

Interpreting differences among transcripts

C. M. Sperberg-McQueen, Black Mesa Technologies LLC

Claus Huitfeldt, Universitetet i Bergen

DH 2018, México D.F., 28 June 2018

Slides, handout:

<http://mlcd.blackmesatech.com/mlcd/2018/Talks/MexicoDF-201806/index.{es,en}.{xml,xhtml}>

<http://mlcd.blackmesatech.com/mlcd/2018/Talks/MexicoDF-201806/{handout,volante}.pdf>

1 Context

- What is the *logical structure* of transcription? (*Not*: how should we transcribe?)
- What do transcripts tell us? (*Not*: are they telling the truth?)
- Can we provide a rational reconstruction of the concept of transcription?
- Can an account of transcription be integrated into an account of the meaning of markup?
- For example: how does one say, in formal logic, “This `sec` element transcribes a block of text in the manuscript identified in the MECS-WIT header”?

2 Key concepts

- type, token, marks; document, text
- atomic and compound types and tokens
- type inventory, type system
- reading, inferred reading
- transcription policy

3 Notation

- **E** = the exemplar
- **T** = the transcript
- **A**, **B**, **C** = specific transcripts
- [**T**] = unpublished transcript
- ***T** = transcript constructed for this talk

4 Example: Ludwig Wittgenstein

[A]	[B]	C
munonyqi	wunouyqi	muuvnyzi

At most one of these can be correct.

Moral: a transcript reflects a *reading* of *E*.

Moral: a *reading* maps tokens to types (inter alia).

5 Example: Tumba Edithae, Magdeburg

[A] INPULSU HOC TEMPLUM AB OTHONE MAGNO DIVO
CAESARE CONGIUGE FUNDATUM EST OBIIT ANNO
CHRISTI

DCCCC XLVII°

DIVE REGINE ROMANORUM EDIT ANGLIE REGIS ECMVNDI
FILIE HIC OSSA CONDUNTUR CUIUS RELIGIOSI

AMORIS

B DIVE · REGINE · RO[MA]NOR[UM] · EDIT · ANGLIE ·
REGIS · ECMVNDI · FILIE · HIC · OSSA · CO[N]DVNTVR ·
CVIVS · RELIGIOSI

AMORIS

INPVLVS · HOC · TE[M]PLVM · AB · OTHONE · MAGNO
· DIVO · CAESARE · CONIVGE · FV[N]DATV[M] · EST ·
OBIIT · AN[N]O · CHRISTI
DCCCCXLVII

Moral: Texts are not (just) sets of character tokens; they have structure.

Moral: Tokens and types are more than characters: they are also words, sentences, paragraphs, ...

6 Example: Jane Addams

*A	*B	C
altho	altho[<i>ugh</i>]	[although]

*In *B, bracketed italics = editorial addition.*

In C, bracketed words = editorial intervention.

Moral: Some tokens in *T* do not claim to have exemplars in *E*.

Moral: Some tokens in *T* have exemplars in *E*, but not their constituents.

Moral: *T* can tell us a word is in *E* without telling us how it's spelled.

7 Example: Sor Juana Ines de la Cruz

A Que aun vista perspicaz en vano apura

***B** Que aun vifta perspicàz en vano apura

Moral: A reading of a document depends on the type system (= set of type inventories) used.

Moral: When different type inventories are used, different type mappings don't entail contradiction.

8 Example: Frida Kahlo

*A	*B
nos juntaremos ya <i>para siempre</i>	nos juntaremos ya <u>para siempre</u>

*In *A, underscoring in E is rendered with italics.*

*In *B, underscoring in E is rendered with underscoring.*

Moral: Transcribers sometimes map one type in *E* to another in *T*. Such type equivalences are not part of the reading of *E*, only part of the transcription policy of *T*.

9 Conclusions

- A reading of a token maps it to a type (Wittgenstein).
- A reading of a document identifies tokens and maps them to types (Tumba Edithae).
- Different readings may use different sets of types (Sor Juana).
- A transcription policy
 - distinguishes normal and special tokens in *E* (Tumba Edithae);
 - distinguishes normal and special tokens in *T* (Addams);
 - defines some equivalences between types (Kahlo).

- From any T , we can reconstruct a reading of E .

10 Wrapping it up

- Transcripts provide information about exemplars partly by reinstantiation, partly by description.
- Reinstantiation is intrinsically digital / based on reproduction of discrete symbols.
Contrast facsimile, which is intrinsically analog.
- Reinstantiation of types is necessarily relative to a given type system, a given rule of selection, and a given reading of E .
- A meaningful statement that “ T transcribes E ” must presuppose a transcription policy and *readings* of both E and T .

11 Type inventories and systems

A *type inventory* I is a set of types.

A *type system* P is a set of type inventories.

12 Readings

A reading R of a token k with respect to a type inventory I is a tuple $R = (k, I, p)$ where:

- k is the token being read.
- I is a type inventory.
- p is in I .

A reading R of a document D is a tuple (D, K, P, M) where:

- D is the document being read.
- K is a set of tokens identified as being in D .
- P is a type system.
- M (the mapping of R) is a set of triples (k, I, p) , where
 - k is in K .
 - I is in P .
 - p is in I .
 - No two triples have the same k and I .
 - There is at least one triple for every k in K .

13 Transcription policies

A transcription policy π is a triple (SE, ST, Q) , where:

- SE is a unary predicate. $SE(k)$ is true iff k is a “special token” in E (not to be transcribed).
- ST is a unary predicate. $ST(k)$ is true iff k is a “special token” in T (lacking exemplar in E).

- Q is a set of pairs (p_E, p_T) ; for purposes of π , type p_E occurring in E will be reinstantiated in T using type p_T . (E.g. underline, italics.)

14 Inferred reading of E

Any T reflects a reading of E and allows us to reconstruct it at least in part. A *reconstruction* of a reading of E given transcript T is a tuple $RR = (E, T, \pi_T, R_T, R_{(E,T)})$, where

- E is the exemplar
- T is the transcript
- $\pi = (SE, ST, Q)$ is the transcription policy reflected in T
- R_T is a reading of T
- $R_{(E,T)} = (E, K_{(E,T)}, P, M_{(E,T)})$ is a reading of E

In practice R_T should be compatible with π .

$R_{(E,T)}$ is the reconstructed reading of E .

15 One last thing ...

To say “This <sec> element transcribes a block of text in the manuscript identified in the MECS-Wit header”, one can write (oversimplifying slightly)

$$\begin{aligned}
 & (\exists b : \text{Token})(\forall d : \text{Document})(\text{identifies}(/ \text{doc/catno}, d) \Rightarrow \\
 & \quad (\text{token-in-document}(b, d) \\
 & \quad \wedge (\exists R_D : \text{reading}) \\
 & \quad \quad (\exists K_D : \text{set Tokens}) \\
 & \quad \quad (\exists P_D : \text{typesystem}) \\
 & \quad \quad (\exists M_D : \text{tt-mapping}) \\
 & \quad \quad \quad (R_D = (d, K_D, P_D, M_D) \\
 & \quad \wedge (\exists R_T : \text{reading}) \\
 & \quad \quad (\exists K_T : \text{set Tokens}) \\
 & \quad \quad (\exists P_T : \text{typesystem}) \\
 & \quad \quad (\exists M_T : \text{tt-mapping}) \\
 & \quad \quad \quad (R_T = (/ , K_T, P_T, M_T) \\
 & \quad \wedge (\exists \pi : \text{transcription-policy}) \\
 & \quad \quad (\exists S_E : \text{unary-predicate}) \\
 & \quad \quad (\exists S_T : \text{unary-predicate}) \\
 & \quad \quad (\exists Q : \text{type-type-function}) \\
 & \quad \quad \quad (\pi = (S_E, S_T, Q) \\
 & \quad \wedge (M_T(.) = M_D(b) \vee M_T(.) = Q(M_D(b)))))))))
 \end{aligned}$$