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—Preussicher Kulturbesitz for access to and permission to reproduce images of some relevant fragments. All images are copyright from *Depositum der Berlin Brandenburgischer 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 ‹pyrkʾymn›, which actually occurs in the aforementioned Old Turkic document written in Syriac script, should have been read as *b*(x)*r-kȧymän* (*bergȧymän*), while a fictitious syntagm such as *‹pyrxʾymn› should have been read as *b*(x)*rqaymån*.

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):

$\mathfrak{W}$ = at the underlying representation level, a $\mathfrak{W}$-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;

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

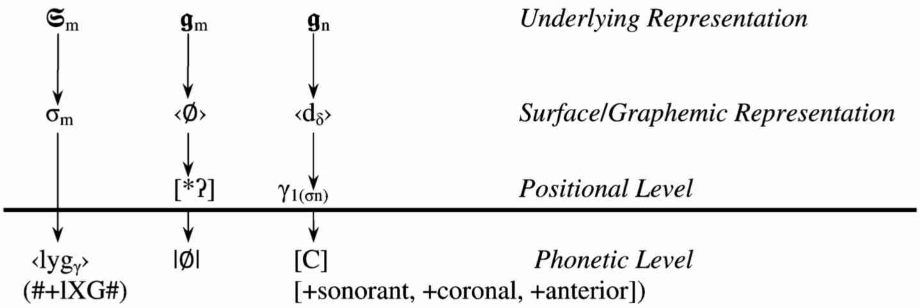
followed by a list of binary place-predicates (beyond the usual Boolean operators and quantifiers):

$\lhd$ = immediate domination
$\lhd^*$ = domination
$\prec$ = linear precedence (left-of)
$\nprec$ = linear subsequence (not left of)

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

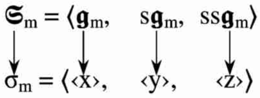
3 For the sake of economy, *viz.* in order to not weigh strings down, in the following we will write $\mathfrak{g}_{(1)}$, or even $\mathfrak{g}_{n}$ instead of $\mathfrak{g}_{n\text{-}m(1)}$, $\mathfrak{g}_{p(n)}$—where the leftmost subscript, non-bracketed numerical variable expresses the ordinality 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 ($\mathfrak{S}_m \lhd \sigma_m$), ($\mathfrak{g}_m \lhd$ ‹Ø›) and ($\mathfrak{g}_n \lhd^*$ [C]).

The atomic formula (x ≺ 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



we may say that ‹x› ≺ ‹z›—since, obviously, $\mathfrak{g}_m \prec ss\mathfrak{g}_m$, and since $\mathfrak{g}_m \lhd$ ‹x› and $ss\mathfrak{g}_m \lhd$ ‹z›—while, *z does not* precedes *y* (‹z› ⊀ ‹y›).

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 $\gamma_{1(\sigma n\text{-}m)}$, 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 [*ʔ]; 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 ‹Ø›) in surface representation; in fact, it turns out to have no phonetic correspondence (|Ø|).

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 ($\mathfrak{g}_m \lhd \langle\emptyset\rangle \lhd [*ʔ] \lhd |\emptyset|$). 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:

$$(\forall \mathfrak{W}_p)(\forall \mathfrak{S}_{(1)},\mathfrak{S}_{(n\text{-}m)},n)\ [\bigwedge_{i=1}^{n} \mathfrak{W}_p \lhd \mathfrak{S}_i \longrightarrow \bigvee_{i<j=n} \mathfrak{S}_{i=j}] \quad (1a)$$

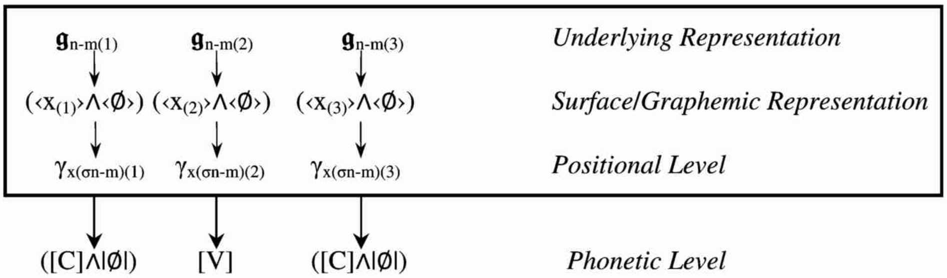
– which says that there can be at most n $\mathfrak{S}$ under a $\mathfrak{W}$;

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

which, roughly speaking, mean that every $\mathfrak{S}$-object is invariably and only constituted by the triplet $\langle \mathfrak{g}_{(1)}, \mathfrak{g}_{(2)}, \mathfrak{g}_{(3)} \rangle$;[5]
Both statements (1c) and (1d) are found to be true:

$$(\forall \mathfrak{g}_{n\text{-}m})(\forall x_{(1)},23)\ [x_{(1)} \prec x_{(2)} \prec x_{(3)} \longrightarrow ((\bigwedge_{i=1}^{3} (\langle x_{(i)} \rangle)\mathfrak{g}_{n\text{-}m})\ (\vee\ \mathfrak{g}_{n\text{-}m} = \lhd \langle\emptyset\rangle))] \quad (1c)$$

$$(\forall \mathfrak{S}_{n\text{-}m})(\forall \mathfrak{g}_{(1)}, \mathfrak{g}_{(2)}, \mathfrak{g}_{(3)})\ [\mathfrak{g}_{(1)} \prec \mathfrak{g}_{(2)} \prec \mathfrak{g}_{(3)} \longrightarrow ((\bigwedge_{i=1}^{3} (C)\mathfrak{g}_{(i)} \vee \mathfrak{g}_{(i)} = \lhd^* |\emptyset|) \wedge ((V)\mathfrak{g}_{(i)}))] \quad (1d)$$



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

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 Utterances[6]—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: ‹pyrm𐰞q› *bir* {-mAK} (131.1) CVC.CVC *versus* ‹p𐰞rm𐰞ʕ𐰞y̆› *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›	‹𐰞›	‹ʕ›	‹∅›	‹𐰞›	‹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 $\mathfrak{g}_{n\text{-}m}$-object may be represented by zero in surface representation—provided that the general statement [($\mathfrak{g}_{n\text{-}m(2\pm1)}$ ⊲* [C]) ⟶ ⊤)] implies that (∄$\mathfrak{S}_{n\text{-}m}$|$\mathfrak{g}_{n\text{-}m(1)} \in \mathfrak{S}_{n\text{-}m} \wedge \mathfrak{g}_{n\text{-}m(1)}$ ⊲* [V])—vocalic-onset syllables are here conventionally represented as $[*ʔ][V][C]$, with a fictitious glottal stop [ʔ] in onset position, encoded by a zero-surface representation ($\mathfrak{g}_m$ ⊲ ‹∅› ⊲ [*ʔ]). Furthermore, σ_n syllables exhibiting a $[C][V]$ phonetic structure are here conventionally represented as $[C][V][*ʔ]$, with a fictitious glottal stop [ʔ] in coda position.

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

‹𐰞ybyn͡gk𐰞ƺ› (124.1), *ev* {+Iŋ} {+KA}, but also ‹𐰞ybyn͡g𐰞ƺ› (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 syllabication take place [...]".

C	V	C	C	V	C	C	V	C
‹∅›	‹⁊y›	‹b›	‹∅›	‹y›	‹n͡g›	‹k›	‹⁊›	‹∅›

C	V	C	C	V	C	C	V	C
‹∅›	‹⁊y›	‹b›	‹∅›	‹y›	‹n͡g›	‹∅›	‹⁊›	‹∅›

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: t_1, t_2; d_1, d_2 etc.), we will adopt an arbitrary set of alphabetic symbols, taken from the Greek alphabet.

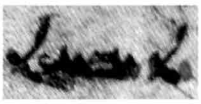
1.1.1.

‹d_{δ}› comes ultimately from Syriac ‹d› (‹ܕ›); ‹t_{θ}› comes from Syro-Sogdian ‹θ›, which in turn derives from Syriac ‹t› (‹ܬ›); ‹t_{τ}› comes ultimately from Syriac ‹ṭ› (‹ܛ›). 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 ‹ܐ› are here merged into the encoding symbol ‹⁊›: the isolated form, the final form and the simplified final form. The grapheme ‹⁊›, when followed by a grapheme encoding for a [+cons] segment, may encode for one of the following vocalic segments: [ɑ], [a] ([ɐ]) (cf. *infra*, § 2.1.1).

The sequences ‹⁊y› and ‹⁊w› are intrinsically ambiguous, since, as digraphs, they may encode for single vocalic segments. Cf. , for example, ‹⁊y⁊x› *ayaq* (124.3, 4, 7), *versus* ‹⁊ybynγ⁊ƶ› (124.3), *ev* {+In} {+KA}:



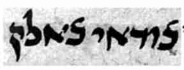
C	V	C	versus	C	V	C
‹Ø›	‹ˀ›	‹y›		‹Ø›	‹ˀy›	‹b›

The sequence ‹ˀwy› is also intrinsically ambiguous; when interpreted as a trigraph, it encodes for a single vocalic segment, cf. for example ‹ˀwytˀkˀymn̊› (128.2) *ötä* {-GAy} *m*(*ä*)*n*:

C	V	C	C	V	C	C	V	C	C	V	C
‹Ø›	‹ˀwy›	‹t›	‹Ø›	‹ˀ›	‹k›	‹Ø›	‹ˀ›	‹y›	‹m›	‹Ø›	‹n›

1.1.3.

‹x›():[9] This grapheme is foreign to Syriac script, but occurs in Sogdo-Syriac writing[10] (,[11] )[12] as well as in early Neo-Persian manuscripts written in Syriac script ().[13] It is clearly derived from Syriac ‹k› and, at any rate, has nothing to do with Avestan  ‹x›.[14] In Modern Judeo-Persian manuscripts we observe a parallel phenomenon: the grapheme ‹x› (‹خ›) appears to be encoded by means of a "modified" grapheme ‹k› (‹כ›) 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 & Chimeddhorji 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 ‹p› and ‹f›.

and obstruent *begadkephat* allophonic variants (these latter being the zero-marked ones):[16]

Deut XI 11: וְהָאָרֶץ אֲשֶׁר אַתֶּם עֹבְרִים שָׁמָּה לְרִשְׁתָּהּ אֶרֶץ הָרִים וּבְקָעֹת

1.1.4.

‹g_γ›: as inferred from the immediate evidence—an inference which obviously must be thoroughly examined—the grapheme ‹g_γ›, derived from the ancestral Syriac grapheme ‹g› (‹ܓ›), appears to encode for a [+dorsal, -back] consonant ([x̟]). Therefore, in order to prevent any confusion with the Old Turkic consonant represented as *γ* (for example, in Gabain's transcription system), we avoid adopting the transcription encoding usually employed for Sogdo-Syriac (‹ܓ›→‹γ›). As far as the sonorant nasal dorsal [ŋ] is concerned, the graphemic cluster ‹ܢܓ› is transliterated as ‹n͡g›.

1.1.5.

‹ʕ›: cf. ‹˥wʕwl›[17] *oġul* (130.3); ‹ṣyʕ˥yx˥› *čïġay* {+KA} [ʧɨɢayxa:] (128.1, 131.6); ‹ṣyʕ˧y̽ny̽› (124.2; 128.6); ‹p̱˥rm˥ʕ˥y̽› *bar* {-mA -GAy} (128.1); ‹syʕymlyxꝫ (127.5) *sïġïm* {+lXG}; ‹[xy]lʕxyn͡g˥› *ḳïlġaḳ* {+iŋA} (126.12, cf. *infra*); ‹pwlʕ˥ly› *bol* {-GAlI} (132.4). Outside genuine Syriac contexts, this grapheme appears to occur exclusively in intervocalic position (or after a lateral sonorant) to encode for a combinatory (allophonic) variant of a [+dorsal, +back] consonant. An erratic occurrence (though always as a combinatory variant of /q/) is ‹x˥θyʕ› *ḳatïġ* (130.1) *versus* ‹˥ryx› *arïġ* (124.8) (*versus* ‹nyθyg_γ› *ne-teg*, 124.8). Cf. *infra*, §§ 1.2.3 and 1.4.4.2.

1.1.6.

‹ḥ›: this grapheme seems only to occur in Syriac lexemes, such as ‹rḥym›, ‹mšyḥ˥› (cf. *infra*).

1.1.7.

‹q›: apparently, outside genuine Syriac contexts (e.g. 124.11: ‹kl‿dmšqʾ› → ‹dqryrʾ›, cf. Zieme [2013: 101]: Mt X 42: καὶ ὃς ἂν ποτίσῃ → ψυχροῦ), and not taking into account Sogdian loan-words such as *tarsāk* (‹$t_r$rsq̊lyg_γ› 125.6; ‹$t_r$rsq˥› (128.10, 132.5), occurrences of grapheme ‹q› seem erratic. But cf. *infra*, § 1.2.1.

16 This is exactly the opposite of the function exhibited by *qypyå* or *råpé* diacritics, cf. Morag (1962: 33, note 57; 34, note 59).

17 Zieme (2015: 157, f. 4r l. 075): "ʾwkwl".

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't›: ‹xwšnwt$_{\theta}$ fryt$_{\theta}$⌉˛t$_{\theta}$¸[19] kwyn͡gwl|pyrl⌉› *ḳušånu̇t frïtat köŋül birlȧ* '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 [xusɒnt].[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 *[xušɒnɨt] > [xušɒnʉt] in Turkic mouths. Cf. *infra*, § 3.1.4. Cf. also the following occurrence: ‹˛k¸yšy¸˛nyn͡g¸kwyn͡gly¸nyn͡g˛˛xwšn¸wt$_{\theta}$|[xy]]lʕxyn͡g⌉›[23] *kiši* {+nIŋ} *köŋli* {+nIŋ} *ḳušånu̇t ḳïlġaḳ* {+iŋA} (126.11–12, cf. § 2.1.2) 'for the happiness of a human heart'.

The second lexeme is a phonetic calque from Sogdian ‹fryt't› *frītāt*,[24] 'love', cf. Gharib (1995: 157b, no. 3969, 3977).

Manichean		Sogdo-Syriac		Turko-Syriac	
 ‹k›	 ‹x›	Ø	 ‹x›	 ‹k›	 ‹x›
 ‹p›	 ‹f›	 ‹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 *utθaduš*, 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 *ï* […] 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 ʿ anzunehmen".

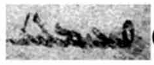
24 Therefore, it is definitely not to be read *qutluγ*, as Zieme does in Mutō (2008: 243), nor even *birtäg* (Zieme 2015: 160).

1.2. A morphological survey

1.2.1. Morpheme #+lXG#

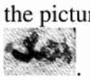
‹yˀrlyx› *yarlïġ*[25] (125.8);

‹yˀrl˛y˛xˀlsˀr› *yarlï(ġ)ḳa* {-sAr} (128.5); ‹yˀrlyxˀlmˀmyšẑ› *yarlï(ġ)ḳa* {-mA} {-mIš} (124.9), [26] ‹˛yˀ˛rlyxˀyw˛r˛› [27] (125.9), ‹yˀrl˛y˛xˀywr› (131.3) *yarlï(ġ)ḳa* {-yUr};

‹syʕymlyx̊›  (127.5).[28]

Let us examine the syntagmatic context of this occurrence: ‹snyn͡g˽kwṣwn͡g˽ ytₜmyšṣˀ | pyr˛m˛yšyn͡g˽pyr˽myn͡gw ˽tₜn͡grykˀ | syʕymlyx˽tₜwrwr› *säniŋ küčüŋ yitmiš* {-čA} *bermiš* {-Iŋ} *bir meŋü täŋri* {+kA} *sïġïm* {+lXG} *turur* ‘your (best) tribute to the unique eternal God, (given) to the utmost of your capacity, is the *highness of heart*’.[29] *Sïġïm* {+lXG} is a denominal syntagm which we hypothesize to be derived from the Sogdian compound-word ‹sqˀ-mˀn›, *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 *[sqɑmɑn] ([sqɑmən]) to *[səɢɑmən] > [səɢəmən] ≈ [sɨɢɨmɨn], this latter issue exhibiting the expected splitting of the onset consonant cluster /sq/. The resulting syntagm would have been easily analyzed as √*sïġïm* {+In}.

25 Regarding this lexeme, cf. Clauson (1972: 966b(–967a)): “Although morphologically a P.N./A. in *-lïğ*, it cannot be so explained etymologically, and this fact, taken with the fact that in Manichaean and Uyg. script it is habitually spelled *yrlğ*, less often *yrlïğ*, 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): “*yarlïqamaqï*”; 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): *yarlïqayay-m<ä>n*, afterward corrected by Zieme (2013: 101); Zieme (2015: 154, f. 1v, 021).

28 Mutō (2008: 244): *suqïmlïq*; Zieme (2015: 156, f. 3v, 065): *sy‘ymlyx*; 160: *sïxïmlïx*; 164: “*sy‘ymlyx*. Vermutlich, wenn man *sïxïmlïg* liest, eine Ableitung von *sïg-* ‘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 (?)’.

‹rḥym̊lyx›[30] (131.7; ‹rḥym̊› 127.11, 129.10) a denominal syntagm composed of the Syriac loan-word *rḥem* ‘mercy, compassion’. Cf. §§ 1.2.4, 2.1.2.

‹ˀwt$_\tau$lyl˛yx¸› *utlï* +lXG (126.9) *versus*

‹ˀlylg$_\gamma$› *elig*: ‹ˀlylg$_\gamma$› (127.9), ‹ˀlylg$_\gamma$n› (124.1).

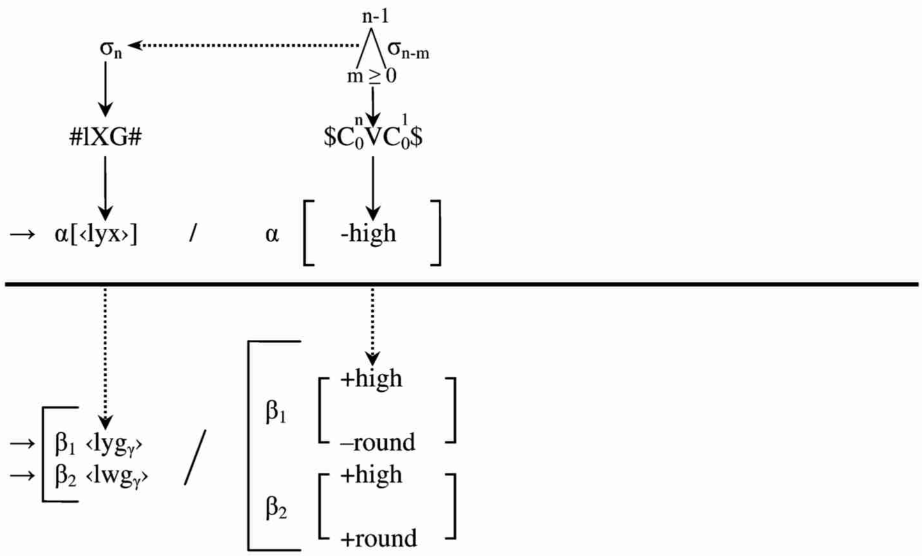
‹t$_\tau$rsq˽lyg$_\gamma$› (125.6), cf. ‹t$_\tau$rsqˀ› (132.5): a denominal syntagm made up of the Sogdian loan-word ‹trsʾq› *tarsāk* ‘Christian’,[31] cf. Gharib (1995: 391a, no. 9667). According to some graphotactic evidence (cf. § 2.1.3), this loan-word appears actually to have been pronounced [tɛrsɛk] by Turkic speakers.

‹ˀdg$_\gamma$wlwg$_\gamma$› *ädgülüg* (125.11); this denominal syntagm occurs within the following syntagmatic context: ‹ˀwl˽ˀdg$_\gamma$wlwg$_\gamma$˽xyl˛myš¸[y]nyn͡glˀwt$_\theta$ly̌sy̌› *ol ädgülüg ḳïlmïš* {+In} {+Iŋ} *utlï* {+sI} ‘reward for his good deed’.[32]

‹kwyn͡gwllwg$_\gamma$› *köŋüllüg* (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>qïmlïq*, correctly translated as 慈悲を; Zieme (2015: 158, f. 4v, l. 091): *rḥym 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 ädgülüg qïlïnč-nïng* | *utlï-sï*; Zieme (2013: 101); Zieme (2015: 154), f. 1v, ll. 022–023: *ʾwl ʾdgwlwg xylmyšynyng* | *ʾwtly sy.*

We can try to better understand the above-mentioned constraint schema by putting forward the following descriptive statement: *we assume that* the morpheme #+lXG#, 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 $\sigma_{n\text{-}m\,[n > m \geq 0]} \in N_{(input)} \wedge \sigma_n \in N_{(output)} \wedge \langle u, u' \rangle \in V(Nuc)$
- input $\wedge$ ($u \wedge$ [αhigh]) $\rightarrow \boxed{io}\, \varphi$ ($u' \wedge$ [αhigh])[34]

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

	‹lyk›, ‹lwk›: [+lik] / [+lük]	l. 11: ‹ʾrklyk› *ärklig*; 13: ‹twykʾʾlyk› *tükällig*
British Library, Ms Or. 8218 (122)[35]		l. 12: ‹kwyclwk› *küčlüg*; 15: ‹pwykwylwk› *bögülüg*
	‹lyq›: [+lïq]	l. 10: ‹tylyq␣lʾr␣nynk› *tïnlïḳlar* {+nIŋ}

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 $(\sigma_n \mid \sigma_n \lhd {*}\#\text{-lXG}\#) \in N_{(input)} \subseteq N_{(output)} \wedge u \in C(Coda) \wedge u' \in V(Nuc)$
- input $\wedge$ ($u \wedge$ [αhigh]) $\rightarrow \boxed{io}\, \varphi$ ($u' \wedge$ [αhigh])

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 #+lXG#) is an input node and a specific daughter node (its coda) is [αhigh], 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 [αhigh] feature". Roughly speaking: *we assume* that the [αhigh] feature actually exhibited by the coda consonant of the morpheme #+lXG# (*via* the graphemic opposition ‹x› *versus* ‹g_γ›) 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#

‹swyzl˥šm˥q› *sözläš* {-mAK} (133.5);

‹pyrm˥q› *bir* {-mAK} (131.1), ‹pyrm˥˻q˼lyg$_\gamma$› (127.10): *bir* {-mAK} {+lXG}, *versus* ‹pyrm˥klyg$_\gamma$› (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›.

‹xylm˥xlyx› *ḳïlmaḳlïḳ*:

‹˥wzwn̽˽ywlx˥ | ˽˥zwx˽pw˽ṣyx˥y̽˽x˥˽l˥p̽ | ˽pwšy̽˽pyrm˥˻q˼lyg$_\gamma$˽$t_\tau$wrwr ˽ṣyx˥y̽ | ˥wyz˥˽rḥym̽˽xylm˥xlyx | $t_\tau$wrwr› *uzun yol* {+KA} *azuḳ bu* †*čïġay* {+KA} *lab buši̇* [here the Sanskrit lexeme लाभ *lābha* is immediately followed by its Chinese translation 布施 *bùshī*, which refers to the Buddhist practice of almsgiving] *bir* {-mAk} {+lXG}‡ [†‡ = 131.6][36] *tur -Ur čïġay üzȧ råhïm ḳïl* {-mAk} {+lXG} *tur* {-Ur} '[the appropriate] provision for the "long trip" is almsgiving to the poor and showing mercy upon the poor' (127.8–12);

‹y˥rm˥x̽› *yar* {-mAK} (129.[5], 6).[37]

1.2.5 Morpheme #+KA#:[38]

‹pwy˥n̽|x˥› *buyan* (< Sanskrit पुण्य *puṇya*) {+KA}: ‹pwy˥n̽|x˥˽s˻˥˼xynyp˽$t_\tau$rsq˽lyg$_\gamma$˽ ˥t_τy|˥wṣ˻wn˼› *buyan* {+KA} *saḳïn* {-Ip} *tärsäk* {+lXG} *at* {+I} *üčün* 'thinking (to do) a good deed for (the reason of) being a Christian [lit.: because of his Christian name]' (125.5–7).[39]

‹ṣyʕ˥yx˥› *čïġay* {+KA} (128.1, 131.6).

36 Regarding the syntagm #*lab buši̇ 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).

‹ˈlyṣwrmš̽kˈl› *içür* {-mIş} {+KA} (128.4), ‹ˈlyṣwrsˈlr› *içür* {-sAr} (125.5).
‹pyzkˈl› *biz* {+KA} (127.8²).[40]
‹pyrkˈl› *bir* {+KA} (128.2, 4).
‹t$_τ$n͡grykˈl› *täŋri* {+KA} (127.4) (cf. ‹t$_θ$n͡gry̽›, 127.1).
‹ˈlyβyn͡gkˈl› (124.1) *ev* {-Iŋ} {+KA} *versus* ‹ˈlyβyn͡gˈl› (124.3).

1.2.6 Morpheme #-GAy#[41]

‹pˈlrmˈlʕˈly̽› *bar* {-mA} {-Gay} (128.1).
‹pyrmˈlkˈlymw› *ber* {-mA} {-GAy} mU (128.9).
‹pyrkˈlymn› *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 #+lXG# 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 ynkʾ› http://vatec2.fkidg1.uni-frank furt.de/vatecasp/Maitrisimit_0-2.htm: r16(MaitrGeng1196), v06(MaitrGeng1756)] *versus* ‹pyz kʾ› [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 𐰋𐰃𐰔𐰚𐰀 ‹b²İzk²A› seems the only possible one according to the sole available drawing (Yïldïrïm & Aydïn & Alimov 2013: 101, http://img856. images hack.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 [± high], before it reaches σ_n:

Stem →	$\sigma_{n\,[Kaisse]}$	stem	σ_{n-1}	$\sigma_{n\,[Kaisse]}$
√‹ʾly›	‹lyg$_\gamma$›	√‹ṣyʕʾly›		‹xʾ›
√‹ʾwt$_\tau$ly›	‹lyx›	√‹ywl›		‹xʾ›
*√‹rḥym›	‹lyx›	*√‹pwyʾn̆̊›		‹xʾ›
*√‹synym›	‹lyx›	√‹ʾlyṣwr›	‹mš̊›	‹kʾ›
√‹yʾr›	‹lyx›	√‹pyz›		‹kʾ›
		√‹pyr›		‹kʾ›
		√‹t$_\tau$n͡gry›		‹kʾ›

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]-(±high) is applied[44] —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, $\sigma_{n\text{-}m}$) will turn out to be the head trigger of a *leftward* (regressive) harmony (Spread-L[αF]: 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:

Iff $((\forall \mathfrak{S}_x | \mathfrak{S}_x \in \mathfrak{W}_z) \wedge (\forall \sigma_{n\text{-}m[n > m \geq 0]} | \ \mathfrak{S}_x \lhd \sigma_{n\text{-}m}))$
- $\wedge\ ((\forall \mathfrak{S}_y | \mathfrak{S}_y \in \mathfrak{W}_z) \wedge (\forall \sigma_{n\text{-}p[n > p \geq 0;\ p \neq m]} | \ \mathfrak{S}_y \lhd \sigma_{n\text{-}p}))$
- $\wedge\ [(\sigma_{n\text{-}m} \prec \sigma_{n\text{-}p}) \vee (\sigma_{n\text{-}p} \prec \sigma_{n\text{-}m})] \wedge [(\sigma_{n\text{-}m} \in N_{(input)}) \wedge (\sigma_{n\text{-}p} \in N_{(output)})]$
- $\wedge\ [(u'_{input} \in M(Onset) \vee (u'_{input} \in P(Coda))] \wedge (u''_{output} \in M(Nuc))$

input $\wedge\ (u' \wedge [\alpha high]) \rightarrow \boxed{io}\ \varphi\ (u'' \wedge [\alpha high])$ (4)

This we may loosely read: "If a syllable $\sigma_{n\text{-}m}$ is an input node and a specific daughter node (its onset or its coda) exhibits a certain [αhigh] feature, then each daughter node *u*" of every correspondent output $\sigma_{n\text{-}p}$ (namely, its syllabic nucleus) exhibits the same [αhigh] feature". Such a statement is to be considered a formal description of a [αhigh] 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 [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 $\mathfrak{W}_n$—where[45]

$$[(\mathfrak{W}_n = \langle \mathfrak{S}_1 \ldots \mathfrak{S}_n \rangle \in W |\ W \subset S \supset S_\sigma) \wedge (\forall \mathfrak{S}_{n\text{-}m[n > m \geq 0]} \in \mathfrak{W}_x |\ \mathfrak{S}_{n\text{-}m} \in S_\sigma) \cdot (\mathfrak{S}_{n\text{-}m} = \langle \mathfrak{g}^{\o}_{n\text{-}m}, s\mathfrak{g}^{\o}_{n\text{-}m}, ss\mathfrak{g}^{\o}_{n\text{-}m} \rangle)]^{46} = Cond_1$$

which we may read: "for every sequence of graphemic strings $\mathfrak{W}_n = \langle \mathfrak{S}_1 \ldots \mathfrak{S}_n \rangle$ 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 $s\mathfrak{g}^{\o}_{n\text{-}m}$ 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}: ‹$t_\theta$wrwr› (124.10; 127.5, 10, [12]), *versus* ‹$t_\tau$wrwr› (128.11, 131.7);

täŋri: ‹t_τn͡gry› (126.3) *versus* ‹t_θn͡gry̆› (127.1);

bitig: ‹pyt$_\tau$yg$_\gamma$ƻt$_\theta$⅂› *bitig* {+DA}: ‹nyt$_\theta$g$_\gamma$ƻ kymƻ ⅂ryxƻ ⅂wn͡glywnƻ nwm̊ƻl pyt$_\tau$yg$_\gamma$ƻ t$_\theta$⅂ƻ mšyḥ⅂ƻ mrn̊ƻ y⅂rlyx⅂m⅂myšƻl mwƻt$_\theta$wrwr›, *netäg kim arïγ ewangelyōn nom bitig* {+DA} *mšīḥā māran yarlï(ġ)ḳa* {-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* ‹pyt$_{\theta}$˪y˩g$_\gamma$› (131.3).

We may provisionally argue that the opposition between ‹t_θ› (‹⅂nt$_\theta$⅂x› *andaḳ* [128.10]) and ‹t_τ› (‹⅂rt$_\tau$wx› *artuḳ* [(127.2]) does not necessarily involves 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 *δāl* 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 $\mathfrak{S}_{n\text{-}m} = \langle \mathfrak{g}^{\o}_{n\text{-}m}, s\mathfrak{g}^{\o}_{n\text{-}m}, ss\mathfrak{g}^{\o}_{n\text{-}m} \rangle$ is intended to be equivalent to $\mathfrak{S}_{n\text{-}m} = \langle \mathfrak{g}^{\o}_{n\text{-}m(1)}, s\mathfrak{g}^{\o}_{n\text{-}m(2)}, ss\mathfrak{g}^{\o}_{n\text{-}m(3)} \rangle \rangle$; 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 Uygur […] *-mAmIš* appears rarely, and only in very late Uygur 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$_{\delta}$›? It seems to occur only in the following contexts:

a: ‹ꞁd$_{\delta}$› *äd* (130.4), ‹ꞁd$_{\delta}$yn› *adïn* (128.4), ‹ꞁd$_{\delta}$g$_{\gamma}$wlwg$_{\gamma}$› *ädgülüg* (125.11), ‹ꞁyd$_{\delta}$mꞁƻsꞁrƻsn› *ïdma* {-sAr} {-sAn} (124.5); ‹ꞁyd$_{\delta}$mꞁsꞁr› *ïdma* {-sAr} (128.7); ‹ywd$_{\delta}$wp› *yüd* {-Xp}: here is its syntagmatic context: ‹swyǀ yꞁzwxny␣ ywd$_{\delta}$-wp␣kyt$_{\theta}$ꞁrt$_{\tau}$ꞁṣy̌ǀ ṣyʕꞁy␣xꞁ␣lꞁp␣pwšy␣pyrmꞁklyg$_{\gamma}$ǀ t$_{\tau}$wrwr› (131.4-5),[50] *suy* [here the chinese lexeme *zui* 罪 is glossed by its Turkic translation] *yazuḳ* {+nI} *yüd* {-Xp} *keδär*[51] {-DAčI} *čïġay* {+KA} *lab buši̇̈ bir* {-mAK} {+lXG} *tur* {-Ur}: which may be translated as 'Almsgiving to the poor (*čïġayḳa lab buši̇̈ birmaklig*) will remove the *charging of sin* (*suy yazuḳnï yüdüp*)'.[52]

b: ‹pwld$_{\delta}$ỵ̄› *bol* {-DI} (124.6), ‹ꞁknd$_{\delta}$y› *ikindi* (133.4).

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

(5) Environment of ‹d$_{\delta}$›:

Iff $Cond_1$[53] $\wedge$ ($\forall \mathfrak{g}_{\text{n-m(x)}} | \ \mathfrak{g}_{\text{n-m(x)}} \lhd$ ‹d$_{\delta}$›)[54]

input→ [(($\mathfrak{S}_{\text{n-m}} = \mathfrak{S}_1$) $\wedge$ ($\mathfrak{g}_{1(1)} \lhd$ (‹ꞁ› $\vee$ ‹y›) $\wedge$ ($\mathfrak{g}_{1(2)} \lhd$ (Ø $\vee$ ‹y› $\vee$ ‹w›) $\wedge$ ($\mathfrak{g}_{1(3)} \lhd$ ‹d$_{\delta}$›)) $\vee$ ($\mathfrak{g}_{\text{n-m(1)}} \lhd$ ‹d$_{\delta}$›) $\wedge$ ($\mathfrak{g}_{\text{n-m-1(3)}} \lhd$*[C][+sonorant, +coronal, +anterior])] = Env_1[55]

which may be read:

"given $Cond_1$ (i.e., for each graphemic string $\mathfrak{S}_{\text{n-m}} = \langle \mathfrak{g}_{\text{n-m(1)}}, \mathfrak{g}^{\varnothing}_{\text{n-m(2)}}, \mathfrak{g}_{\text{n-m(3)}} \rangle$ which encodes for a CVC syllable $\sigma_{\text{n-m}}$);

and for every occurrence of grapheme ‹d$_{\delta}$› as surface representation of the underlying object $\mathfrak{g}_{\text{n-m(x)}}$;

→ the grapheme ‹d$_{\delta}$› will occur *exclusively* in syllable σ_1, this latter being *exclusively* represented either by the strings ‹ꞁØd$_{\delta}$›, ‹ꞁyd$_{\delta}$› or ‹ywd$_{\delta}$›—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$_{\delta}$ as onset of the syllable which immediately follows $\sigma_{\text{n-m}}$, the underlying object $\mathfrak{g}_{\text{n-(m-1)(3)}}$ 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 "$\forall \mathfrak{g}_{\text{n-m(x)}} | \ \mathfrak{g}_{\text{n-m(x)}} \lhd \in N$[input])" is here held always to be implied by the simple statement "$\forall \mathfrak{g}_{\text{n-m(x)}}$".

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

Besides the aforementioned contexts, the grapheme ‹d_δ› occurs only in some Sogdian loanwords, such as ‹$d_\delta$ynd$_\delta$˥r› *dïndar* (133.1).[56] This lexeme is calqued directly from Sogdian ‹δynδʾʾr› (cf. Gharib [1995]: 149b). One thing that stands out is that not only does the Aramaic grapheme ‹d› 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 ‹d_δ› would have represented the voiced apico-dental continuant [ð].

1.4.3.1.

To sum up: the Turco-Syriac graphemes ‹t_τ›, ‹t_θ› and ‹d_δ›—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:

[+tense] ◄·························► [-tense]

‹t_τ› ‹t_θ› ‹d_δ›

▼ ▼ ▼

[t] [θ] [ð]

1.4.3.2.

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

[+tense] ◄·························► [-tense]

‹x› ‹ʕ›

▼ ▼

[χ] [ɢ]

1.5.

Let us compare the two formulas which describe the environments of ‹d_δ› and ‹ʕ› (see §§ 1.1.5 and 1.4.2):

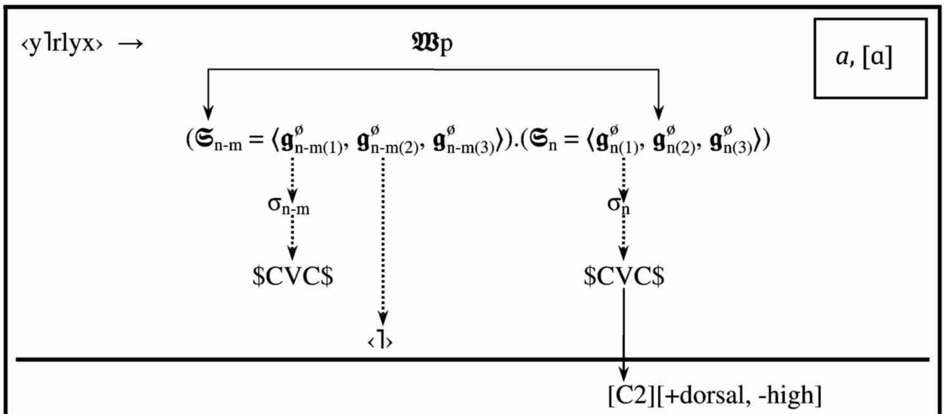
56 Here I must rely upon Zieme's transcription, as I cannot draw any reliable information from the photograph provided by Yoshida & Chimeddhorji (2008: 363).

(6) Environment of ‹d_δ›:
Iff $Cond_1 \wedge (\forall \mathfrak{g}_{n\text{-}m(x)} \mid \mathfrak{g}_{n\text{-}m(x)} \lhd$ ‹d_δ›$)$
input→ $[((\mathfrak{S}_{n\text{-}m} = \mathfrak{S}_1) \wedge (\mathfrak{g}_{1(1)} \lhd$ (‹ᒣ› $\vee$ ‹y›$) \wedge (\mathfrak{g}_{1(2)} \lhd (\emptyset \vee$ ‹y› $\vee$ ‹w›$)$
$\wedge (\mathfrak{g}_{1(3)} \lhd$ ‹d_δ›$)) \vee (\mathfrak{g}_{n\text{-}m(1)} \lhd$ ‹ d_δ ›$) \wedge (\mathfrak{g}_{n\text{-}(m\text{-}1)(3)} \lhd^*$ [C][+son, +coronal, +ant])] =
Env_1

(7) Environment of ‹ʕ›:
Iff $Cond_1 \wedge (\forall \mathfrak{g}_{n\text{-}m(x)} \mid \mathfrak{g}_{n\text{-}m(x)} \lhd$ ‹ʕ›$)$
input → $[(\mathfrak{g}_{n\text{-}m(3)} \lhd$ ‹ʕ›$) \wedge (\mathfrak{g}_{n\text{-}(m+1)(1)} \lhd$‹ᒣ›$) \vee (\mathfrak{g}_{n\text{-}m(1)} \lhd$ ‹ʕ›$)$
$\wedge (\mathfrak{g}_{n\text{-}(m\text{-}1)(3)} \lhd^*$ [C][+son, +coronal, +ant])] = *Env_2*

2.0. A partial vowels inventory

2.1.1.



(9)
Iff $Cond_1 \wedge (\mathfrak{g}_{n(3)} \lhd^*$ [C][+dorsal, -high]$) = Cond_{4a}$
$\wedge (\forall \mathfrak{g}^{\emptyset}_{n\text{-}m(2)} \mid \mathfrak{g}^{\emptyset}_{n\text{-}m(2)} \lhd$ ‹ᒣ›$)$[57] $= Cond_{5a}$
input → $(\mathfrak{g}^{\emptyset}_{n\text{-}m(2)} \lhd^*$ [V] [-front, +low, +back, -round]).

"Given $Cond_1$; if $\mathfrak{g}_{n(3)} \in \mathfrak{S}_n$ is the underlying representation of a [+dorsal, -high] consonant ([χ]), coda of syllable σ_n; for every $\mathfrak{g}_{n\text{-}m(2)}$ (the input) which dominates an occurrence of grapheme ‹ᒣ›—i.e. which is not zero in surface representation (immediately dominates a non-zero surface-representation)—and which therefore is the underlying representation of a non-disambiguated [αfront] vocalic nucleus of $\sigma_{n\text{-}m}$ (cf. §§ 0.2, 1.1.2);
→ this $\mathfrak{g}_{n\text{-}m(2)}$ should encode for [ɑ] ([-front, +low, +back, -round])".

57 Cf. above, note 50.

2.1.2.

‹rḥym̊lyx› *råhïmlïġ* → $\mathfrak{W}_p$

å, [ɒ]

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

(10)

Iff $Cond_1 \wedge Cond_{4a} \wedge (\forall \mathfrak{g}^{\emptyset}_{n\text{-}m(2)} \mid \mathfrak{g}_{n\text{-}m(2)} \lhd \langle\emptyset\rangle) = Cond_{5b}$

input → ($\mathfrak{g}_{n\text{-}m(2)} \lhd$*[V][-front, +low, +back, +round])

Given *Cond₁* and $Cond_{4a}$; for every $\mathfrak{g}_{n\text{-}m(2)}$ (the input) which *is* zero in surface representation (which immediately dominates a zero surface-representation);
→ this $\mathfrak{g}_{n\text{-}m(2)}$ should encode for [ɒ] ([-front, +low, +back, +round]).
Cf. ‹[xy]lʕøx› [χɨlɢɒχ] (126.12), ‹t$_θ$øbꞁr› *tåv* {-Ar} (130.5); ‹ꞁydmꞁƻsꞁrƻsøn› *ïdma* {-sAr} *sån* (124.5).

2.1.3.

‹swyzlꞁšmꞁq› *sözlȧš* {-mAK} → $\mathfrak{W}_p$

ȧ [a], [ɐ]

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

(11)

Iff $Cond_1 \wedge (\mathfrak{g}_{n(3)} \lhd$*[C][+dorsal, +high, +tense]) $= Cond_{4b} \wedge Cond_{5a}$

input → ($\mathfrak{g}^{\emptyset}_{n\text{-}m(2)} \lhd$*[V][-front, +low, -back])

"Given *Cond₁*; if $\mathfrak{g}_{n(3)}$ is the underlying representation of a [+dorsal, +high, +tense] consonant ([k]), coda of syllable σ_n; and given $Cond_{5a}$;
→ then $\mathfrak{g}^{\emptyset}_{n\text{-}m(2)}$ should encode for [a] ([-front, +low, -back])".

2.1.4.

‹t$_τ$rsq˽lyg$_γ$› *tärsäklig* → $\mathfrak{W}_p$

ä, [ɛ]

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

(12)

Iff $Cond_1 \wedge (\mathfrak{g}_{n(3)} \lhd$*[C][+dorsal, +high, -tense]) $= Cond_{4b} \wedge Cond_{5b}$

input → ($\mathfrak{g}^{\emptyset}_{n\text{-}m(2)} \lhd$*[V] [+front, -low, -back, -tense])

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

2.1.5.

‹⅂ylyg$_\gamma$› *elig* → $\mathfrak{W}_p$ | *e*, [e]

$$(\mathfrak{S}_{n\text{-}m} = \langle \mathfrak{g}^{\emptyset}_{n\text{-}m(1)}, \mathfrak{g}^{\emptyset}_{n\text{-}m(2)}, \mathfrak{g}^{\emptyset}_{n\text{-}m(3)} \rangle).(\mathfrak{S}_n = \langle \mathfrak{g}^{\emptyset}_{n(1)}, s\mathfrak{g}^{\emptyset}_{n(2)}, ss\mathfrak{g}_{n(3)} \rangle)$$

(13)
Iff $Cond_1 \wedge Cond_{4b} \wedge (\mathfrak{g}^{\emptyset}_{n\text{-}m(2)} \lhd (\langle y\rangle \vee \langle ⅂y\rangle) = Cond5_c$
input → $(\mathfrak{g}^{\emptyset}_{n\text{-}m(2)} \lhd^*$[V] [+front, -low, -back, +tense])

"Given $Cond_1$ and $Cond_{4b}$; and given $Cond_{5c}$;
→ then $\mathfrak{g}_{n\text{-}m(2)}$ should encode for [e] ([+front, -low, -back, +tense])"
Cf. ‹⅂ykỹ› *eki* (128.2, 8), but also ‹nyg$_\gamma$w› (128.12) *negü*, ‹nyt$_\theta$yg$_\gamma$› (124.8) *ne-täg*.

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

Iff $Cond_1 \wedge ((Cond_{4a} \vee Cond_{4b}) \wedge (Cond5_a \vee Cond_{5b}))$
input → $(\mathfrak{g}_{n\text{-}m(2)} \lhd^*$[V] [-round]) ∨ $(\mathfrak{g}_{n\text{-}m(2)} \lhd^*$[V] [+round])
∨ $(\mathfrak{g}_{n\text{-}m(2)} \lhd^*$[V] [-front, +low]) ∨ $(\mathfrak{g}_{n\text{-}m(2)} \lhd^*$[V] [+front, -low]).

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:

$(\exists \mathfrak{g}_{n\text{-}m(2\pm1)} | \mathfrak{g}_{n\text{-}m(2\pm1)} \lhd^*$[C][+dorsal, -high) = $Cond_{4c}$
$(\exists \mathfrak{g}_{n\text{-}m(2\pm1)} | \mathfrak{g}_{n\text{-}m(2\pm1)} \lhd^*$[C][+dorsal, +high) = $Cond_{4d}$

we obtain (15):

Iff $(Cond_{4c} \vee Cond_{4d}) \wedge (Cond5_a \vee Cond_{5b})$
→ $(\mathfrak{g}_{n\text{-}m(2)} \lhd^*$[V] [-round]) ∨ $(\mathfrak{g}_{n\text{-}m(2)} \lhd^*$[V] [+round])
∨ $(\mathfrak{g}_{n\text{-}m(2)} \lhd^*$[V] [-front, +low]) ∨ $(\mathfrak{g}_{n\text{-}m(2)} \lhd^*$[V] [+front, -low]).

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:

$\Gamma = \Gamma_{input} \cup \Gamma_{output} = \{\ \gamma_1 = \mathfrak{g}^{\sigma}{}_{n\text{-}m(2\pm1)} \in C;$
$\gamma_a = \mathfrak{g}^{\sigma}{}_{n\text{-}m(2)} \lhd^*[V_x]\ |(([V_x] \in V_{\neg\emptyset}) \wedge (V_{\neg\emptyset} \not\ni \langle\emptyset\rangle));$
$\gamma_z = \mathfrak{g}^{\sigma}{}_{n\text{-}m(2)} \lhd \langle\rceil\rangle;$
$\gamma_b = \mathfrak{g}^{\sigma}{}_{n\text{-}m(2)} \lhd^*[V_x]\ |\ [V_x] \in V_{[\text{-front, +low, -back}]};$
$\gamma_c = \mathfrak{g}^{\sigma}{}_{n\text{-}m(2)} \lhd^*[V_x]\ |\ [V_x] \in V_{[\text{+front, -low, -back, -tense}]};$
$\gamma_d = \mathfrak{g}^{\sigma}{}_{n\text{-}m(2)} \lhd^*[V_x]\ |\ [V_x] \in V_{[\text{-front, +low, +back, -round}]};$
$\gamma_e = \mathfrak{g}^{\sigma}{}_{n\text{-}m(2)} \lhd^*[V_x]\ |\ [V_x] \in V_{[\text{-front, +low, +back, +round}]};$
$\gamma_f = (\mathfrak{g}^{\sigma}{}_{n\text{-}m(2)} \lhd \langle\emptyset\rangle) \wedge (\mathfrak{g}^{\sigma}{}_{n\text{-}m(2)} \lhd^*[V_x]);$
$\gamma_2 = \mathfrak{g}^{\sigma}{}_{n\text{-}m(2\pm1)} \lhd^*[C_x]\ |\ [C_x] \in C_{[\text{+dorsal, -high}]};$
$\gamma_3 = \mathfrak{g}^{\sigma}{}_{n\text{-}m(2\pm1)} \lhd^*[C_x]\ |\ [C_x] \in C_{[\text{+dorsal, +high}]};$
$\gamma_4 = (\mathfrak{g}^{\sigma}{}_{n\text{-}m(2\pm1)} \lhd \langle\emptyset\rangle) \wedge (\mathfrak{g}^{\sigma}{}_{n\text{-}m(2\pm1)} \lhd^*[C_x]);$
$\sigma_{1(p = 2 \pm 1)};$
$\sigma_{n\text{-}m(p = 2 \pm 1)};$
$\sigma_{n(p = 2 \pm 1)}\ \}$

where C is (the set of all the $\mathfrak{g}$-objects which immediately dominate) [= *Statement*$_{\mathfrak{g}1}$] graphemes encoding for [+cons] segments; $C_{[\text{+dorsal, +high}]} \subset C$ is [*Statement*$_{\mathfrak{g}1}$] graphemes encoding for [+cons, +dorsal] segments specified for the [+high] feature; $C_{[\text{+dorsal -high}]} \subset C$ is [*Statement*$_{\mathfrak{g}1}$] graphemes encoding for [+cons, +dorsal] segments specified for the [-high] feature; V is [*Statement*$_{\mathfrak{g}1}$] graphemes encoding for [+vocalic] segments; $V_{\neg\emptyset} \subset V$ is [*Statement*$_{\mathfrak{g}1}$] *non-zero* graphemes encoding for [+vocalic] segments; $V_{[\text{-front, +low}]} \subset V$ is [the set of all the $\mathfrak{g}$-objects which dominate = *Statement*$_{\mathfrak{g}2}$] [+vocalic] segments specified for the [-front, +low] features; $V_{[\text{+front, -low}]} \subset V$ is [*Statement*$_{\mathfrak{g}2}$] [+vocalic] segments specified for the [+front, -low] features; $V_{[\text{-round}]} \subset V$ is [*Statement*$_{\mathfrak{g}2}$] [+vocalic] segments specified for the [-round] feature; $V_{[\text{+round}]} \subset V$ is [*Statement*$_{\mathfrak{g}2}$] [+vocalic] segments specified for the [+round] feature; $\sigma_{1(p = 2 \pm 1)}$, $\sigma_{n\text{-}m(p = 2 \pm 1)}$, $\sigma_{n(p = 2 \pm 1)}$ 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:

$((\forall\gamma_z)\ (\gamma_1 \prec \gamma_z) \vee (\gamma_z \prec \gamma_1)) = A$
$((\forall\gamma_f)\ (\gamma_1 \prec \gamma_f) \vee (\gamma_f \prec \gamma_1)) = B$

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

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

Then, we can implement a Deterministic Finite State Transducer OTR = (Q, Σ, Γ, δ, #, q0, qf) 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 $\gamma_4$occur in a given cell when, respectively, a [+vocalic] or a [+cons] segment is encoded by ‹Ø› in surface representation.

3.1.1.

First example: let us consider the string ‹pyrkʅymn› *ber* {-GAy} {mAn} (128.5). As a first step, we 'interpret' it according to our CVC syllabification model (statement 1):

‹p›	‹y›	‹r›	‹k›	‹ʅ›	‹y›	‹m›	‹n›
‹x_1›	‹x_2›	‹x_3›	‹x_4›	‹x_5›	‹x_6›	‹x_7›	‹x_8›
$\mathfrak{g}_{1(1)}$	$\mathfrak{g}_{1(2)}$	$\mathfrak{g}_{1(3)}$	$\mathfrak{g}_{2(1)}$	$\mathfrak{g}_{2(2)}$	$\mathfrak{g}_{2(3)}$	$\mathfrak{g}_{3(1)}$	?

Now, since (s‹x_7› ⊲ [C]) ⟶ ($\mathfrak{g}_{n-m(2)}$ ⊲ ‹Ø›), it is the case that:

$\mathfrak{g}_{1(1)}$	$\mathfrak{g}_{1(2)}$	$\mathfrak{g}_{1(3)}$	$\mathfrak{g}_{2(1)}$	$\mathfrak{g}_{2(2)}$	$\mathfrak{g}_{2(3)}$	$\mathfrak{g}_{3(1)}$	$\mathfrak{g}_{3(2)}$	$\mathfrak{g}_{3(3)}$
↓	↓	↓	↓	↓	↓	↓	‹Ø›	↓
[$C_{1(1)}$]	[$V_{1(2)}$]	[$C_{1(3)}$]	[$C_{2(1)}$]	[$V_{2(2)}$]	[$C_{2(3)}$]	[$C_{3(1)}$]	[$V_{3(2)}$]	[$C_{3(3)}$]

We then transform it into the following TM finite tape:

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

Let OTR starts:

01. $\delta(q_0, \gamma_{1\sigma1(1)}) = (q_1, Ø, +1)$
02. $\delta(q_1, \gamma_a\sigma_{1(2)}) = (q_1, Ø, +1)$
03. $\delta(q_1, \gamma_1\sigma_{1(3)}) = (q_1, Ø, +1)$
04. $\delta(q_1, \gamma_3\sigma_{n-m(1)}) = (q_2, Ø, +1)$
05. $\delta(q_2, \gamma_z\sigma_{n-m(2)}) = (q_3, \gamma_b\sigma_{n-m(2)}, +1)$
06. $\delta(q_3, \gamma_1\sigma_{n-m(3)}) = (q_2, Ø, +1)$
07. $\delta(q_4, \gamma_1\sigma_{n(1)}) = (q_2, Ø, +1)$
08. $\delta(q_2, \gamma_f\sigma_{n(2)}) = (q_4, \gamma_c\sigma_{n(2)}, +1)$
09. $\delta(q_4, \gamma_1\sigma_{n(3)}) = (q_5, Ø, -1)$
10. $\delta(q_5, \gamma_c\sigma_{n(2)}) = (q_5, Ø, -1)$
11. $\delta(q_5, \gamma_1\sigma_{n(1)}) = (q_5, Ø, -1)$
12. $\delta(q_5, \gamma_1\sigma_{n-m(3)}) = (q_5, Ø, -1)$
13. $\delta(q_5, \gamma_b\sigma_{n-m(2)})) = (q_5, Ø, -1)$
14. $\delta(q_5, \gamma_3\sigma_{n-m(1)}) = (q_5, Ø, -1)$
15. $\delta(q_5, \gamma_1\sigma_{1(3)}) = (q_5, Ø, -1)$
16. $\delta(q_5, \gamma_a\sigma_{1(2)}) = (q_5, Ø, -1)$
17. $\delta(q_5, \gamma_1\sigma_{1(1)}) = (q_f, Ø, 0)$

	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	$(q_1, Ø, +1)$	$(q_1, Ø, +1)$	$(q_1, Ø, +1)$		
q_2				$(q_2, Ø, +1)$	
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	$\gamma_1\sigma_{n\text{-}m(3)}$	$\gamma_1\sigma_{n(1)}$	$\gamma_f\sigma_{n(2)}$	$\gamma_1\sigma_{n(3)}$
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 ($\delta(q_2, \gamma_z\sigma_{n\text{-}m(2)}) = (q_3, \gamma_b\sigma_{n\text{-}m(2)}, +1)$) 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)*rkȧymän* [berkɐymɛn].

3.1.2.

Another example is: ‹t_τꞁxꞁ$t_\tau$yn͡gt_θꞁ› (127.2). The lexeme ‹t_τꞁxꞁt_τ› being a loan-word based on Syriac *taḥt* ‘under, beneath’ (which eventually became [taχat] in Turkic mouths), should possibly not be translated as ‘worship’:[58] ‹pyr␣myn͡gw␣t_θn͡grÿ̆ ␣syn͡g| t_τꞁxꞁ$t_\tau$yn͡gt_θꞁ␣ꞁr$t_\tau$wx␣ nꞁrsf̈r|$t_\tau$ylꞁmꞁš̈› *bir meŋü täŋri sän* {+Iŋ} *taḳat* {+Iŋ} {+DA} *artuḳ nȧ (ȧ)r* {-sAr} *tilȧ* {-mAz} (127.1–3) ‘An eternal God does not require anything more than your *submission*’.

‹t_τ›	‹ꞁ›	‹x›	‹ꞁ›	‹t_τ›	‹y›	‹n͡g›	‹t_τ›	‹ꞁ›
‹x_1›	‹x_2›	‹x_3›	‹x_4›	‹x_5›	‹x_6›	‹x_7›	‹x_8›	‹x_9›

Now, since (s‹x_3› ⊲ [V]) ⟶ ($\mathbf{g}_{2(1)}$ ⊲ ‹∅›), but also (s‹x_5› ⊲ [V]) ⟶ ($\mathbf{g}_{2(1)}$ ⊲ ‹∅›) (see statement (1)), so it is the case that:

58 As did Mutō (2008: 244): “崇拝を”. But cf. Zieme (2015: 164): “[…] *taxat* ‘Gehorsam’, vgl. osm. *taat* (طاعت) ‘act of obedience to God; act of piety’ (Redhouse 2011, 1072a) (< ar. √ *ṭāʿ*)”. Though perfectly conceivable, and semantically convincing—although *taxat* would have appeared in conjunction with *tapaγ*, cf., at any rate, Zieme (2015: 160, line 104): *qurbana tapınur*—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: [*tɒʕæt] > [*tɑɢɑt] > [tɑχɑt]. On the other hand, a hypothetical lexeme **taγat* could well have been written ‹t_τꞁʕꞁt_τ›, but hardly ‹t_τꞁxꞁt_τ› (cf. §§ 1.1.5 and 1.4.3.2).

$g_{1(1)}$	$g_{1(2)}$	$g_{1(3)}$	$g_{2(1)}$	$g_{2(2)}$	$g_{2(3)}$	$g_{3(1)}$	$g_{3(2)}$	$g_{3(3)}$	$g_{4(1)}$	$g_{4(2)}$
↓	↓	↓	‹Ø›	↓	↓	‹Ø›	↓	↓	↓	↓
$[C_{1(1)}]$	$[V_{1(2)}]$	$[C_{1(3)}]$	$[C_{2(1)}]$	$[V_{2(2)}]$	$[C_{2(3)}]$	$[C_{3(1)}]$	$[V_{3(2)}]$	$[C_{3(3)}]$	$[C_{4(1)}]$	$[V_{4(2)}]$

The resulting TM finite tape is:

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

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

Again, OTR may start:

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

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

3.1.3

A third example: ‹ꞁkndδy› (133.4):

‹ꞁ›	‹k›	‹n›	‹d_δ›	‹y›
‹x_1›	‹x_2›	‹x_3›	‹x_4›	‹x_5›

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

‹∅›	‹ʾ›	‹k›
$\gamma_1\sigma_{1(1)}$	$\gamma_2\sigma_{1(2)}$	$\gamma_1\sigma_{1(3)}$

Then, according to (1b):

‹∅›	‹n›
$\gamma_4\sigma_{n\text{-}m(1)}$	$\gamma_1\sigma_{n\text{-}m(x)}$

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

‹∅›	‹∅›	‹n›
$\gamma_4\sigma_{n\text{-}m(1)}$	$\gamma_f\sigma_{n\text{-}m(2)}$	$\gamma_1\sigma_{n\text{-}m(3)}$

Eventually,

‹∅›	‹ʾ›	‹k›	‹∅›	‹∅›	n	‹d_δ›	‹y›	‹∅›

from the resulting TM finite tape is:

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

OTR's output: *ȧkänd(*x*)* [ɐkɛndi].

3.1.4.

As a final example, let us examine the controversial instance ‹xwšnwt$_\theta$› (126.2), cf. § 1.1.9. From the initial string

‹x›	‹w›	‹š›	‹n›	‹w›	‹t_θ›
‹x_1›	‹x_2›	‹x_3›	‹x_4›	‹x_5›	‹x_6›

we obtain the following TM tape:

‹x›	‹w›	‹š›	‹∅›	‹∅›	‹n›	‹∅›	‹w›	‹t_θ›
$\gamma_2\sigma_{1(1)}$	$\gamma_a\sigma_{1(2)}$	$\gamma_1\sigma_{1(3)}$	$\gamma_4\sigma_{n\text{-}m(1)}$	$\gamma_f\sigma_{n\text{-}m(2)}$	$\gamma_1\sigma_{n\text{-}m(3)}$	$\gamma_4\sigma_{n(1)}$	$\gamma_a\sigma_{n(2)}$	$\gamma_1\sigma_{n(3)}$

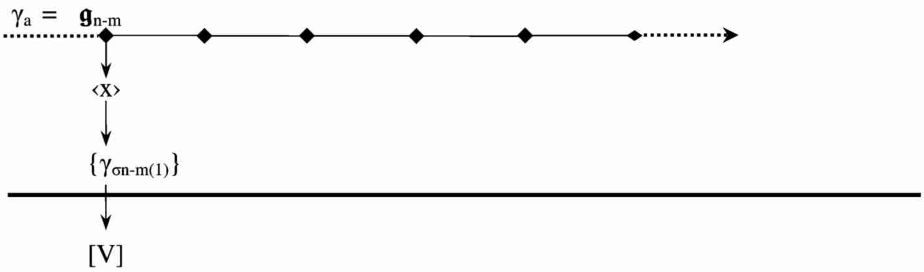
OTR's output: *q(*x*)šån(*x*)t*.

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 $(\forall\mathfrak{g}_{n\text{-}m}) \longrightarrow (\exists!\varphi_{n\text{-}m} \mid \mathfrak{g}_{n\text{-}m} \vDash \varphi_{n\text{-}m})$, i.e. there is exactly one particular linear relationship that turns out to be satisfied for each mother node (17).

This means that iff $(\varphi_{n\text{-}m} = \varphi_{n\text{-}p}) \longrightarrow (\mathfrak{g}_{n\text{-}m} = \mathfrak{g}_{n\text{-}p})$.

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.

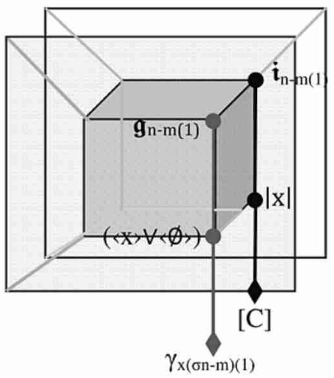
4.1. A brief excursus: *mod-b*

In the following, brief excursus, we will employ the function $\varphi_{n\text{-}m}$, 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 $\gamma_x\sigma_{n\text{-}m(2\pm1)}$ is a mere positional label which encapsulates some input-information about the "nature" of the correlated $\mathfrak{g}$-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 $\mathbf{i}_{n\text{-}m(1)}$, the "real-world" counterpart of $\mathbf{g}_{n\text{-}m(2)}$, as the neural state correlated with a specific graphic gesture.

Thus, since we empirically observe the following:

$$(\forall|x|\,|\ \mathbf{i}_{n\text{-}m(1)} \lhd |x|)\cdot((s|x| \lhd [C]) \rightarrow (\mathbf{g}_{n\text{-}m(2)} \lhd \langle\emptyset\rangle))$$

"for every $\mathbf{i}_{n\text{-}m(1)}$-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' $\mathbf{g}_{n\text{-}m(1)}$-object ($\mathbf{g}_{n\text{-}m(2)}$) is represented by ∅ in surface representation";

$$((\forall\langle x\rangle|\ \mathbf{i}_{n\text{-}m(3)} \lhd |x|)\cdot(s|x| \lhd [V] \vee s|x| \lhd [C][+\text{sonorant}, +\text{nasal}, -\text{labial}])$$
$$\longrightarrow (\mathbf{g}_{n\text{-}m+1(1)} \lhd \langle\emptyset\rangle))$$

"for every $\mathbf{i}_{n\text{-}m(3)}$-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 $\mathbf{g}_{n\text{-}m(3)}$-object (*viz.* $\mathbf{g}_{n\text{-}m+1(1)}$, which dominates the onset of syllable $\sigma_{n\text{-}m+1}$) is represented by ∅ in surface representation";

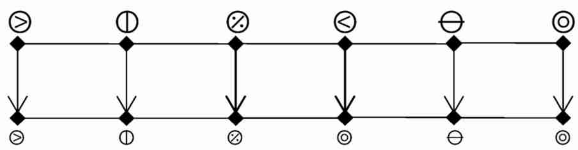
$$((\forall|x|\,|\ \mathbf{i}_{n(2)} \lhd |x|)(\nexists s|x|) \longrightarrow (\mathbf{g}_{n(3)} \lhd \langle\emptyset\rangle))$$

"for every $\mathbf{i}_{n(2)}$-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 $\mathfrak{g}_{n(2)}$-object ($\mathfrak{g}_{n(3)}$) is represented by $\emptyset$ in surface representation";

we can generalize by saying that, when in the world of concrete objects, a certain $\mathfrak{i}_{n\text{-}m(x)}$ satisfies the property $\varphi_{n\text{-}m(x)}$ ($\mathfrak{i}_{n\text{-}m(x)} \models \varphi_{n\text{-}m(x)}$), we infer (logical tunneling) that in our representational universe is $\mathfrak{g}_{n\text{-}m(2\pm1)} \models \varphi_{n\text{-}m(\emptyset)}$.

4.2

Let now consider a well-ordered set of totally arbitrary symbols, namely the alphabet $\Gamma_x = \{$⧁, ⦶, ⧀, ◎, ⊖, ⧁$\}$. 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.



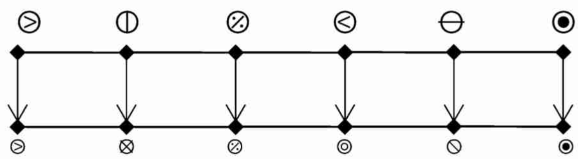
Since, according to (17), $(\forall \mathfrak{g}_{n\text{-}m}) \longrightarrow (\exists! \varphi_{n\text{-}m} | \mathfrak{g}_{n\text{-}m} \models \varphi_{n\text{-}m})$, the above mentioned graph implies the following, peculiar set of conditions:

$\Phi_x = \{$⧁$\models\varphi_{⧁}$, ⦶$\models\varphi_{⦶}$, ⊘$\models\varphi_{⊘}$, ⧀$\models\varphi_{⧀}$, ⊖$\models\varphi_{⊖}$, ◎$\models\varphi_{◎}\}$

Thus, $(\forall \Gamma_x) \longrightarrow (\Gamma_x \models \Phi_x)$

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 *‹p'rmk› should have been read as *bärmäk*, while a fictitious syntagm such as *‹p'rmx› should have been read as *barmåq* (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 *‹p'rmk› ⟶ *bärmäk*, the relationship *‹p'rmx› ⟶ *barmåq* may be, in its turn, represented as follows (19):



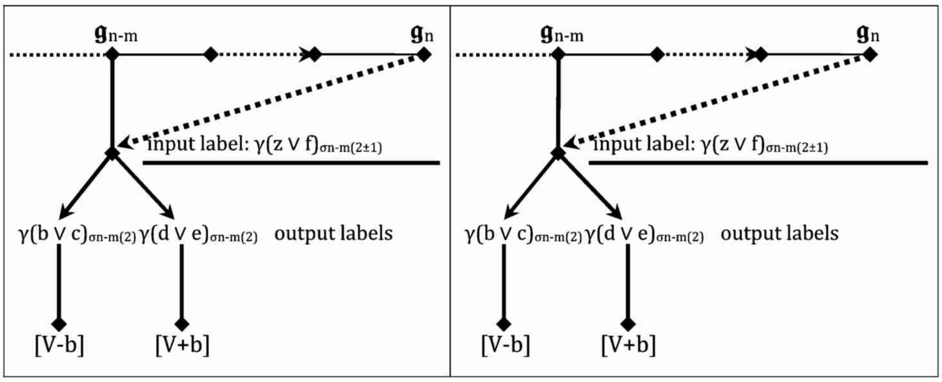
The problem arises that graph (19) disproves statement (17). We are forced to reformulate the representation enunciated in § 4.1 in the following terms:

Each $\mathfrak{g}_{n\text{-}m}$ 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 $\sigma_{n\text{-}m(2\pm1)}$) 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):

$\Phi_{x1} = \{\ (⧁\models\varphi_\gamma \wedge ⧁\models \varphi_\sigma \wedge ⧁\models \varphi_\alpha);$	$\Phi_{x2} = \{\ (⧁\models\varphi_\gamma \wedge ⧁\models \varphi_\sigma \wedge ⧁\models \varphi_\alpha);$
$(⦶\models\varphi_\gamma \wedge ⦶\models \varphi_\sigma \wedge ⦶\models \varphi_\alpha);$	$(⦶\models\varphi_\gamma \wedge ⦶\models \varphi_\sigma \wedge ⦶\models \varphi_\alpha);$
$(⊘\models\varphi_\gamma \wedge ⊘\models \varphi_\sigma \wedge ⊘\models \varphi_\alpha);$	$(⊘\models\varphi_\gamma \wedge ⊘\models \varphi_\sigma \wedge ⊘\models \varphi_\alpha);$
$(⧀\models\varphi_\gamma \wedge ⧀\models \varphi_\sigma \wedge ⧀\models \varphi_\alpha);$	$(⧀\models\varphi_\gamma \wedge ⧀\models \varphi_\sigma \wedge ⧀\models \varphi_\alpha);$
$(⊖\models\varphi_\gamma \wedge ⊖\models \varphi_\sigma \wedge ⊖\models \varphi_\alpha);$	$(⊖\models\varphi_\gamma \wedge ⊖\models \varphi_\sigma \wedge ⊖\models \varphi_\alpha);$
$(◎\models\varphi_\gamma \wedge ◎\models\varphi_\sigma \wedge ◎\models \varphi_\alpha)\ \}$	$(◉\models\varphi_\gamma \wedge ◉\models \varphi_\sigma \wedge ◉\models \varphi_\alpha)\ \}$

What we *actually* observe can be illustrated by the following graph:

$(⦶\models \varphi_\alpha)_{\Phi x1};\ (⊖\models \varphi_\alpha)_{\Phi x1}$ $(⦶\models \varphi_\alpha)_{\Phi x2};\ (⊖\models \varphi_\alpha)_{\Phi x2}$



The following statement captures some aspects of the above reformulated representation:

$$\text{iff}\quad (\exists\Phi_{x1})(|\Phi_{x1}| - |(◎\models\varphi_\gamma \wedge ◎\models \varphi_\sigma)| = |\Phi_{x2}| - |(◉\models\varphi_\gamma \wedge ◉\models \varphi_\sigma)|)$$

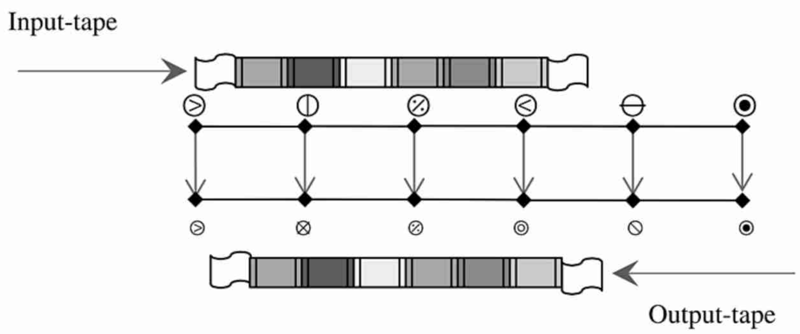
$$\longrightarrow\quad [\exists\varphi_\alpha{}^{\Phi x1}|\ ((⦶\models \varphi_\alpha) \in \Phi_{x1}))$$

$$\wedge\ ((\ominus \vDash \varphi_\alpha) \in \Phi_{x1}))]\cdot[(\text{⦶} \vDash \varphi_\alpha{}^{\Phi x1} \neq \text{⦶} \vDash \varphi_\alpha{}^{\Phi x2})$$
$$\wedge\ (\ominus \vDash \varphi_\alpha{}^{\Phi x1} \neq \ominus \vDash \varphi_\alpha{}^{\Phi x2})]\ (20)$$

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 ‹ʾ› and ‹Ø›, when occurring in the strings *‹pʾrmØk› (to be read as *bärmäk*) and *‹pʾrmØx› (to be read as *barmåq*), is the alternation of the graphemes ‹k› and ‹x› at the end of the sequence ‹pʾrmØ›.

Such an otherwise insightful observation does not capture the core meaning of statement (20), which pivots around the inequalities $\text{⦶} \vDash \varphi_\alpha{}^{\Phi x1} \neq \text{⦶} \vDash \varphi_\alpha{}^{\Phi x2}$, and $\ominus \vDash \varphi_\alpha{}^{\Phi x1} \neq \ominus \vDash \varphi_\alpha{}^{\Phi x2}$.

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˛m¸yšyn͡g˽pyr˽myn͡gw* (127.4): *bermiš* {-Iŋ} *bir meŋü.*

The following example is definitely much more insidious. Let us consider again some syntagmatic contexts in which the morpheme {+DA} occurs:

‹t$_\tau$ꞁxꞁt$_\tau$y$\widehat{ng}$t$_\theta$ꞁ› (127.2), cf. § 3.1.2;

‹sw˛β˛sꞁmyšynt$_\theta$ꞁ›:

‹pyr⌴ꞁṣyp⌴sw˛β˛sꞁmyšynt$_\theta$ꞁ⌴t$_\tau$yg$_\gamma$|yꞁl$\widehat{ng}$wz⌴pyrṣꞁn̊⌴swyyx⌴swβ̊[60]| ꞁyṣwrsꞁr› *bir ač* {-Ip} *suvsa* {-mIš} {+In} {+DA} *teg yalŋuz bir čan* (< Chinese 盞) *soyïḳ*[61] *suv ičür* {-sAr} '[…] when he is thirsty, (he) let (him) drink only a cup of cold water […]' (125.3–5).[62] Cf. ‹pyr⌴ṣꞁn̊⌴swyyx|swβ̊⌴ꞁyṣwrmyš› (128.3–4).

‹pyt$_\tau$yg$_\gamma$ꝫt$_\theta$ꞁ› bitig {+DA}, cf. § 1.4.1:

stem	σ_{n-2}	σ_{n-1}	$\sigma_{n\ [Kaisse]}$
*√‹t$_\tau$ꞁxꞁt$_\tau$›		‹y$\widehat{ng}$›	‹**t**$_\theta$ꞁ›
√‹swβsꞁ›	‹myš›	‹yn›	‹**t**$_\theta$ꞁ›
√‹pyt$_\tau$yg$_\gamma$›			‹**t**$_\theta$ꞁ›

In paragraph 3.1.2 we considered the string ‹t$_\tau$ꞁxꞁt$_\tau$y$\widehat{ng}$t$_\theta$ꞁ› to be calculable by our Finite State Transducer OTR since we assumed that, for every occurrence of the grapheme ‹ꞁ› 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 . soyïq suv* […]; Zieme (2013: 101): *bir čan soyïx suv*; Zieme (2015: 163, l. 016).

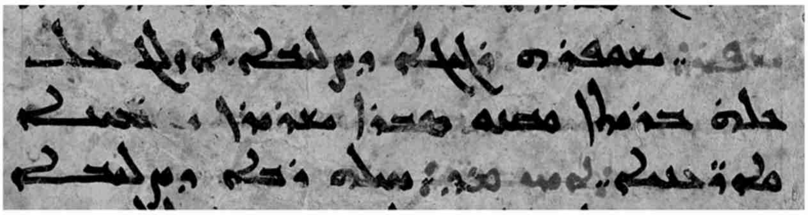
61 Starostin, Dybo & Mudrak (2003: 1336): OTurk. *soɣïq*.

62 Zieme (2015: 154, ll. 15–17; 161, note 561).

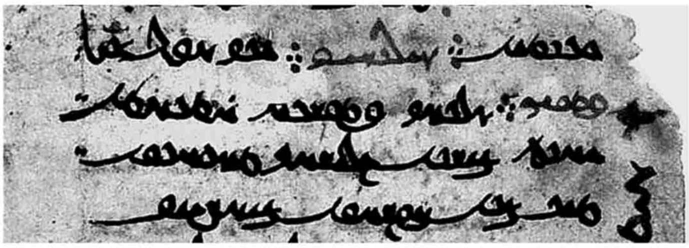
Dickens, Mark 2013. The importance of the Psalter at Turfan. In: Li & Winkler (2013: 357–380).
Erdal, Marcel 2004. *An old Turkic grammar*. Leiden: Brill.
Eraslan, Kemal 2012. *Eski Uygur Türkçesi grameri*. (Atatürk Kültür, Dil ve Tarih Yüksek Kurumu Türk Dil Kurumu Yayınları 1047.) Ankara: Türk Dil 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: Farhangan 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. 2014^2. *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. *Manuscrits 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, Rajeen & Ullmann, Jeffrey D. 2001^2. 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-ġaršūnī 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 & Chimeddhorji 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

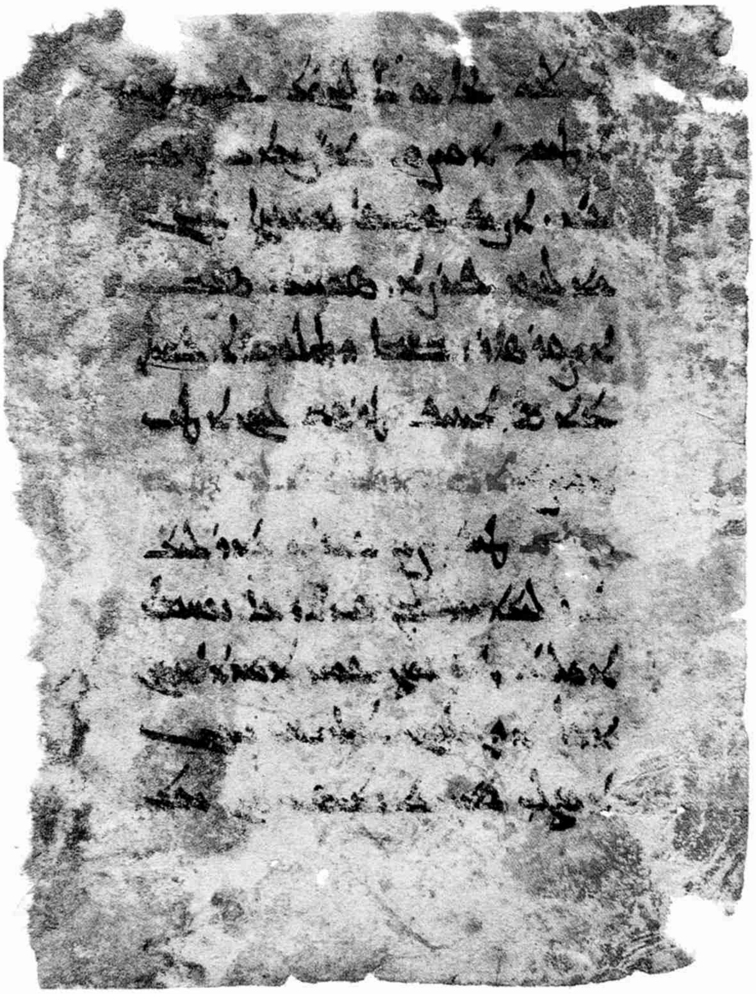


Specimen of a Syriac ms in Syriac script: Ms Berlin, Museum für Asiatische Kunst, MIK III 111v = T II B 37, cf. Hunter & Dickens (2014: 354-355, no. 394).



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 Bulayïq; 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".