, Freie Universität Berlin, has carried out comprehensive and pioneering studies on Turkish as spoken in Germany. For many years she has collaborated with Lars Johanson on building a network of scholars studying the development of

spoken varieties of Turkish in Northwestern Europe (TINWE). At the workshop, she spoke about continuity and contact-induced change in pronominal and demonstrative usage in Turkish in Germany, summarizing her findings from her empirical studies of three generations of children and adolescents in Berlin.

Fikret Turan, Istanbul University, compared grammars of modern Turkish used as textbooks in the undergraduate and postgraduate degree programs in Turkish Language and Literature in Turkey. He demonstrated certain differences in their approach, interpretation, and overall methodologies concerning the presentation of Turkish sentence types and phrase structures.

Delio V. Proverbio, Vatican Apostolic Library, Rome, has in recent years published a series of studies analyzing the writing systems used for East Old Turkic texts. In his talk, he analyzed graphotactic regularities in Old Turkic texts in Syriac script. The written version is published in this issue of *Turkic Languages*.

Gülschen Sahatova, Aristotle University of Thessaloniki, dealt with the 18th century Turkmen text *Destan Sayatlî Hemra*. She gave a short description of the content of this literary work and analyzed some linguistic aspects of the text: orthography, morphophonemics, and the morphological inventory.

Mustafa Uğurlu, University of Muğla, and Ahmet Aydemir, Hacettepe University, received their PhD degrees in Mainz under the supervision of Lars Johanson. Uğurlu gave a comprehensive review of research on Turkic languages in Turkey. Aydemir presented interesting results of his first fieldwork in a Tuvan community in Western Mongolia.

Sema Demir Aslan, Hacettepe University, has visited Mainz University as a research fellow. Her studies in the field of Turkish and Turkmen verbal systems are based on Lars Johanson's viewpoint aspect theory. The written version of her talk on viewpoint aspect in Turkmen is published in this issue of *Turkic Languages*.

Nurettin Demir, Hacettepe University, also a former PhD student in Mainz, arrived with a special gift, the Turkish translation of Lars Johanson's book *Aspekt im Türkischen*, published in Uppsala in 1971. The Turkish translation *Türkçede Görünüş* (Ankara: Grafiker) will help Turkish scholars to get acquainted with this work. Demir has also translated a number of other publications by Lars Johanson into Turkish, for instance *Türkçe Dil İlişkilerinde Yapısal Etkenler* (Türk Dil Kurumu, 2007) (original title *Structural Factors in Turkic Language Contacts*, 2002) and *Türk Dili Haritası Üzerinde Keşifler*, 2002 (original title *Discoveries on the Turkic Linguistic Map*, 2001).

A memorable contribution to the workshop was Göran Hammarström's talk on *New Linguistic Fundamentals*. Göran Hammarström, emeritus professor at the University of Melbourne, has on several occasions been guest professor in Roman Linguistics in Mainz. His acquaintance with Lars Johanson goes back to the years 1959–1960, when Hammarström was professor of phonetics in Uppsala and Lars Johanson was a student.

The participants agreed that it is vital to continue the fruitful cooperation in the field of Turkic linguistics in the future. The next project will be a comprehensive *Encyclopaedia of Turkic Languages and Linguistics* to be published by Brill.

Tunguso-Sibirica

Herausgegeben von Michael Weiers

39: Dolgor Guntsetseg

Differential Case Marking in Mongolian

2016. XIV, 204 pages, 16 diagrams, 10 ill.,
26 tables, pb
170x240 mm
ISBN 978-3-447-10611-5
©E-Book: ISBN 978-3-447-19494-5
each € 48,- (D)

Mongolian is a regular DOM (Differential Object Marking) language: the accusative case does not always occur on direct objects. This book investigates the phenomenon starting from the research question of how the Mongolian pattern is influenced by factors that cross-linguistically trigger DOM, such as referentiality, animacy, and topicality. It shows that the examination of any one of these factors on its own is not sufficient, but rather that DOM emerges from a complicated interaction of these features. In addition to DOM, Mongolian also exhibits a specific type of Differential Subject Marking (DSM), in which the subjects of embedded clauses (including adverbial clauses) occur with the accusative case. This is the second issue investigated in the study. Apart from the features already mentioned, sentence structure turns out to be relevant here. More specifically, the adjacent occurrence of main and embedded subjects is identified as a crucial factor for triggering DSM. Both observations about DOM and DSM in Mongolian can be brought together in the generalization that the accusative case in Mongolian is used to distinguish between two arguments not only within a clause but also across clause boundaries. The book provides a detailed analysis of the relevant components of Mongolian grammar, and its findings are supported by extensive experimental studies with a large number of native speakers to ensure the highest quality of linguistic evidence.

40: Andrew Shimunek

Languages of Ancient Southern Mongolia and North China

A Historical-Comparative Study of the Serbi or Xianbei Branch of the Serbi-Mongolic Language Family, with an Analysis of Northeastern Frontier Chinese and Old Tibetan Phonology

2017. L, 519 pages, hc
170x240 mm
ISBN 978-3-447-10855-3 Ca. € 98,- (D)
In Vorbereitung / In Preparation

This is the first book on the Serbi-Mongolic language family and the first modern linguistic study of the Serbi (Xianbei 鮮卑) peoples, whose conquest of North China took place at approximately the same time as the Germanic and Hunnic *Völkerwanderung* into the former Western Roman Empire. The findings presented in this book – the first rigorous and systematic unified theory on the origins of the Mongolic and Serbi languages – add substantially to our understanding of the linguistic geography of Eastern Eurasia, and to the ethnolinguistic history of the Mongolic peoples and their neighbors, including speakers of Chinese, Japanese-Koguryoic, Tibeto-Burman, Tungusic, possibly Indo-European, and later, Turkic. This book also enhances our understanding of attested Middle Chinese, Early Old Mandarin, and Old Tibetan phonology. Moreover, it is the first study to present linguistic sketches of Taghbach (Tuoba 拓跋), Tuyuhun 吐谷浑 ('Azha 𐰽𐰺𐰍), and Kitan (Qidan 契丹), and to systematically compare Kitan and Mongol morphological and syntactic paradigms, resulting in the first reconstruction of Common Serbi-Mongolic phonology, morphology, lexicon, and syntax. Readers interested in Mongolia, the Mongols, North China, Central Eurasia, the Tibetan Empire, languages of Asia, historical linguistics, and history will find this book to be a useful resource.

HARRASSOWITZ VERLAG · WIESBADEN

www.harrassowitz-verlag.de · verlag@harrassowitz.de