Homer Multitext project: guide for editors

This document describes the HMT project's use of TEI XML in the summer of 2017.

Tier 1: Transcription

The transcription tier handles issues of paleographic status. Is a reading clear? How do we transcribe the specific orthographic choices of the manuscript? What belongs in the text edition and what is recorded elsewhere.

What doesn't belong in a text edition?

Text editions of the *Iliad* or scholia texts should not contain anything in the following categories:
1. critical signs: located next to the Iliadic lines in the Venetus A
2. connecting signs: sometimes written next to scholia and corresponding main text
3. metacomments: numerical or other symbolic notation that further explains the text (e.g. reordered lines)

Orthography

Regarding the following issues of orthography, we have very specific conventions for how to display them in Unicode. These examples are unambiguous and therefore do require mark up. We do, however, want to be consistent in how we handle them in our editions.

accentuation and breathings
: We record all accents and breathings as given in the manuscript (not "corrected" to modern usage). This category includes diaeresis, and markings for short or long vowel quantities (breve and macron).

iota subscripts
: Iotas are consistently written adscript in most of our manuscripts: we write them subscript where that is usual in modern practice without implying that the manuscript does so.
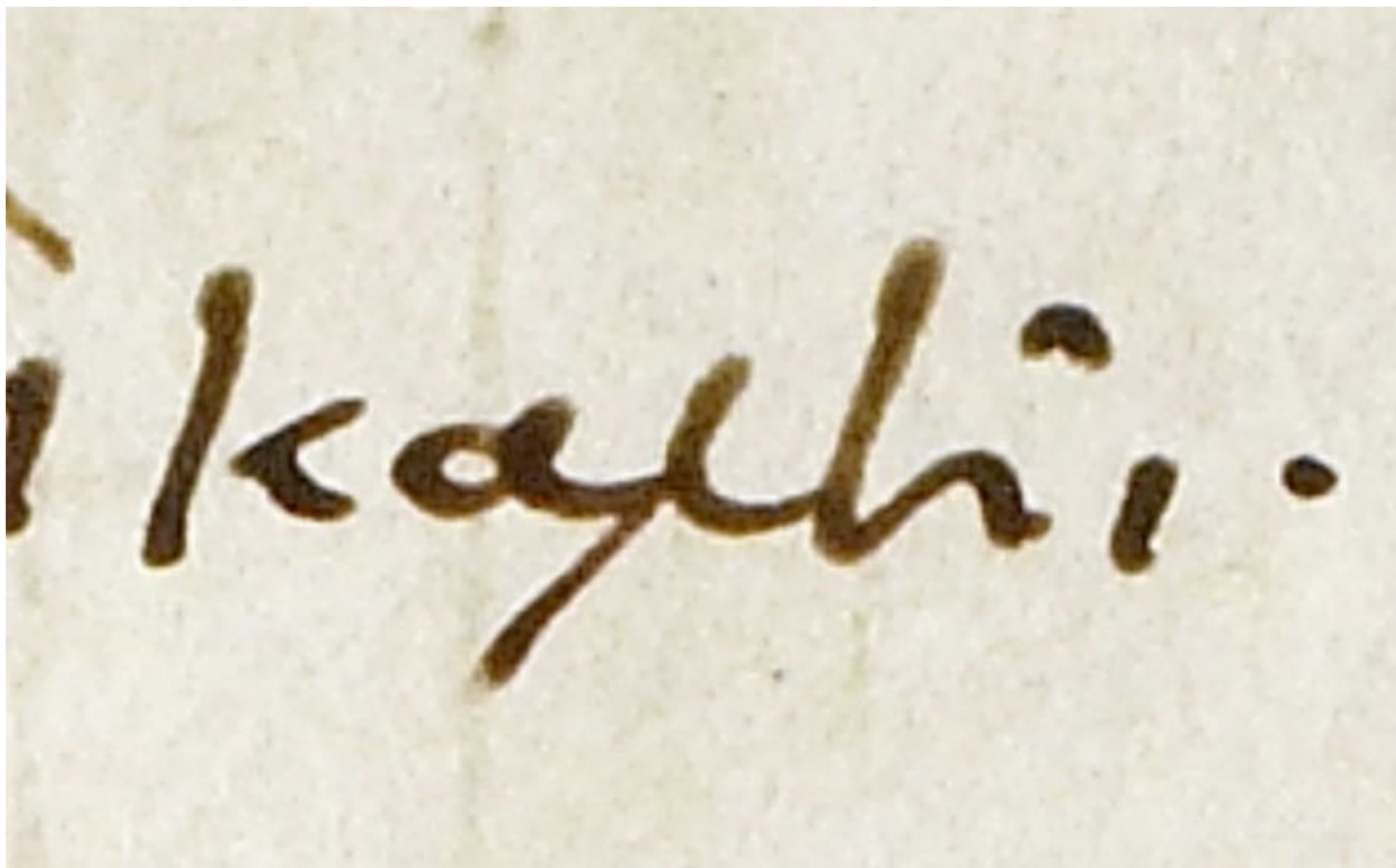
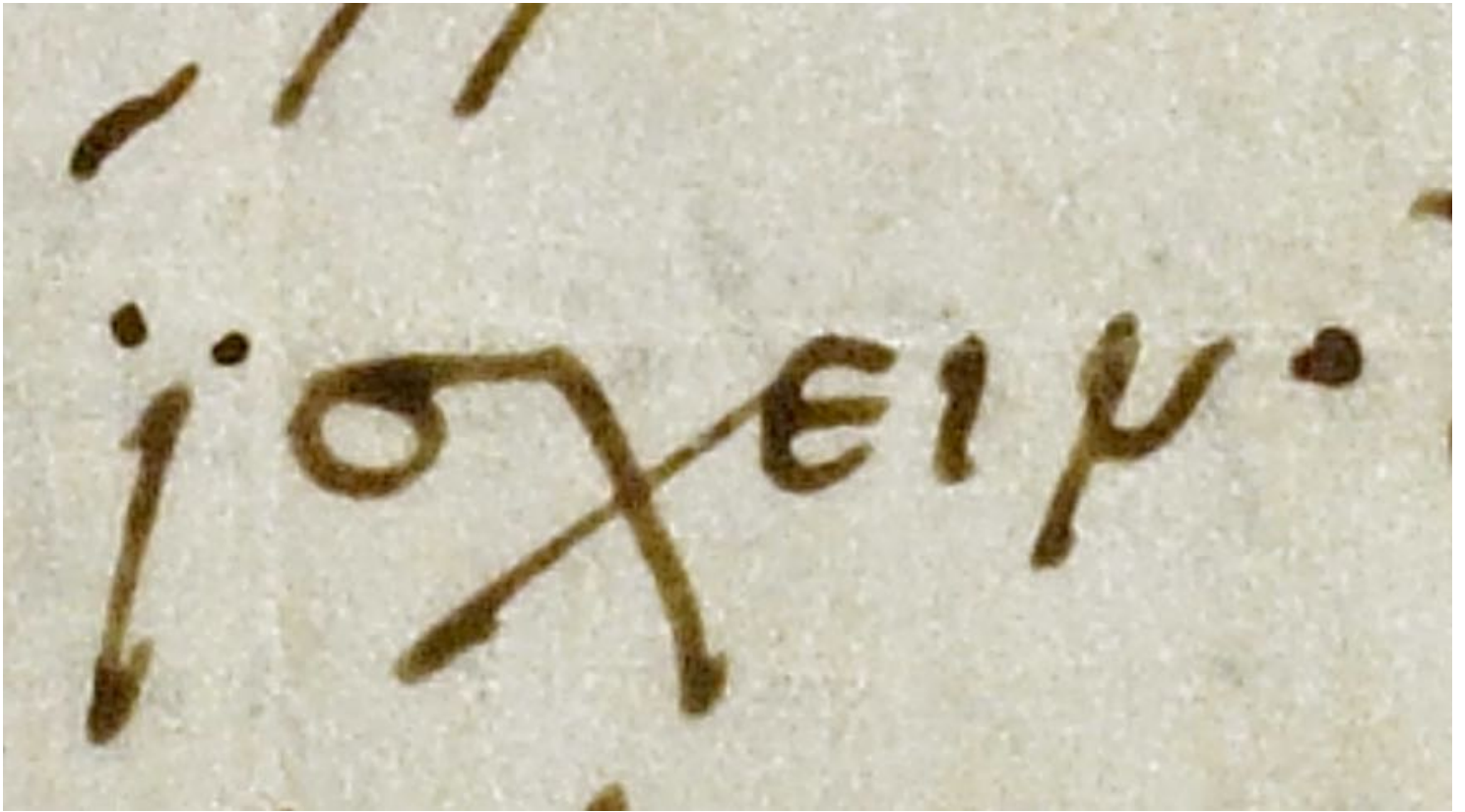Example: καλῇ (urn:cts:greekLit:tlg0012.tlg001.msA:1.604)

ligatures and variant letter forms
: Our transcription does not note ligatures or variant letter forms: readers interested in these visual features of the manuscript should consult the associated high-resolution images. We treat regular abbreviations for morphological endings, for prepositions, and for the conjunction καί as variant letter forms, and silently expand these in the transcriptions. See more on variant letter forms under abbreviations.

Diacritical Marks

Most Greek keyboards are not able to handle more than two diacritical marks, so in instances where you need and accent, breathing, and a diacritical mark, use combining characters which are typically found in special character menus in most basic text editors.

diaereses on vowels with breathing and/or accent should use Unicode "Combining Diaeresis" character.

Example: ἴσχειν (urn:cts:greekLit:tlg0012.tlg001.msA:9.352)

**macra** should use Unicode "Combining Macron" character.
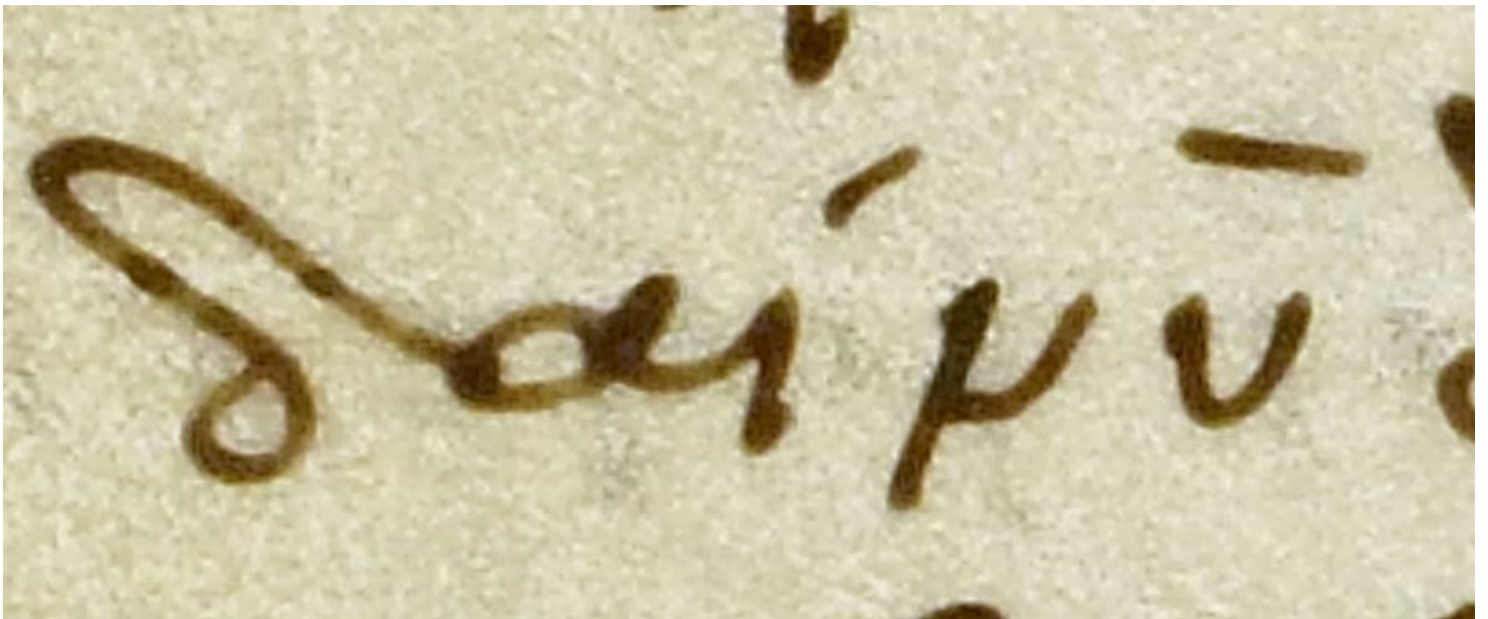
Example: δαίνῡ {urn:cts:greekLit:tlg0012.tlg001.msA:9.70)

brevia should use Unicode "Combining Breve" character

urn:cite:hmt:edguide:brevefloat

elisions

Of the many possible Unicode characters for elision, use only Unicode straight single quote. If you use Sophokeys to enter Greek, note that this is **not** the character it generates automatically when entering Greek.
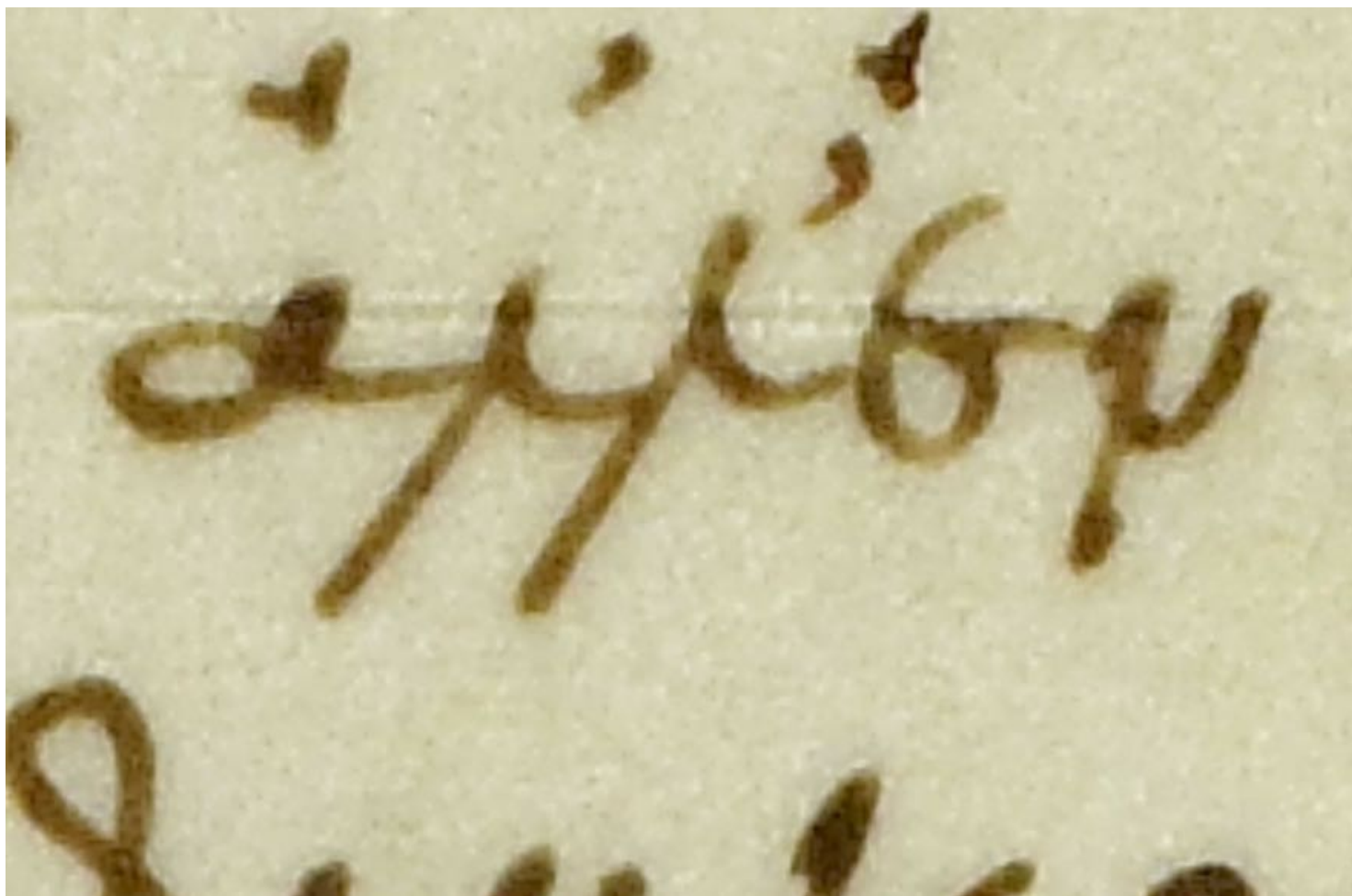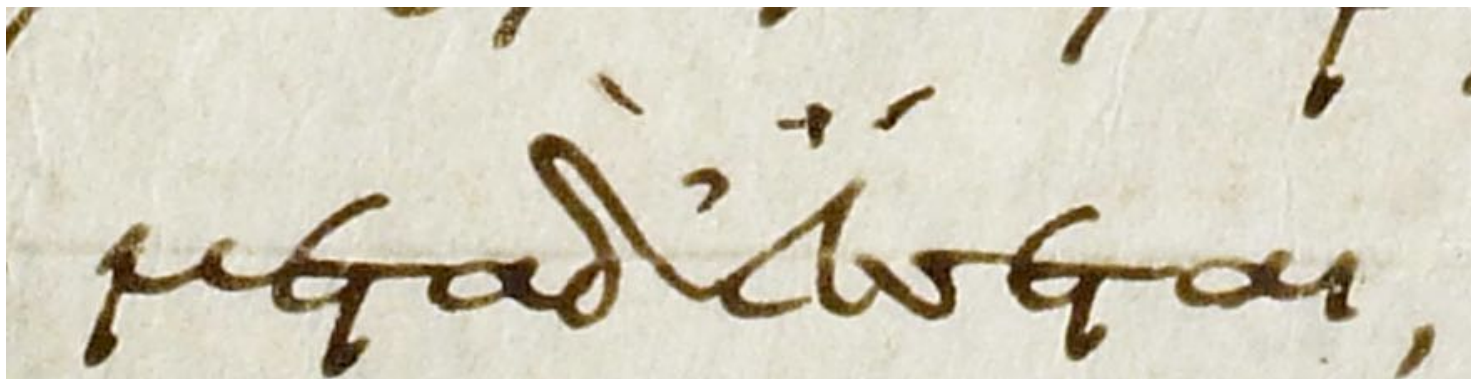

Example: ἄλλ' ἐν (urn:cts:greekLit:tlg0012.tlg001.msA:9.235)

"floating" grave accents

Floating grave accents in the manuscript can appear on an elided syllable.
"Float" a grave accent character between the last letter of the word and the elision mark.

Example: μετα δ`` ἔσσεται
(urn:cts:greekLit:tlg0012.tlg001.msA:9.131)

Paleographic Status

While readers always take many kinds of contextual information into consideration when reading a text, HMT editors should distinguish three levels of legibility based solely on the paleographic clarity of the reading.

1. *clear*. The letter is unambiguously legible, based on paleographic considerations alone. It may be incompletely preserved, but the visible remains cannot be read as any other character.
2. *unclear*. Part of the letter is visible, but taken by itself cannot be unambiguously read. That is **paleographically** ambiguous. You might be able to make an educated guess, but that is not a diplomatic edition.
3. *missing*. No trace of the letter remains, but it is clear from the context that one or more letters were originally present. This occurs in instances where there is damage to the manuscript such as a hole or extreme fading, even in specialized lighting.

Clear readings

Enter the text with no additional markup

Unclear readings

Use TEI `unclear` element. If the `unclear` element breaks up a word token, wrap the entire work in a TEI `w` element. Note that `w` is only necessary if the word is broken up. If a whole word is unclear, `w` is not necessary and should not be included.
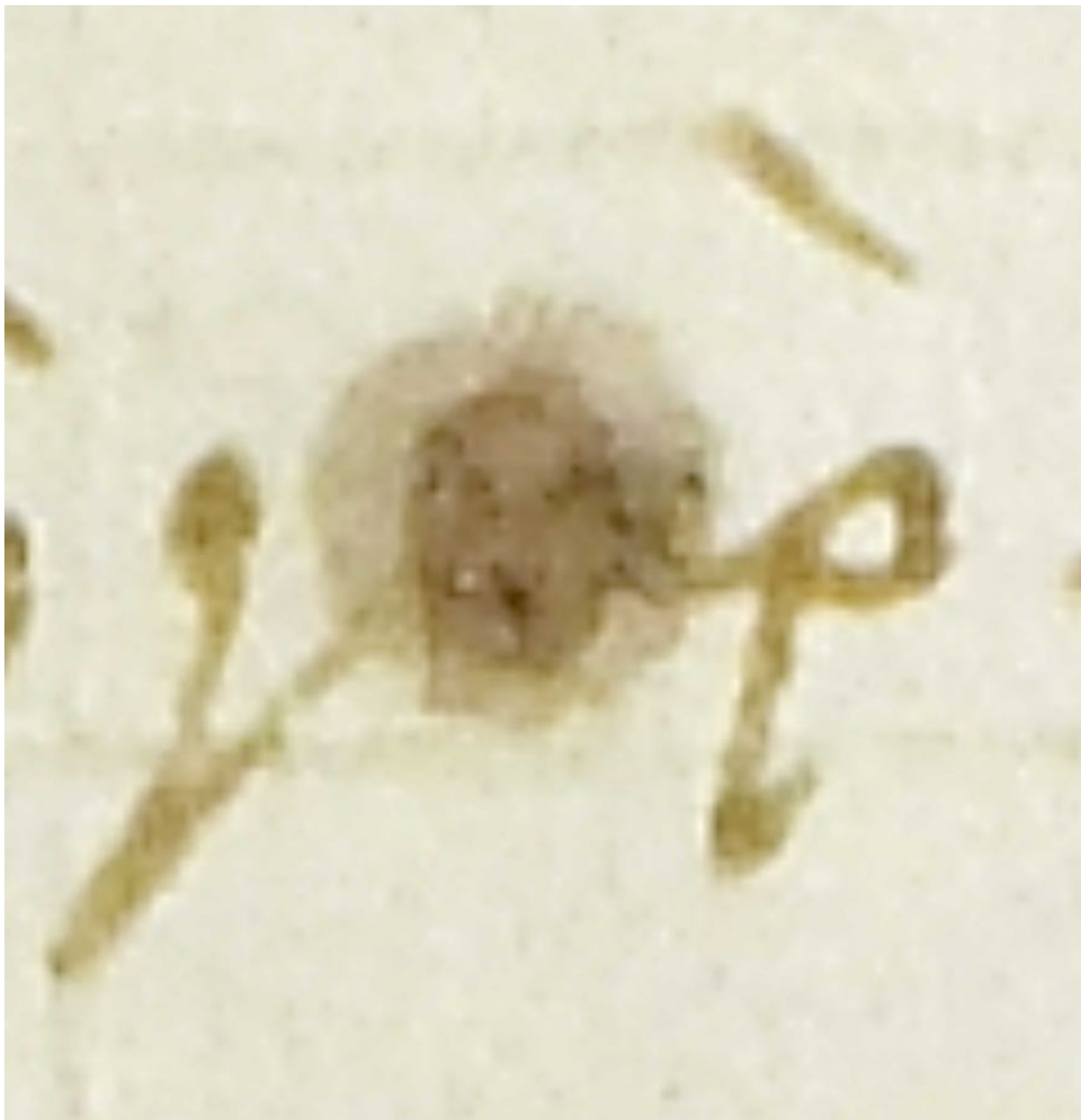
Example:

<w>γ<unclear>ὰ</unclear>ρ</w>

(urn:cts:greekLit:tlg5026.msA.hmt:8.132)

Missing

Use TEI `gap` element. Use `@unit` attribute for the type of the gap (for example, "letters" or "words"), and `@extent` to indicate how many units are likely included in the gap.

Example:

```
<gap unit="letters" extent="3"/>
```

Tier 2: Tokenization

The tokenization tier disambiguates our edition at the word level. Typically a machine can read tokens by splitting at white space. However, there are a few exceptions that require mark up.
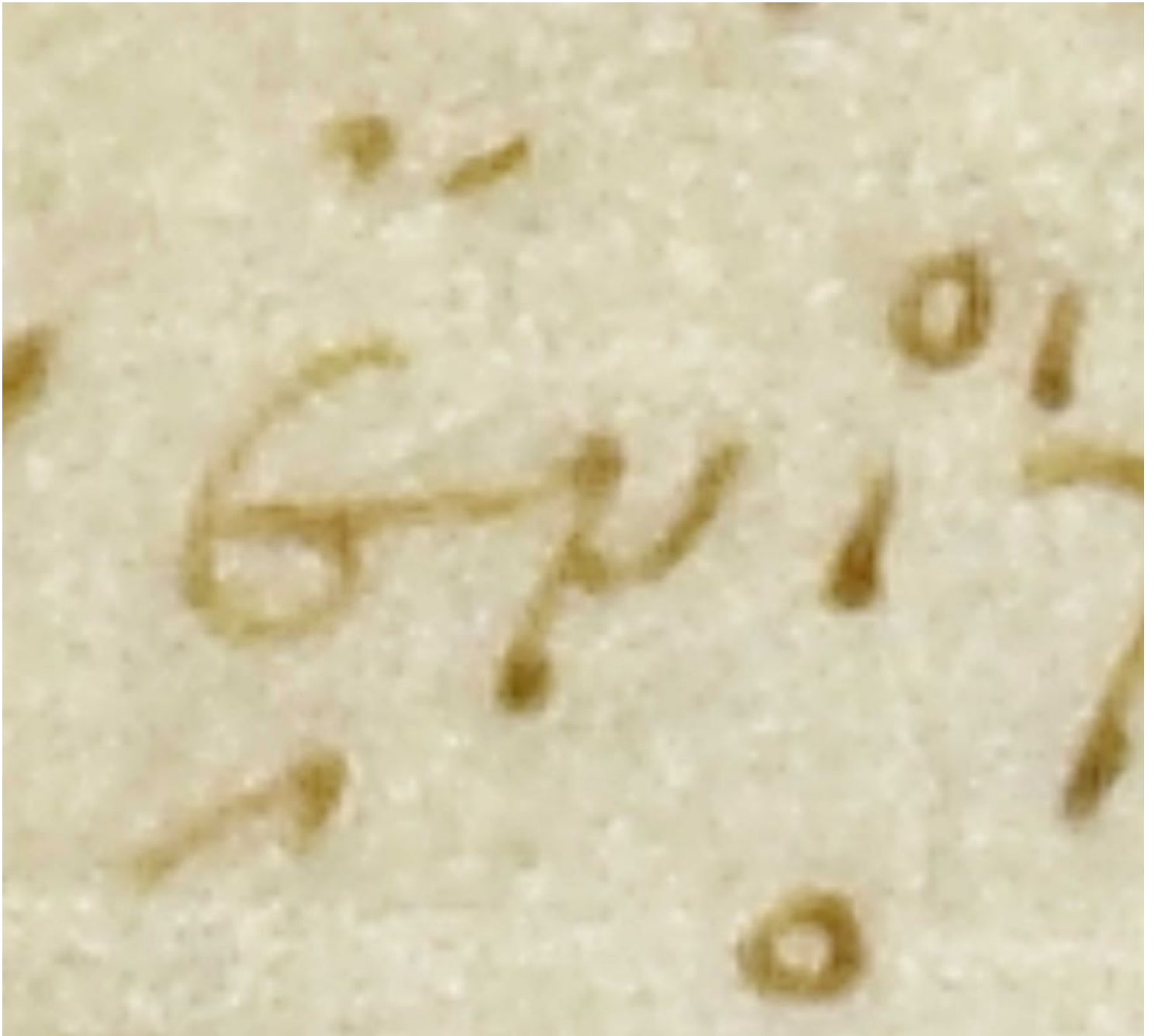
Word

There are a few cases where we cannot rely on whitespace to split our tokens because of other TEI mark up.

paleographic clarity

As seen on the transcription level, `gap` and `unclear` can sometimes split words. In these situations the whole word needs to be explicitly wrap in the `w` element so that our edition remains machine-readable.

scribal additions

Use TEI `add` element, including `@place` attribute with value `supralinear`
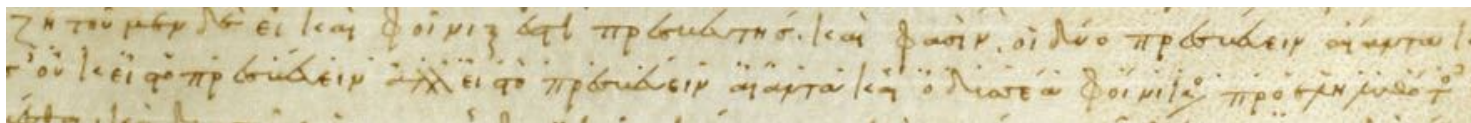
Example:

`<w>ἔνι<add place="supralinear">οι</add></w>`

(urn:cts:greekLit:tlg5026.msA.hmt:4.9)

In cases where the addition does not split a word, you would not need to use `w`

Scribal Deletions

Use TEI `del` when the scribe has either crossed out content, erased, or marked it with "deletion dots"
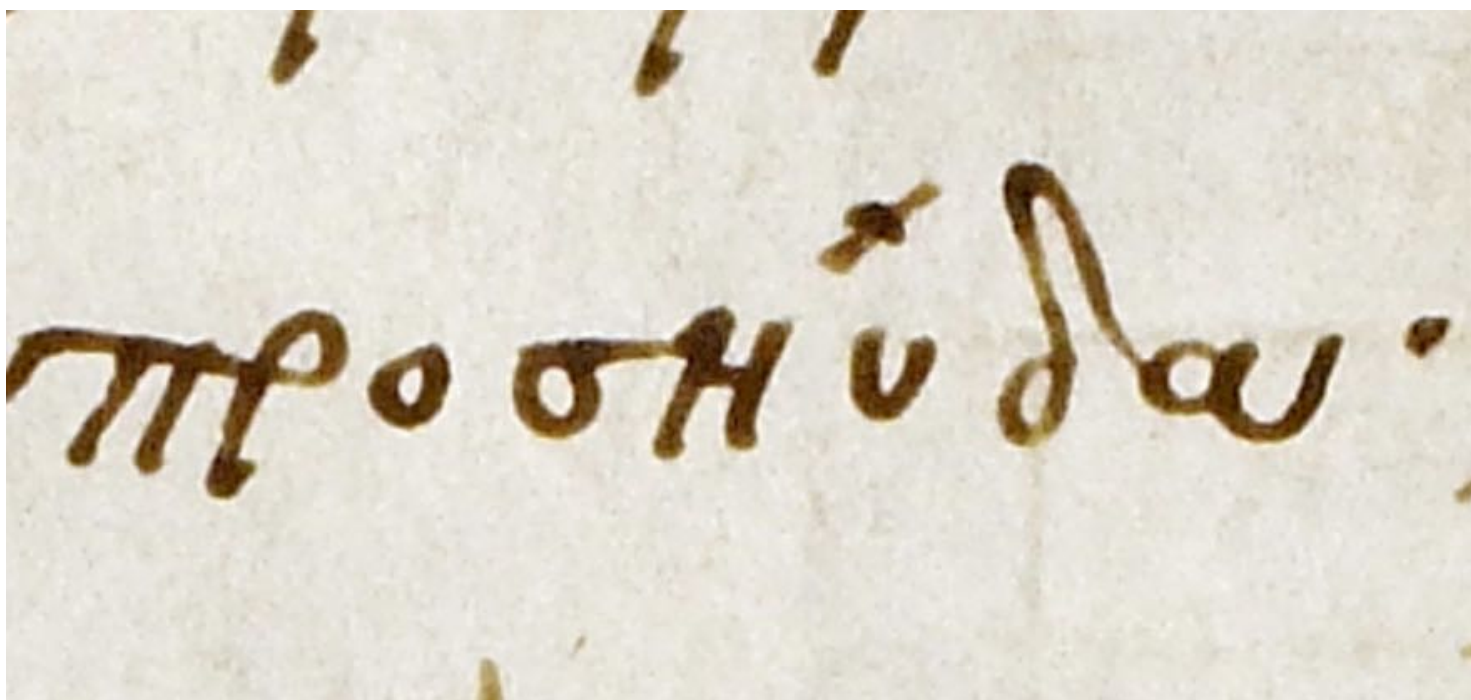
Example:

~~οὐκ εἰς το *πρεσβεύειν* ἀλλ εἰς τὸ *πρεσβεύειν* Αἴαντας καὶ Ὀδυσσέα Φοίνικος *προέληλυθότος*~~ (urn:cts:greekLit:tlg5026.msA.hmt:9.193)

Scribal Corrections

Use TEI `sic/corr` pair for corrections when the wants to correct a reading that is unintelligible in the text; group the pair in a TEI `choice` element.

Example: accent corrected by the original scribe

<choice><sic>*προσηῦδα*</sic><corr>*προσηύδα*</corr></choice>

(urn:cts:greekLit:tlg0012.tlg001.msA:4.337)

Please note that this mark up is *only* used by corrections we believe were made by the scribe. That means you cannot offer your own corrections and the scribe could be wrong by our orthographic standards. This is a diplomatic editions and we do not want to make our own judges on grammatical correctness at this stage of editing.

Scribal Multiforms

Use the TEI `orig`/`reg` pair to identify the reading in the main text and the alternative reading; group them with TEI `choice`. The scribe sometimes writes only the letters that are to be changed to create the alternate reading. We put the fully expanded word in the `reg` element.

Example:

`<choice><orig>μεθειέμεν</orig><reg>μεθιέμεν</reg></choice>`

(urn:cts:greekLit:tlg0012.tlg001.msA:4.351)

Alternative readings are typically strings of characters or even single letters. You will use the whole word like in the example above.

Abbreviations

Abbreviations without mark up

Regular terminating syllables

Regular terminating syllables (often morphological endings, but not always appearing at the end of a word) are treated as a variant letter form, and are not specially marked in any way. See the paleography guide if you are unfamiliar with the common abbreviations.
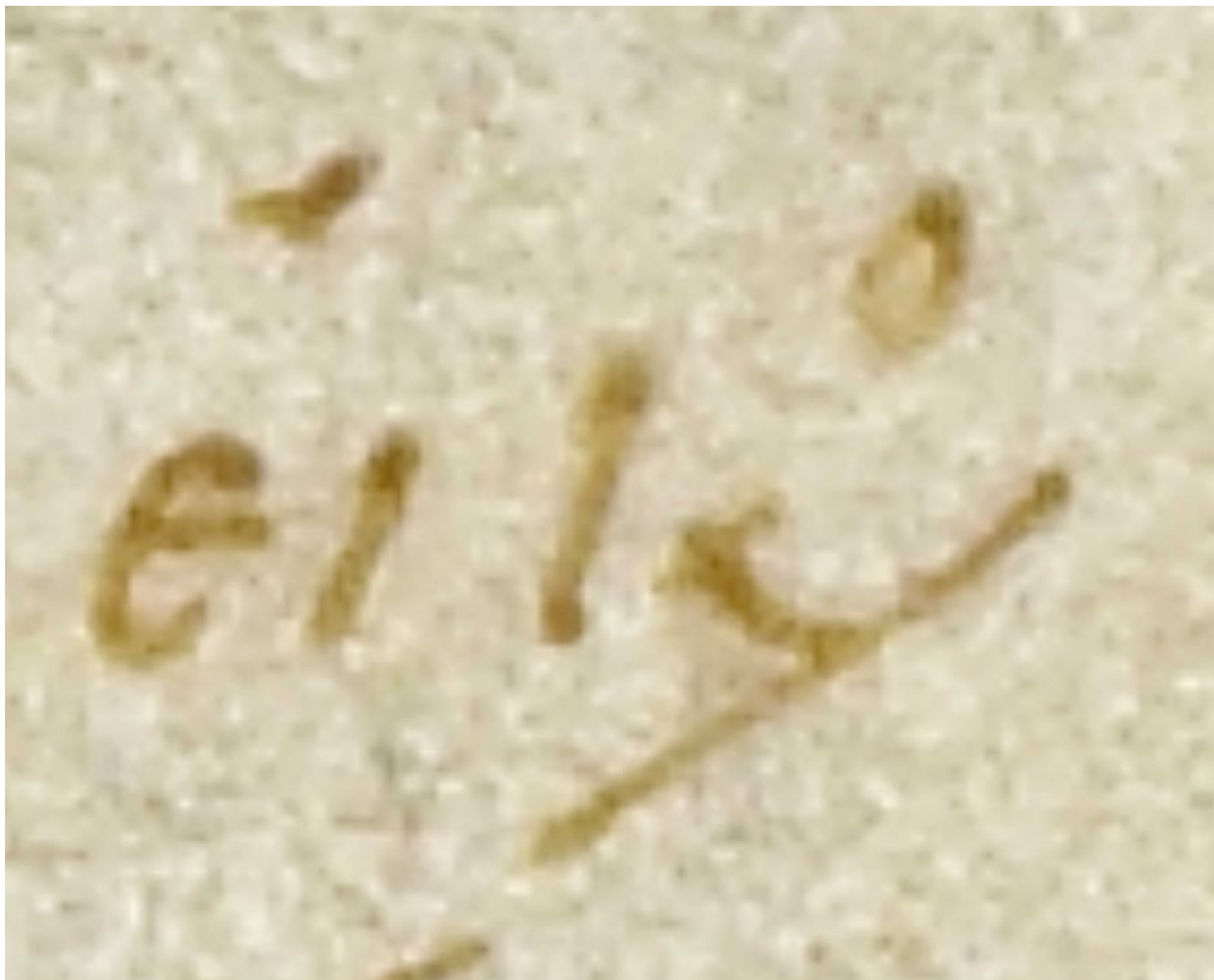
Example:

εἰκος, where a raised omicron is a brief writing for the morphological ending -ος
(urn:cts:greekLit:tlg5026.msA.hmt:4.478)

Symbol Abbreviations

Regular abbreviated forms of the whole words (e.g. καί, ὅτι, δέ, γάρ, διά, etc.) are treated as variant letter forms, and are not specially marked in any way. The rule of thumb for determining if it is a symbol abbreviation is if the symbol cannot be completely broken down into distinquishable letter forms.

Examples:

ς

(urn:cts:greekLit:tlg5026.msA.hmt:4.437)

(urn:cts:greekLit:tlg5026.msA.hmt:4.8)
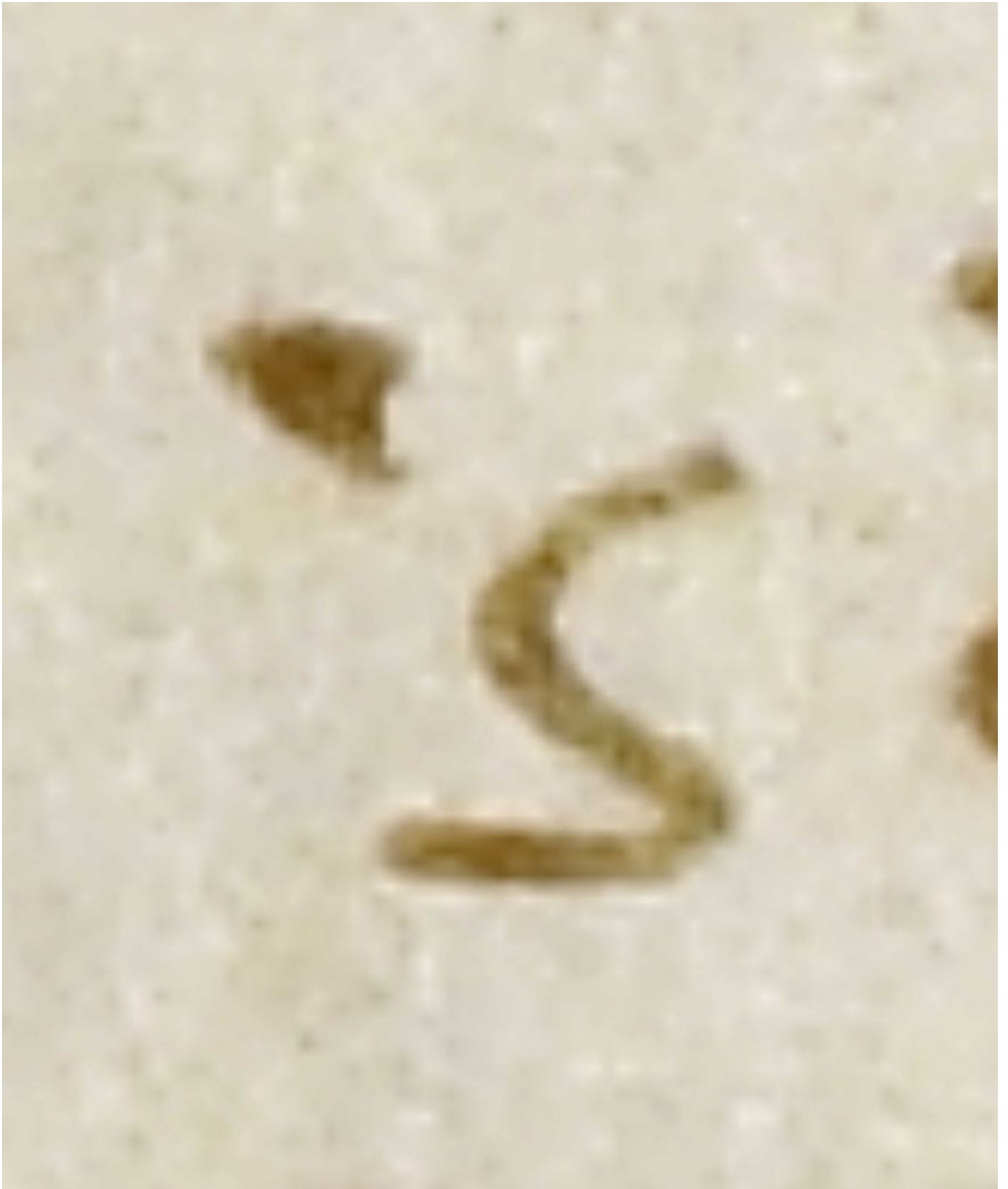
Abbreviations that Require Mark up

paleographically ambiguous abbreviations

Unlike the examples above, the marks that denote these abbreviations do not always indicate the same missing letters. Therefore, because they are paleographically ambiguous, they require mark up.

- Use the TEI `abbr/expan` pair, wrapped in `choice`

Example:

`<choice><abbr>ουτ</abbr><expan>οὐτῶς</expan></choice>`

(urn:cts:greekLit:tlg5026.msA.hmt:4.495)
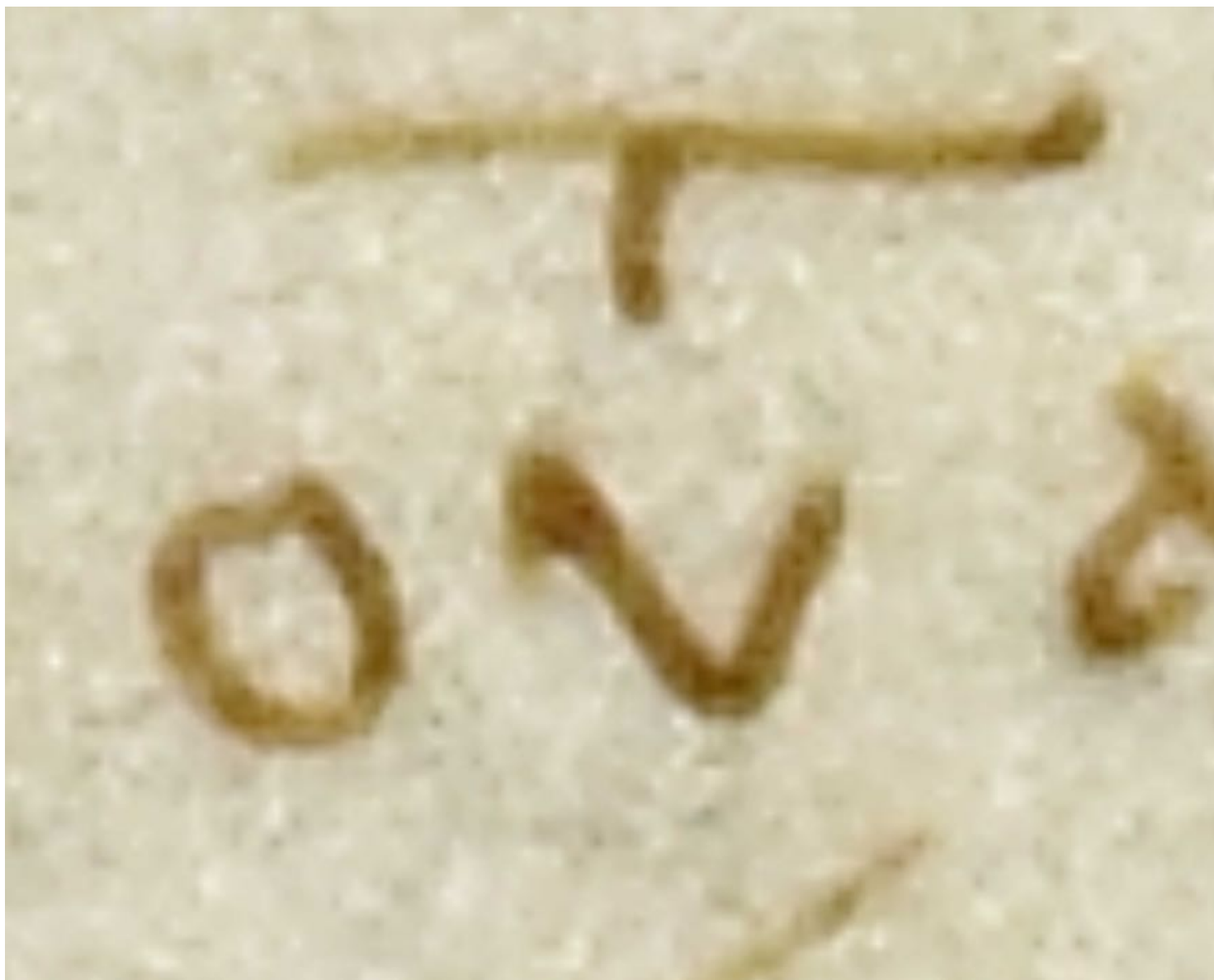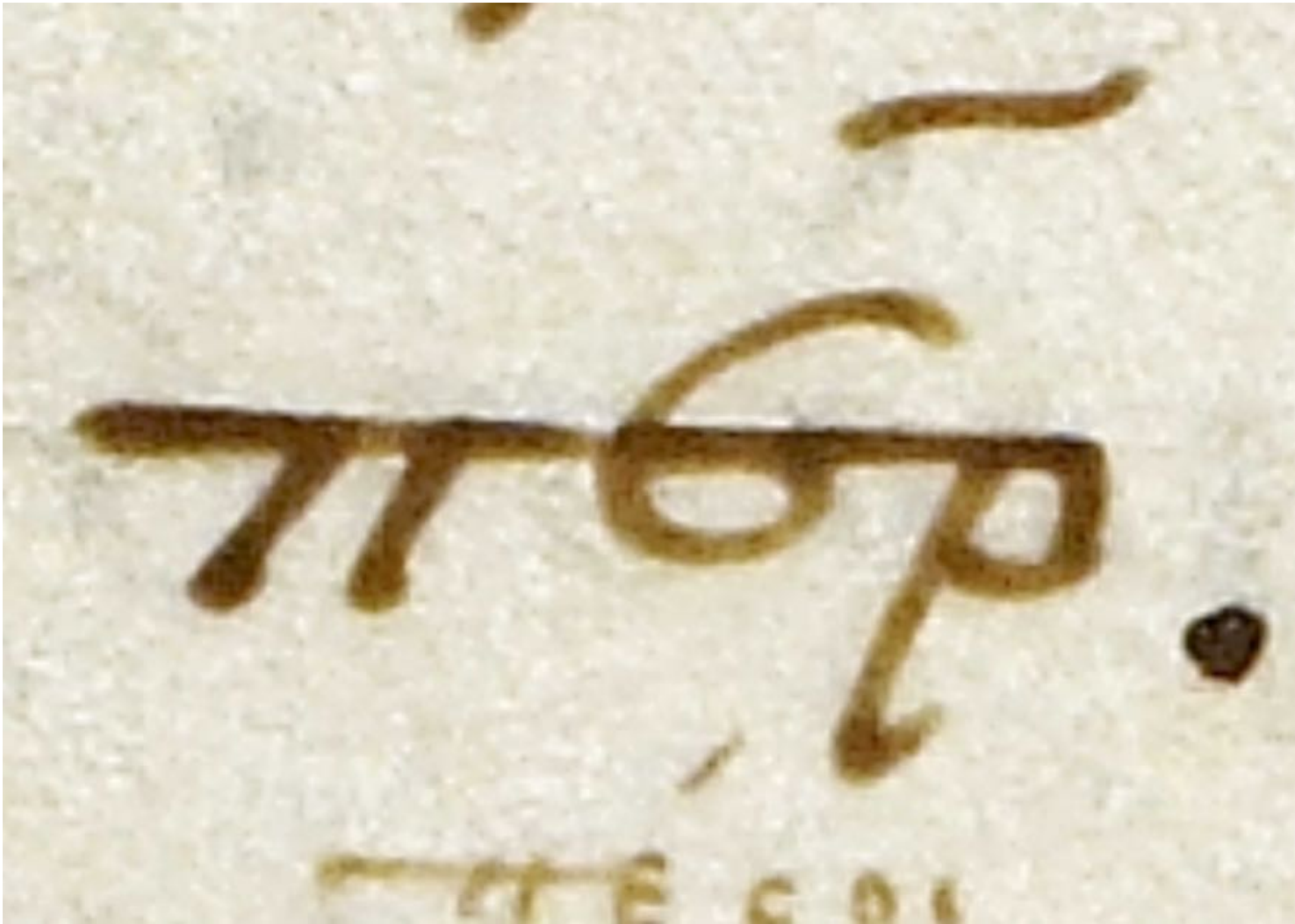
Example:

`<choice><abbr>πϵρ</abbr><expan>πάτϵρ</expan></choice>`

(urn:cts:greekLit:tlg0012.tlg001.msA:4.288)

Numbers

Use TEI `num` with `@value` attribute. Numbers are often denoted with a horizontal bar by the scribe to indicate that they are not lexical content.

<num value="50">v</num>

(urn:cts:greekLit:tlg5026.msA.hmt:1.42)

note: Remember that Dindorf and Erbse will typically take a Milesian numeral and write the full Greek word, but we want a fully diplomatic edition. Also remember that book numbers are treated as titles, not numbers in TEI.

Tier 3: Editorial disambiguation

Tier 3 disambiguates the text on an editorial level. This class of mark up is primarily named entities.

Named entities (proper nouns and adjectives)

personal names

For people, mythological or historical, use TEI `persName` element; include an `@n` attribute with the full URN value from the [reference table of identifiers for personal names][pers]. If a person does not yet exist in the authority list, create an issue in the repository to request its addition.

Example:

`<persName n="urn:cite:hmt:pers.pers1">Ἀχιλλεύς</persName>`

(urn:cts:greekLit:tlg0012.tlg001.msA:1.58)

- Note in cases where more than one person is refered to, such as the Atreidai, you double wrap personal names

Example:

`<persName n="urn:cite:hmt:pers.pers22"><persName n="urn:cite:hmt:pers.pers119">Ἀτρεΐδαι</persName></persName>`

(urn:cts:greekLit:tlg0012.tlg001.msA:1.17)

place names

For places, mythological or historical, use TEI `placeName` element; include on the `@n` attribute has a full URN value from the [reference table of identifiers for place names][place]. If a place does not yet exist in the authority list, create an issue in the repository to request its addition.
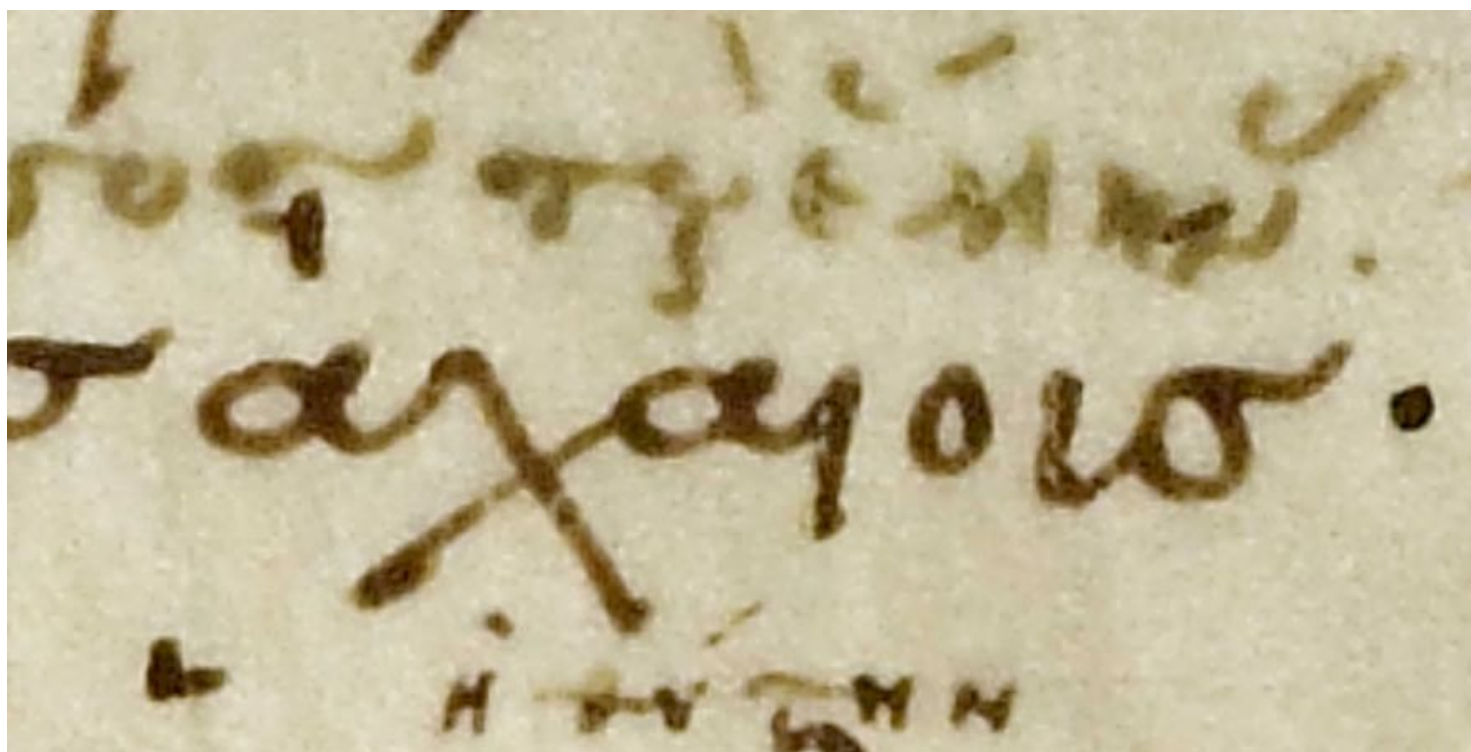
Example:

`<placeName n="urn:cite:hmt:place.place114">Ἄργεϊ</placeName>`

(urn:cts:greekLit:tlg0012.tlg001.msA:1.30)

ethnic groups

These are exclusively for groups of people that can be traced either to a geographic place or an emponymous ancestor. Use TEI `rs` element. Include a `@type` attribute with value `ethnic`, and `@n` attribute with an identifier from the [reference table for place names][place] or the [reference table for personal names][pers], when you need to use an eponymous ancestor (such as the Danaans).

Example:

`<rs type="ethnic" n="urn:cite:hmt:place.place96">Ἀχαιούς</rs>`

(urn:cts:greekLit:tlg0012.tlg001.msA:1.61)

- Since some ethnic names could be attributed to a location or an eponymous ancester (e.g. Trojans

link to Tros or Troy), always take the geographic location before resorting to an eponymous ancestor.

- Note that the ethnic mark up is only for people or groups of people (e.g. Trojan man or Trojans). We would not use it to describe a dialect (e.g. Ionic) or an inanimate object. Those are treated as normal vocabulary items that can be searched.
- Groups that cannot be tied to a geographic place or an eponymous ancestor are also treated as vocabulary items (e.g. Centaurs, Myrmidons, Gorgons, Giants, Fates, etc).
- Adjectival forms of people's names are also not tagged (ergo, we can tag 'Homer' but not 'Homeric')
- We also do not tag unclear epithets, even if the identity can be determined from context. For example "Phoebus Apollo" is ok, but the "Earthshaker" to refer to Poseidon is not. If you are in doubt, refer to the description of the indiviual in the authority list. If the epithet is not listed there, you can request it, via Issue Tracker and our editorial guide team can render a judgement.

astronomical bodies

: Use TEI `rs` element. Include a `@type` attribute with value `astro`, and `@n` attribute with an identifer from the reference table for astronomical bodies.

Example:

`<rs type="astro" n="urn:cite:hmt:astro.1">Ὠρίωνος</rs>`

(urn:cts:greekLit:tlg0012.tlg001.msA:18.486)

Tier 4: Discourse disambiguation

Quotation and citation

unidentified quoted words, phrases, or passages

Use TEI `q` for quoted words, phrases, or passages. A quoted selection must be at least one full parsable Greek work. Quoted phrases can ranges from hypothetical phrases to multiforms to quotations from works that we cannot cite. The scribe will not make quotes visually distinct. Use our print editions to help you, as well as context.

Example:

`<q>πρὸς ὃ</q>`

(urn:cts:greekLit:tlg5026.msA.hmt:1.1498)

identifiable quotations

Use TEI `ref/q` pair wrapped in a `cit` element. On `ref` element use `@type` attribute =`urn`. Inside `ref` element give the full cts urn for the quote (i.e. the text reference). The most common references will likely be *Iliad* or *Odyssey* quotes. If you have other citations, see the lists of citable [authors] and [works]. If you have a known text that is not referenced, create an [issue] in the repository. The scribe will not make quotes visually distinct. Use our print editions to help you, as well as context.

Example:

`<cit><ref type="urn">urn:cts:greekLit:tlg0012.tlg001:17.453</ref><q>ἔτι γάρ σφισι κῦδος ὀρέξω</q></cit>`

(urn:cts:greekLit:tlg5026.msA.hmt:4.8)

quoted strings

Quoted strings are defined as strings of letters that should not be considered parseable Greek words. One example might be the scholion talking about the use of the letter sigma. We do not want the machine to treat a single sigma as a Greek word so use TEI `rs` element with `@type` attribute = `waw` ("word-as-word"). Quoted strings are often, but not always, made visually distinct via a horizontal bar. They can be difficult to distinguish from the horizontal bars that denote numbers and you have to determine from context. The neuter article, prepositions, and other vocabulary related to writing will be your clue.

Example:

`<rs type="waw">v</rs>`

(urn:cts:greekLit:tlg5026.msA.hmt:1.1498)

Other Discourse Analysis

titles

If the title refers to a known, citable work, use TEI `ref` with `@type='urn'` and `@n` attribute with a full URN for the work
- Use TEI `title` element (wrapping `ref` if that is given). `ref` is not necessary if it is a work or portion of a work we do not have a URN for. You can find a full list of our citable authors and works. If a known work is not present, please create an issue in the repository to request its addition.

Example:

`<title><ref type="urn" n="urn:cts:greekLit:tlg0012.tlg002:">Ὀδυσσεία</ref></title>`

(urn:cts:greekLit:tlg5026.msA.hmt:1.4)

- Note that book numbers are considered titles, as are common names for books or portions of books (e.g. the Catalog referring to the Catalog of Ships in Book 2).

Titles that are not citable with a cts URN can still be cited with an object URN found at citable works.

Example:

`<title n="urn:cite2:hmt:citedworks:work2">Κυπρία</title>`

Like the named entities, new citable works can be proposed using the Issue Tracker on the authority