



**From Translation to Creation:  
Changes in Ethiopic Style and Lexicon from  
Late Antiquity to the Middle Ages**

Project working papers 3

GeTa Annotation Tool  
User Manual

(by Eugenia Sokolinski)

Hamburg, October 2019

# GeTa Annotation Tool: User Manual

## Contents

GeTa Annotation Tool: User Manual .....	1
1. General Introduction.....	3
1.1. GeTa tool.....	3
1.2. Technical prerequisites .....	3
1.3. Annotation .....	4
1.4. Transliteration.....	6
2. Working with GeTa: Basic options .....	7
2.1. Starting the tool and GUI .....	7
2.2. Opening a document: New document.....	7
2.3. Opening a document: Annotated document.....	9
2.4. Saving document.....	10
2.5. Export formats .....	12
3. Working with GeTa: Text metadata.....	14
4. Working with GeTa: Processing text.....	15
4.1. Tokenization.....	15
4.2. Transliteration adjustment .....	17
4.3. Edits to fidal text .....	18
4.4. Errors.....	19
4.5. Comments.....	20
4.6. Editorial mark-up .....	20
4.7. Line breaks and page breaks.....	21
5. Working with GeTa: Help functions .....	24
5.1. Show window.....	24
5.2. Tools: Search .....	24
5.3. Tools: Undo function.....	26
5.4. Tools: Statistics.....	26
5.5. Help: Linguistic Annotation Colours.....	26
5.6. Help: Verify Index.....	26
5.7. Help: Correct Line Break Nr .....	27
6. Working with GeTa: Linguistic annotation.....	27
6.1. Basic annotation: POS.....	27
6.2. Linking to dictionary.....	28
6.3. Completing annotation .....	30
6.4. Figure 46 'Deep' annotation .....	30

6.5.	Local and global changes .....	32
7.	Working with GeTa: Text structure.....	33
7.1.	Introducing levels.....	33
7.2.	Assigning spans .....	34
7.3.	Editing spans .....	36
8.	Working with GeTa: Named entities.....	36
8.1.	Places, persons, works .....	36
8.2.	Global annotation .....	39
8.3.	Editing NE annotation .....	40
8.4.	Titles, dates, abbreviations .....	40
8.5.	Hiding NE annotation .....	40
9.	Working with GeTa: particular examples of annotation.....	41
9.1.	Disambiguation .....	41
9.2.	Defining the gender and number in nouns.....	42
9.3.	Using the comment function .....	43
	Glossary of terms .....	45

## 1. General Introduction

### 1.1. GeTa tool

GeTa (Gə'əz Toolkit for Annotation) is an application developed within the framework of the project TraCES: From Translation to Creation: Changes in Ethiopic Style and Lexicon from Late Antiquity to the Middle Ages.<sup>1</sup> The tool supports annotation of texts in Gə'əz (or Classical Ethiopic).

The tool is programmed in Java and stores data in JSON format, while offering data export into several other formats, including ANNIS linguistic visualization platform (<http://corpus-tools.org/annis/>; via an additional converter) and TEI-XML.<sup>2</sup>

### 1.2. Technical prerequisites

#### (1) System requirements

GeTa tool has been tested on the Windows Platform. A cross-platform usage is foreseen but cannot be guaranteed at the moment.

You will need Windows 7 or later and Java 8 or later for the tool to perform properly, and sufficient operative memory, especially for work with longer texts. For a smooth performance, RAM of 8GB and up is desirable.

#### (2) Fonts

GeTa tool uses the 'Ethiopic Unicode' and 'Titus' fonts; please install the fonts before using GeTa.

#### (3) Source text

The source text (prior to annotation) should be in .txt format with UTF-8 encoding. It may contain exclusively Ethiopic characters (*fidal*) in a Unicode font.<sup>3</sup> The accepted punctuation marks are limited to the word separator (:) and the 'four dots' full stop (:). No spaces before punctuation marks are accepted. A single space must follow the punctuation marks. Make sure that the 'four dots' correspond to the single Unicode (1362) and not to two times : (If any of these conditions is not fulfilled the text may fail to load).

*Note that the tool has been tested with texts containing 35 000 graphic units or less. If you want to work with a longer text, it may be reasonable to split it into parts first, as otherwise the processing time may become unnecessarily slow.*

---

<sup>1</sup> Funded by the European Research Council under the European Union's Seventh Framework Programme, grant agreement no. 338756.

<sup>2</sup> The tool has been previously introduced on various occasions; see e.g. C. Vertan, 'Bringing Gə'əz into the Digital Era: Computational Tools for Processing Classical Ethiopic', in A. Bausi and E. Sokolinski, eds, *150 Years after Dillmann's Lexicon: Perspectives and Challenges of Gə'əz Studies*, Supplement to Aethiopica, 5 (Wiesbaden: Harrassowitz Verlag, 2016), 31–42; S. Hummel, V. Pisani, and C. Vertan, 'Multi-level Digital Annotation of Ethiopic Texts', *Comparative Oriental Manuscript Studies Bulletin*, 4/1 (2018) = *Proceedings of the Workshop Linking Manuscripts from the Coptic, Ethiopian, and Syriac Domain: Present and Future Synergy Strategies Hamburg, 23 and 24 February 2018*, 97–106; S. Druskat and C. Vertan, 'Nachnutzbarmachung von Forschungsdaten und Tools am Beispiel altäthiopischer Korpora', in G. Vogeler, ed., *Kritik der digitalen Vernunft. Abstracts zur Jahrestagung des Verbandes Digital Humanities im deutschsprachigen Raum, 26.02.–02.03.2018 an der Universität zu Köln, veranstaltet vom Cologne Center for eHumanities (CCeH)* (Köln: Universität zu Köln, 2018), 270–273.

<sup>3</sup> The tool currently supports work with vocalized *fidal* script only. In the future annotation of unvocalized Gə'əz and/or Gə'əz written in Sabaic characters (as used in inscriptions) may be enabled.

### 1.3. Annotation

Annotation means that additional information (meta-data) is provided to a document at some level—it can be a letter, a word, a phrase, a paragraph, a section, etc., or also the entire document. This meta-data can be then subject of various search requests. In the case of GeTa, it is possible to

- (1) Provide detailed linguistic information;
- (2) Provide lexical information;
- (3) Mark-up named entities;
- (4) Add structural mark-up;
- (5) Introduce editorial remarks.

Comments can be left at each level of annotation. Note that you can freely choose colours to highlight commented graphic units; commented tokens have grey background.

Each document can be further linked to an external resource where general meta-data on the annotated text is stored.

#### (1) Linguistic annotation

The GeTa application allows thorough morphological analysis since labels (tags) can be assigned to minimal morphological units (tokens) and not to graphic units in their entirety, as is the case with the majority of other annotation tools available. For each token, the users can select one of 33 parts of speech, and subsequently optionally select the relevant features. For the list of the available labels, see Table 1.<sup>4</sup> The possible combinations amount to over 6,700 (4,860 for the common noun alone)—the unique value of the GeTa tool.

#### (2) Lexical annotation

Each lexical unit (word) can be linked to the relevant entry in the online dictionary of Ethiopic (also developed by the TraCES project) accessible at <https://betamasaheft.eu/Dillmann/>.<sup>5</sup>

#### (3) Named entities

In addition to being marked up linguistically as ‘proper names’, names of persons and toponyms (whether consisting of one or several graphic units) can be further annotated by linking them to an external authority file (stored at <https://betamasaheft.eu/>) where extensive meta-data on these entities is stored. Dates and titles can also be tagged.

#### (4) Text structure

The tool allows splitting text in semantic sections of various levels (e.g. chapters, sub-chapters, verses, etc.). The beginning/end of a page (e.g. when following a printed edition) or a line (e.g. in poetry or inscriptions) can also be marked.

#### (5) Editorial mark-up

GeTa is primarily a tool for linguistic annotation, not for text editing, and therefore the editorial mark-up feature is relatively basic. Yet, sometimes (e.g. in case of reconstructed, missing, or doubtful text) it is important to be able to reproduce the existing editorial annotation. This is possible in the tool.

---

<sup>4</sup> More on the grammars used and on the development of the Part of Speech tag set in S. Hummel and W. Dickhut 2016. ‘A Part of Speech Tag Set for Ancient Ethiopic’, in A. Bausi and E. Sokolinski, eds, *150 Years after Dillmann’s Lexicon: Perspectives and Challenges of Gə‘əz Studies*, Supplement to *Aethiopica*, 5 (Wiesbaden: Harrassowitz Verlag, 2016), 17–29.

<sup>5</sup> See also E. Sokolinski, ‘The ‘TraCES’ Project: Towards a New Approach to Studying the Gə‘əz Language’, *Comparative Oriental Manuscript Studies Bulletin*, 4/1 (2018) = *Proceedings of the Workshop Linking Manuscripts from the Coptic, Ethiopian, and Syriac Domain: Present and Future Synergy Strategies Hamburg, 23 and 24 February 2018*,) 59–66.

Table 1. Parts of speech and possible features in the GeTa tool

Class	Subclass	POS	Features and examples
Nominals	Nouns	Common Noun	Gender (masculine/feminine/unmarked, nature/pattern/syntax), Number (pattern/syntax, singular/plural external/plural internal/plural of the plural/unmarked), Case (Nominative/Accusative/Accusative zero/Vocative), State (Absolute/Construct/Construct zero/Pronominal)
		Proper Name	Gender, Number, Case (options as above)
	Pronouns	Independent Personal Pronoun	Person (first/second/third), Gender (masculine/feminine/communis), Number (singular/plural), Case (Nominative/Accusative) · አነ፣ አንተ፣ አንቲ፣ ወአቲ፣/ወአተ፣ ይአቲ፣/ይአተ፣ ...
		Pronominal Suffix	Person, Gender, Number (options as above) · + የ/ + ኒ, +ከ, +ኪ, +ሁ, +ሃ ...
		Subject Pronoun Base	no features · ለሊ+PSuff (but 1c.sg.: ለሊየ፣ or ለልየ፣ or ለሊየ፣)
		Object Pronoun Base	no features · ኪየ+PSuff
		Possessive Pronoun Base	Gender, Number (options as above) · m.sg. ሲአ+PSuff, f.sg. አንቲአ+PSuff, c.pl. አሊአ+PSuff
		Demonstrative Pronoun	Gender, Number, Case (options as above) · near: m.sg. ዝ+/ዘ+ or ዝንቲ፣/ዝንተ፣, f.sg. ዛ+ or ዛቲ፣/ዛተ፣ m.pl. እሉ፣ or እሎንቲ፣/እሎንተ፣, f.pl. እላ፣ (እሎን፣) or እላንቲ፣ (እላንቲ፣)/እላንተ፣ · far: m.sg. ዝኩ፣/ዝኩ፣ or ዝኩቲ፣/ዝኩተ፣, f.sg. አንትኩ፣ /አንትኩ፣ or አንታኩቲ፣/አንታኩተ፣, c.pl. አልኩ፣ or አልኩቲ፣ (አልኩቲ፣)/አልኩተ፣ (አልኩተ፣) or አላኩቲ፣/አላኩተ፣
		Relative Pronoun	Gender, Number (options as above) · m.sg. ዘ+, f.sg. አንተ፣, c.pl. እሉ፣
		Interrogative Pronoun	Number, Case (options as above) · መነ፣/መነ፣ who?, ምንት፣/ምንተ፣ what?, sg. አይ፣/አየ፣ : which?, pl. አያት፣/አያተ፣
		Pronoun of Totality Base	Case (options as above) · ተሰላ+PSuff <i>all, everything</i> · ተሰላንት+PSuff <i>whole, in the whole being</i>
		Pronoun of Solitude Base	Case (options as above) · ባሕቲት+PSuff <i>alone</i>
	Numerals	Cardinal Numeral	Gender (masculine/feminine/unmarked), Number (singular/plural), Case (Nominative/Accusative), State (Absolute/Construct/Pronominal), Logogram (free text field to insert the Arabic numeral)
		Ordinal Numeral	Gender, Number, Case, State, Logogram (options as above)
Verbs		Verb	Person (first/second/third), Gender (masculine/feminine/communis), Number (singular/plural), Tense/Aspect/Mood (perfect/imperfect/subjunctive/imperative/gerund/infinitive) For Infinitive: Case, State (as for nouns above)
Existentials		Existentials Affirmative Base	no features · በ+PSuff <i>there is</i> (ብየ፣ ብከ፣ ብኪ፣ በ፣ or ቦቲ፣ ባ፣ or ባቲ፣ ...)
		Existentials Negative Base	no features · አልበ+PSuff: (አልብየ፣ አልብከ፣ አልብኪ፣ አልቦ፣ or አልቦቲ፣ አልባ፣ or አልባቲ፣ ...)
Particles	Adverbs	Interrogative Adverb	no features · አይቲ፣ where? ማእከ፣ when? እፎ፣ how? ሚመጠን፣ how many?
		Other Adverb	no features · location: ህየ፣ there, አፍአ፣ outside, ዝየ፣ here, ፀጋም፣ left · time: ቀዳሙ፣ before, ካዕበ፣ again, ዮም፣ today, ይእከ፣ now, ኒሰም፣ tomorrow · others: መጠነ፣ as large as, ቀዳሚ፣ at first, ባሕቲ only, አማን፣ truly, ግሙራ፣ entirely
	Preposition		State (nominal and pronominal state) · enclitic: ለ+ for, በ+ with, ከም+ from · independent: ላዕለ፣ above, ምስለ፣ with, ማእከለ፣ between, ቅድመ፣ in front, ጎበ፣ to · compound: በጎበ፣ at, ዘእንበለ፣ without, በአንተ፣ concerning, እንበይነ፣ for the sake of · comparative: እም፣ than, ከም፣ like
	Conjunction		no features · ባሕቲ፣ yet, ሶበ፣ when, ጎበ፣ where, በጎበ፣ (there,) where, አላ፣ but rather, እም፣ since, አመ፣ when, አመ፣ if, እሰመ፣ because, እስከ፣ until, አንዝ፣ while, አው፣ or, ከመ፣ (so) that, ው+ and, that, but · +subjunctive so that
	Interjection		State (nominal/pronominal) · to express woe!፣ አሊ፣ (+ለ+PSuff: አሊልየ፣ or አሊ፣ ሊተ፣ ), ወይ፣ · አ+ oh!, አህ፣ ah!, እንቋዕ፣ ha!, ዩ፣ or ዮ፣ alas!, ጸጥ፣ quiet! · ባሕ፣ hail! (also +PSuff 2m/f: ባሕከ፣ ባሕኪ፣)

Class	Subclass	POS	Features and examples
	Further Particles	Accusative Particle	no features · +ሃ (suffixed only to proper names)
		Affirmative Particle	no features · አወ፡ , +ጓ <i>yes indeed!, certainly!</i>
		Deictic Imperative Particle	Person, Gender, Number · m.sg. ነፃ፡ (ነፃ፡), f.sg. ንፃ፡ m.pl. ንዐ፡, f.pl. ነፃ፡
		Interrogative Particle	no features · +ሁ, +ኦ
		Negative Particle (Base)	no features · ኢ+, አልቦ፡ (= PartNeg+PSuff:3m.sg), አንቢ+PSuff: እንቢየ፡, አኮ፡
		Presentational Particle Base	no features · <i>Behold!</i> ና+PSuff:3m.sg (ናሁ፡), ነው+PSuff:3f.sg (ነዋ፡), ነይ +PSuff (not second person), እን +PSuff
		Quotative Particle	no features · +አ
		Vocative Particle	no features · አ+
		Other Particles	no features · +ሂ <i>and also</i> , + <i>even</i> , + <i>doch!</i> , + <i>even</i> , + <i>to</i> , <i>O doch!</i> , <i>so then!</i> , + <i>now</i> , <i>just</i>
Foreign Material			· For example ሰገድ፡ <i>sagad</i> (element in proper names such as 'Aṣnāf Sagad, regnal name of King Galāwdewos, r. 1540–1559).
Punctuation			

Check paragraph 9 below for some examples of the features in linguistic annotation.

#### 1.4. Transliteration

The GeTa tool is based on the transliteration convention, which is currently used by all projects based at the Hiob Ludolf Centre for Ethiopian Studies at Universität Hamburg. For the transliteration table, see Table 2 (see also <http://betamasaheft.eu/Guidelines/?id=transliteration-principles>).

Table 2. Transliteration principles

Consonants						
ሀ	ha	ገ	ga			
ለ	la	ጠ	ṭa			
ሐ	ḥa	ጰ	p̣a			
መ	ma	ጸ	ṣa			
ሠ	śa	ፀ	ḍa			
ረ	ra	ፈ	fa			
ሰ	sa	ፐ	pa			
ቀ	qa	ቈ	q <sup>w</sup> a			
በ	ba	ኸ	ḥ <sup>w</sup> a			
ተ	ta	ኰ	k <sup>w</sup> a			
ኀ	ḥa	ኰ	g <sup>w</sup> a			
ነ	na	ቨ	va			
አ	'a	ቸ	ča			
ከ	ka	ጀ	ǰa			
ወ	wa	ጨ	ča			
ዐ	'a	ቐ	qa			
ዘ	za	ኸ	ḵa			
የ	ya	ሸ	ša			
ደ	da					
Vowel orders						
1	2	3	4	5	6	7
a	u	i	ā	e	ə	o

## 2. Working with GeTa: Basic options

### 2.1. Starting the tool and GUI

Once you have downloaded the tool (in the .jar format) you are all set, no additional installation is needed.

In order to open the GeTa application, double click on the .jar file in the repository you have downloaded it to, and the main window will open (see Figure 1).



Figure 1

*Note that in the top menu, as well as in all programme submenus, available functions are shown in black; inactive functions appear in grey and cannot be clicked.<sup>6</sup>*

### 2.2. Opening a document: New document

In order to start work with a new document (which has not been previously annotated), choose in the top menu File -> Open -> New fidal file (Figure 2).

*(Make sure that the file contains only the Ethiopic text (using a Unicode-based font), saved in the .txt format with UTF-8 encoding (see also §1.2 above)).*



Figure 2

<sup>6</sup> In most cases, text in grey means that the function is disabled at the current work stage. E.g. prior to loading the text, you cannot 'Close', 'Save', or 'Export' it (fig. 2); after a graphic unit has been tokenized, the 'Tokenise' function is disabled (fig. 15); for a non-infinitive Verb 'State' and 'Case' do not apply and cannot be filled (fig. 33). In some cases, this means that the function is still under development, such as e.g. 'Join GraphicUnits' and 'Join Morphologically' (fig. 10).



By default, the menu shall open in your local user menu; navigate to the folder you have saved your file to and click to open.

You will be prompted to choose the script you shall be working with (South Arabian or Ethiopic: click 'No' to continue working with Ethiopic script) and then to choose whether you are working with vocalized or unvocalized Gə'əz: click 'No' to continue working with vocalized Ethiopic script (Figure 3).

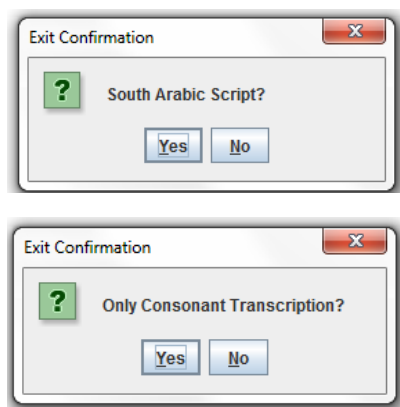


Figure 3

Now the tool is indexing the text and producing the automatic transliteration; this may take some minutes depending on the file size.

Once the indexing and transliteration process are complete, two text frames shall appear in the main interface window: the original *fidal* text on the left and the automatically produced transliteration on the right (see Figure 4).

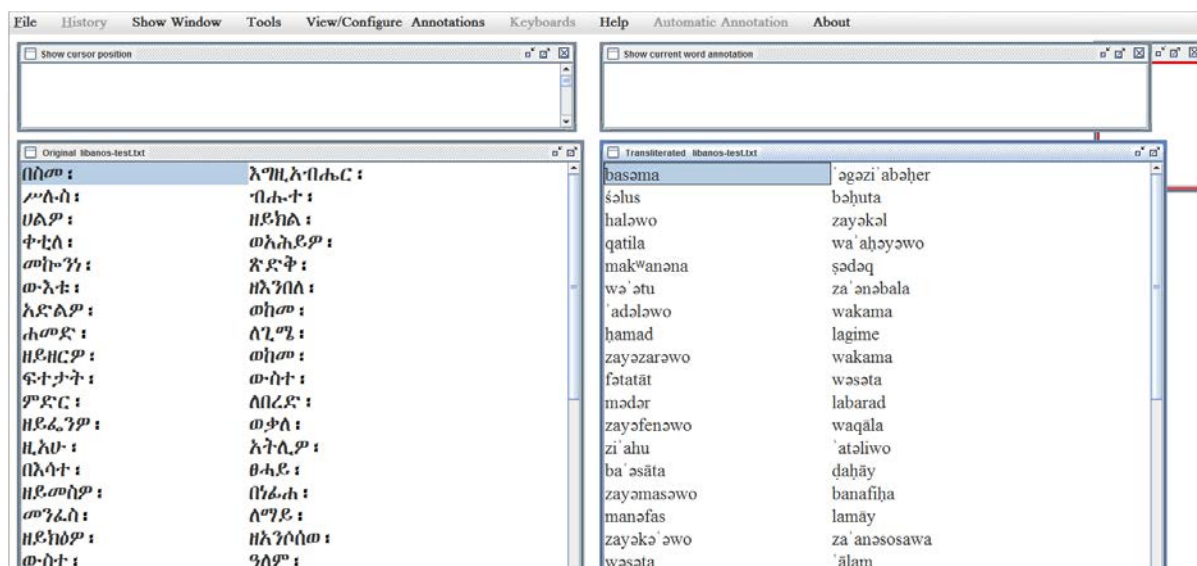


Figure 4

Note that you can adjust the width of each frame manually (e.g. making the frames wide enough to accommodate three horizontal cells, see Figure 5). For long texts, the view adjustment may take some time, but it has to be carried out only once; the next time you open your file, it will open in the last layout you used it in.



During the opening process, the tool controls the consistency of the index; if there are problems, these shall show in the error window (see § 4.4 below).

## 2.4. Saving document

Your file can be saved at any moment of work.

### 2.4.1. Saving for the first time

The very first time you can save the file by going to File -> Save As. You will be asked to browse to the directory on your computer you would like to save the file to and choose the filename (Figure 8). The tool shall then create a project folder bearing the name consisting of the name you chose (e.g. 'genesis') followed by a time stamp for the version control (e.g. 'genesisVER18\_12\_07\_13\_54\_12').

*NB it may be a good idea to select a file name followed by an underline (e.g. 'genesis\_') to generate an easily readable filename (e.g. 'genesis\_VER18\_12\_07\_13\_54\_12').*

The folder contains seven files, the main text file in JSON format and several auxiliary project files (in .ann and .ind formats) generated by the tool. The file names consist of the name you chose and (eventually) an extension specifying the tool version you used (e.g. 'genesisGeTa Ver5\_25Juni18.JSON'; see Figure 9).

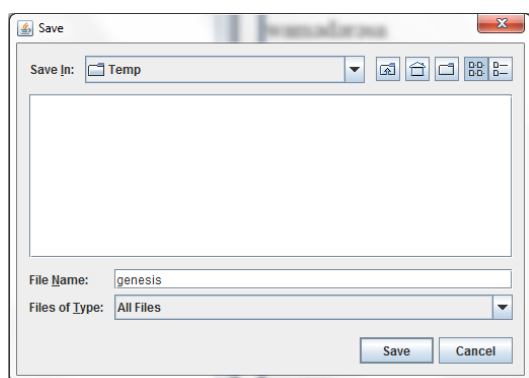


Figure 8

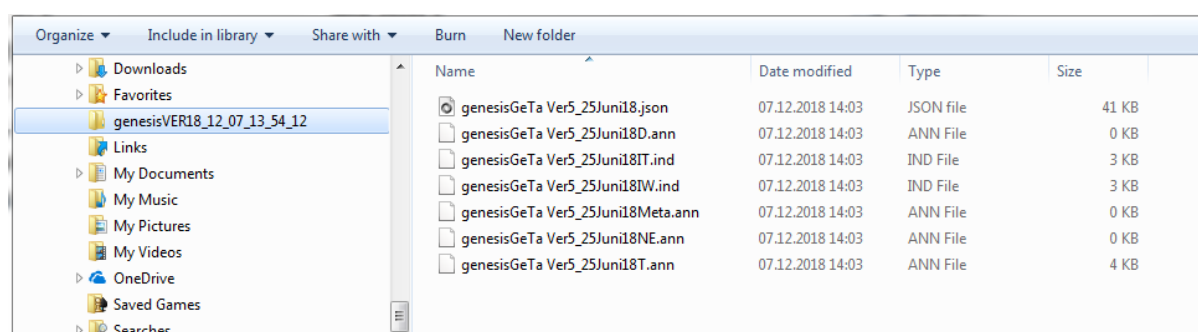


Figure 9

### 2.4.2. Re-saving the document and versioning control

Once a GeTa version of the text has been created on your computer, you can always update it by clicking File -> Save (or Control+S). In this case, the versioning extensions are not modified, and the time stamp remains unchanged, even if changes are introduced later.

For full versioning control, it is recommended to use the 'Save As' function, especially if a series of significant changes have been introduced. You can choose the same filename and yet a new folder with a new time stamp shall be created in your chosen directory (e.g. 'genesisVER18\_12\_21\_11\_51\_10'), with the current version of your annotation project.

If you ever attempt to close the document you have been working on or the tool window without saving, you will be prompted to save your project (see Figure 10). If you click ‘Yes’ in the prompt window, the latest project version shall be updated (without creating a new time-stamped version) and the file closed. If you want to create a new time-stamped folder select ‘Cancel’ and then use the ‘Save As’ function as described above.

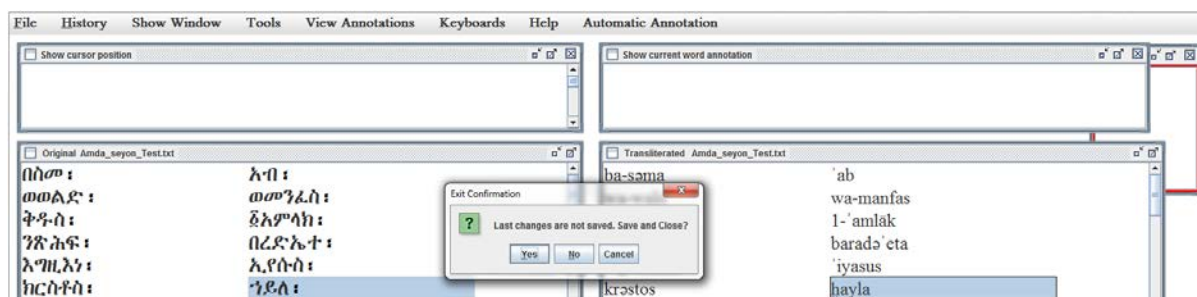


Figure 10

#### 2.4.3. Saving parts

Once your text has been annotated structurally (see § 7 below on assigning text divisions) you can also choose to save only specific sections by selecting File -> Save Parts on the save menu. You will first see a warning message that once you complete the save part operation the current file shall be closed to avoid confusion. Make sure you have saved the file as it is before clicking ‘Yes’.

Subsequently, you will be asked which parts you would like to save (see Figure 11). Select the part(s) and click OK to be taken to the ‘save as’ pop up window where you can select a file name for the new project. A new time stamped project folder will then be generated on your computer and the current project shall be closed to ensure that when you continue your work you know whether you are doing so in the ‘long’ or in the ‘short’ project.

This function is only possible after text spans have been assigned, you cannot simply save passages if these have not been marked up first.

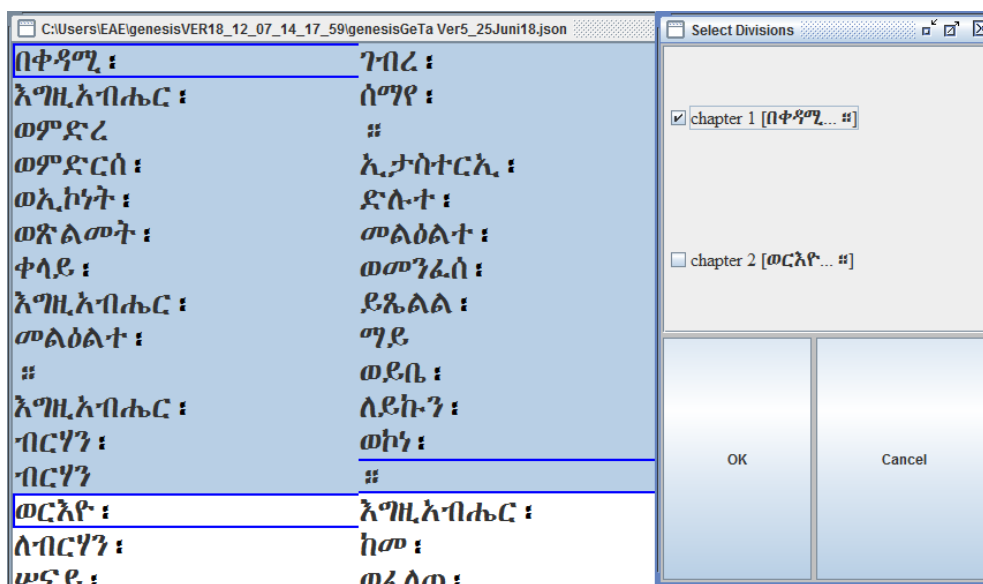


Figure 11

## 2.5. Export formats

In addition to saving your file in the GeTa-own json format you can also choose to export your document to formats that can be used by other applications. Go to the Export submenu in the File menu for the options (Figure 12).

ANNIS (the abbreviation stands for ANNotation of Information Structure) is a web-based platform for search and visualization of linguistic corpora. The ANNIS software can be freely downloaded at <http://corpus-tools.org/annis/> and locally installed and used; the server installation of Hamburg Zentrum für Sprachkorpora hosts some of the texts generated with the GeTa tool within the TraCES project. In order to create ANNIS-readable data, choose the “Export to Converter” option. In this case, a time-stamped Export subfolder containing seven (7) time-stamped files (with the same extensions as the files in your working directory) will be generated inside the directory of your annotated file. Run the Pepper-Grinder Converter application (see the Pepper Grinder user guide) to generate from these new files the ANNIS-readable data.

TEI-XML is the extensible mark-up language standard developed by the Text Encoding Initiative and used by many projects and platforms worldwide. The TEI guidelines can be viewed at <http://www.tei-c.org/guidelines/>. The GeTa tool allows exporting the complete annotation, only the linguistic annotation, or only the text structure with named entities mark-up (the latter is the one most frequently employed by projects publishing TEI-encoded texts for research purposes).

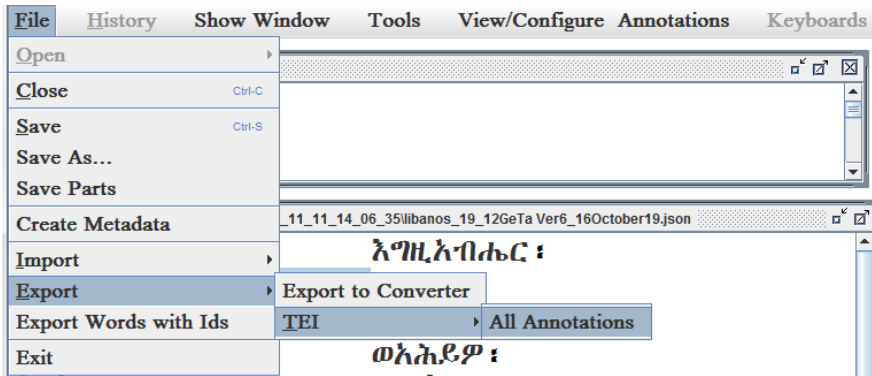


Figure 12

Choose “All Annotations” export to produce a single XML file (with the name \*YOURFILENAME\*TEI.xml) in the directory of your annotated file. The XML file contains all annotations formatted according to the general TEI recommendations; it is not associated to any particular schema (see Figure 13 for the XML file snippet prior to formatting and after formatting).



Figure 13

Note that the links to the authority files (dictionary entries, named entities) are provided in the form of short IDs (e.g. for the dictionary `<f name="lex">L52ad20427e464053b7820f87cf6dfd0d--%#45</f>`; for the named entity `<f name="lit" fVal="ezānā" corresp="#PRS3938Ezana">`) used by the applications <https://betamasaheft.eu/Dillmann> and <https://betamasaheft.eu>, respectively. If you want to use the data on a platform different from <https://betamasaheft.eu>, and maintain the active links, they must be extended to, for the example provided, <https://betamasaheft.eu/Dillmann/lemma/L52ad20427e464053b7820f87cf6dfd0d> and <https://betamasaheft.eu/PRS3938Ezana/>.

The ‘Export Words with Ids’ function is used by developers rather than by annotators. If you use the ‘Export Words with Ids’ function you will generate three new files in your current project folder: (1) a text file containing the entire text in fidal; (2) a text file containing the entire text in transliteration, and (3) a simple XML file and containing your text as a list of words in transliteration supplied with their unique identifiers as automatically assigned by the tool (Figure 14).

```
<w tr="ba-səma" id="Wf2e5da2f-322e-4eab-94df-19c2487220c3"></w>
<w tr="ʾab" id="Wd5067cb3-4640-454a-893f-f1a44e71b26b"></w>
<w tr="wa-wald" id="W40bd3b02-c262-4b36-b5a3-bfa51062e486"></w>
<w tr="wa-manfas" id="Wede71ec2-d171-4149-98c6-e0fa76c12e3f"></w>
<w tr="qəddus" id="W9d7fd567-a8b1-4f06-b3b3-696383a12cbf"></w>
<w tr="1-ʾamlāk" id="W3e71487b-02df-4318-8263-22c66173e1dd"></w>
<w tr="nəṣṣəf" id="Wd84cbcff-b0a4-4c1d-a49a-878dd97e7ecd"></w>
<w tr="ba-radʾeta" id="Wc5dd3162-a3f7-4ae1-a6cc-f46afa4b8338"></w>
<w tr="ʾəgziʾə-na" id="W23302680-5f3a-44e8-b3c5-416b7538e460"></w>
<w tr="ʾiyasus" id="W94e42b68-0b0b-4d14-b2de-f322da2d6aba"></w>
<w tr="krəstos" id="W2fe517a1-6ac8-4949-99c2-2fba30629ad5"></w>
<w tr="ḥayla" id="Wf2c21581-090a-4e4f-a7f7-8b917da9c5eb"></w>
<w tr="wa-mawiʾa" id="Wb043553e-cf52-402f-9204-b06d3f896851"></w>
<w tr="za-gabra" id="W5bc8c904-f398-42f1-bd30-f4e1f15f6e50"></w>
<w tr="ʾəgziʾabəḥer" id="W294f54c0-7056-496b-b9fd-2fb7a6ebdbf9"></w>
<w tr="ba-ʾədawi-hu" id="W6eca1e22-11d4-4c4a-a804-d588d290022d"></w>
<w tr="la-ʿāmda" id="Wb1874f19-a0e2-4649-b8f8-8b2aad6a8976"></w>
<w tr="ṣəyon" id="W4afcd588-5d4c-4c91-bc31-97fa33e705ee"></w>
<w tr="nəḡuša" id="W04e2a76d-d469-43e3-9993-49b3eded3d42"></w>
<w tr="ʾityōpyā" id="Wf0ab39fd-0bdf-4291-8cd4-993cca8d3050"></w>
<w tr="wa-səma" id="W6c29aaf2-2719-4063-a1df-6beea48df1e3"></w>
```

Figure 14

It is possible to process this file by tagging the named entities following the XML schema distributed with the tool (WordsID.xsd) and then import the modified file using the ‘Import -> Import NE’ function in the File menu.

### 3. Working with GeTa: Text metadata

Once a document has been created it must be accompanied by the general meta-data. Choose the 'Create Metadata' option from the File menu. In the new frame, fill the fields as requested (Figure 15).

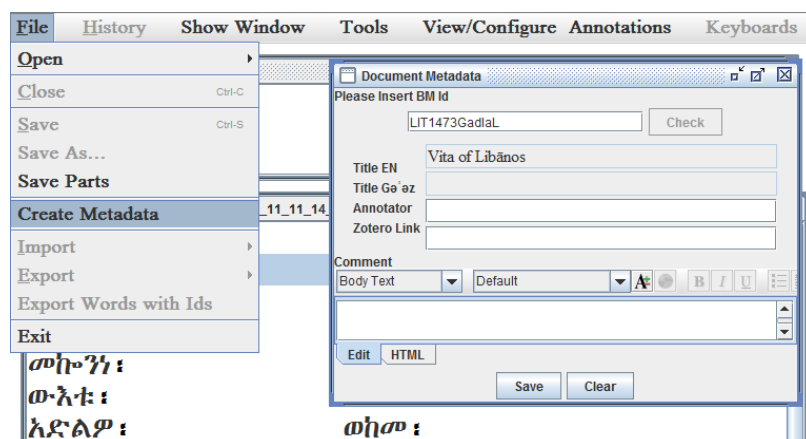


Figure 15

Only minimal data is requested in the GeTa tool, the extensive metadata is stored in TEI XML format on the [betamasaheft.eu](https://betamasaheft.eu) server. Search the [betamasaheft.eu](https://betamasaheft.eu) database to recover the work ID and type it in the upper field. On clicking 'Check' the tool shall connect to the database in real time and retrieve the title information. You can now additionally provide the annotator information, if you are using a published edition, you can also provide a link to the [ethiostudies Zotero](https://www.zotero.org/) library entry for this publication, and enter additional comments.

Check the metadata provided on [https://betamasaheft.eu/](https://betamasaheft.eu) and edit the relevant XML file or inform the project team if any important information is missing and should be added.

If there is so far no fitting record in the [https://betamasaheft.eu/](https://betamasaheft.eu) database, contact the project team so that the record can be created.



#### 4. Working with GeTa: Processing text

Once the transliterated text has been loaded, users should prepare the text for being annotated. This implies correcting possible errors in fidal, adjusting the transliteration (by removing the unnecessary sixth-order vowels and introducing gemination if morphologically (or phonetically) required) and, most importantly, splitting the graphic units into the minimal morphological units (tokens).

*The precise workflow for these actions may vary from user to user. It may be advisable, especially for larger texts, to carry out the tokenization before proceeding to adjusting/correcting the transliteration itself, even if the transcription may feel wrong, as then batch corrections on single tokens are possible. Yet in some cases the transliteration must be corrected before tokenization, see 4.2.1 below.*

##### 4.1. Tokenization

In order to tokenize a graphic unit consisting of more than one morphological units, select the graphic unit in the transliteration frame, and right-click the mouse for a drop down menu. Select 'Tokenise' in the menu (Figure 16).

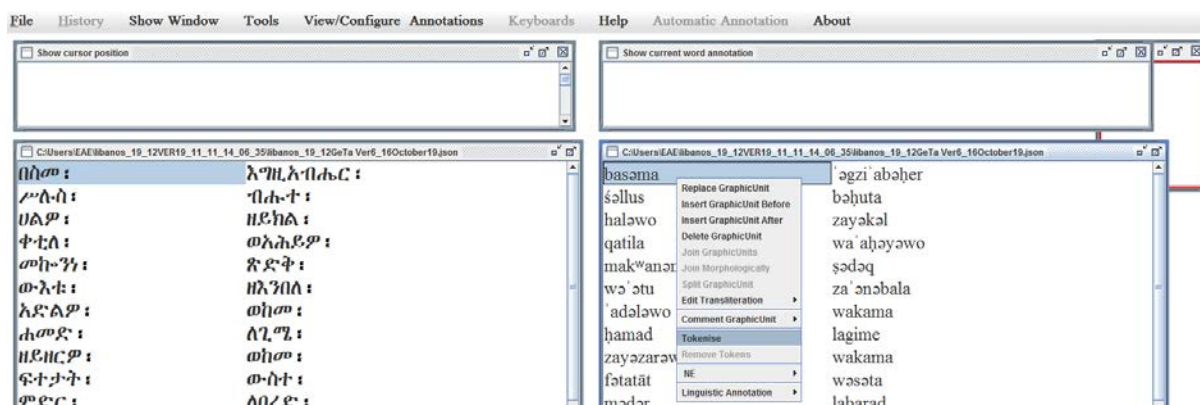


Figure 16

A new frame shall pop-up showing the transliterated letters of the graphic unit. You can now select the last letter of the first token and click 'New token' to introduce a break (marked by a hyphen) (Figure 17).

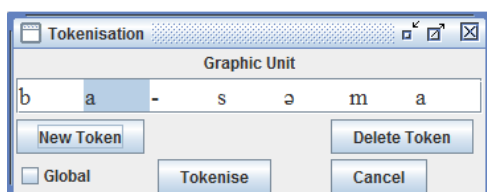


Figure 17. Introducing tokenization

If during the tokenization process you have noticed that the break was introduced erroneously you can delete it by selecting the n-dash and clicking 'Delete token' (Figure 18).

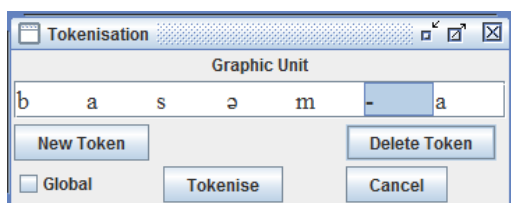


Figure 18. Deleting token during tokenization



Once the word in the frame is fully tokenized you can execute the changes by clicking the ‘Tokenise’ button (Figure 19). If the same tokenization should be applied to all occurrences in the text you can tick the ‘Global’ option.

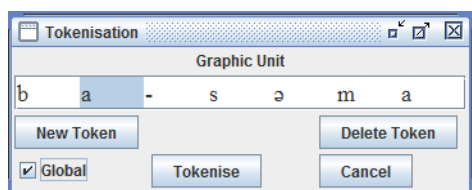


Figure 19. Assigning tokenization

The globally tokenized units appear italicized in the transliteration frame (see Table 3 for the format and colour code), so that the user can easily see what has been done automatically and not manually, and by necessity correct the tokenization (Figure 20).

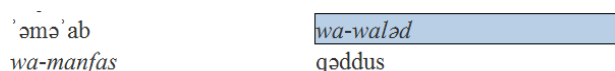


Figure 20. Automatically tokenized words

Once the tokenization has been implemented, it is still possible to edit/delete it, mark the graphic unit in the transliteration window, right-click and select ‘Remove tokens’ from the drop-down menu. A pop-up window shall warn you that if there is any annotation attached to the tokens of the graphic unit, it shall be removed, and the annotation shall have to be eventually reintroduced, so execute ‘Delete Tokens’ only if you are sure (Figure 21).

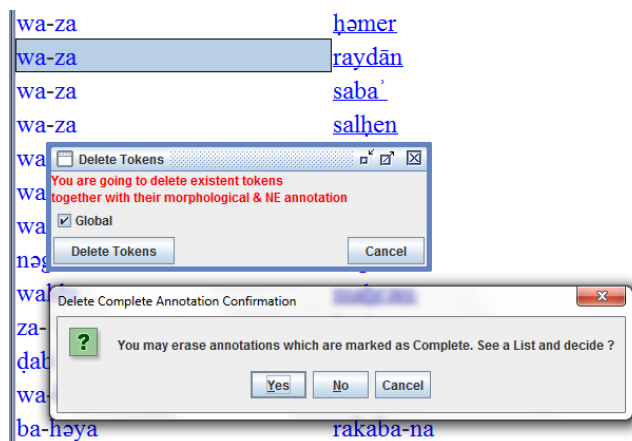


Figure 21. Deleting tokens after tokenization

If the same correction should be applied every time the same graphic unit has been tokenized in the same way, check the ‘Global’ box before performing the ‘Delete Tokens’ action. In this case, you shall be first asked if you want to see the list of all occurrences marked as ‘Complete’. Click Yes to be taken to the list; you can now select manually which instances should be removed (See Figure 22).

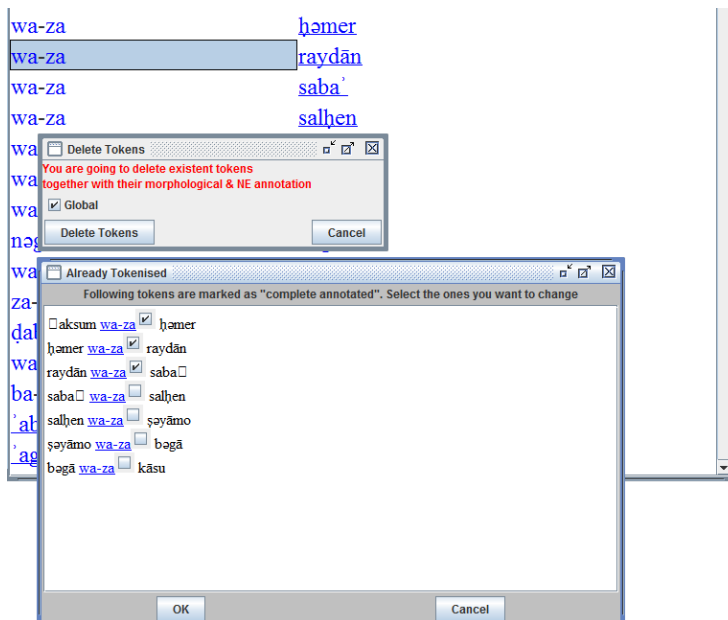


Figure 22

#### 4.2. Transliteration adjustment

In order to correct the automatically generated transliteration (morphological adjustment – without correcting typographic errors that would be also reflected in *fidal*, see §4.3), right-click the relevant graphic unit and choose ‘Edit Transliteration’ from the drop-down menu (Figure 23).

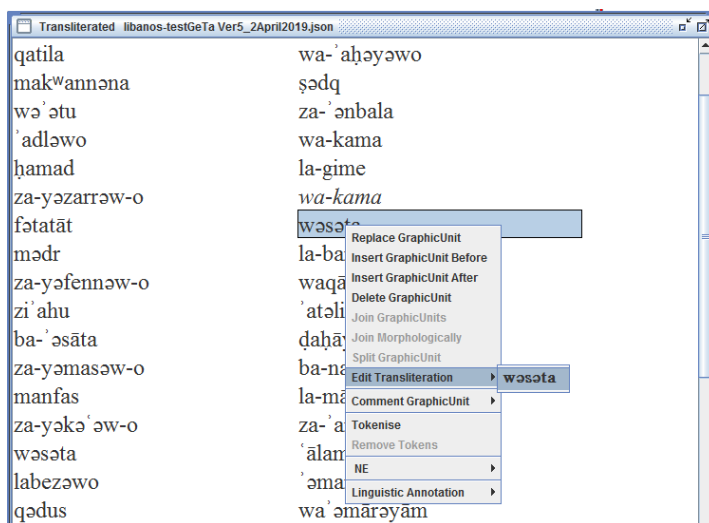


Figure 23

Clicking on the token you would like to edit, you will see a new pop up window. Choose first whether your changes must be applied only locally (to the unique occurrence) or globally (automatically correcting the transliteration of the token throughout the text if the exact string appears elsewhere). You can now select whether you would like e.g. to remove the sixth-order vowel and/or to introduce gemination (Figure 24).

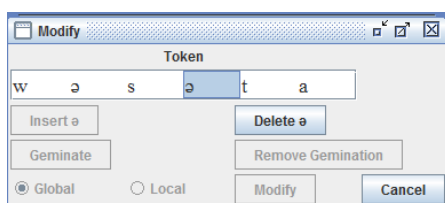


Figure 24. Editing transliteration.

The tool shall now show the preview of changes, highlighting the relevant occurrences (Figure 25).

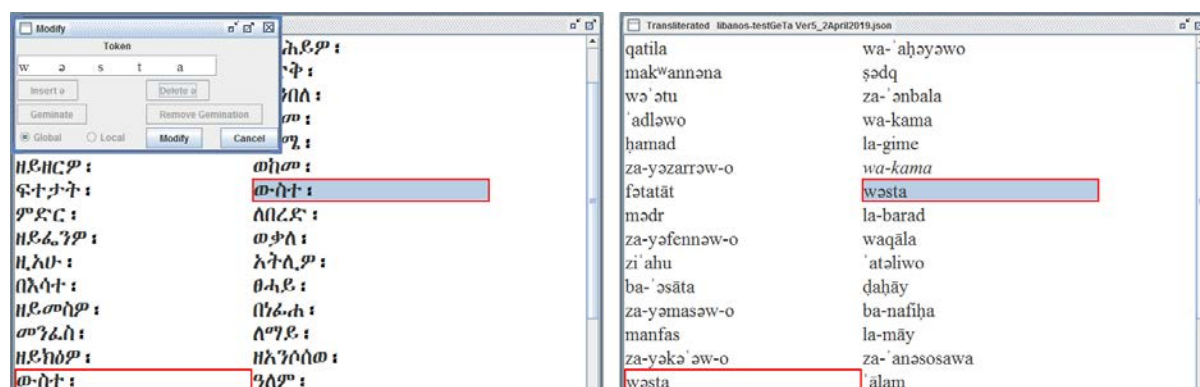


Figure 25

If you are satisfied with the result, you can click ‘Modify’ in order to introduce the changes, or ‘Cancel’ in order not to. Note that once the change has been performed, the automatically corrected tokens are NOT highlighted, so be careful when applying global corrections, especially for already annotated tokens.

The reverse operations (reintroducing the sixth-order vowel or removing gemination) are also possible.

The transliteration adjustments – including global ones – are possible also on already annotated tokens, even if they are marked as ‘Complete’.

#### 4.2.1. Workflow note

NB: You ought to first correct the transliteration in cases where it is relevant for correct tokenization. This happens for example in cases where assimilation of two consonants on the token boundary resulted in a single geminated consonant, represented by one symbol. E.g. in a graphic unit automatically transliterated as ‘*amanafas*’ one should first introduce the consonant gemination of *m* in order to neutralize the assimilation and be able to tokenize correctly ‘*am-man(a)fas*’, with the token boundary between the two consonants. Should one tokenize first (as ‘*a-manafas*’) and geminate afterwards, the result will be the wrong token splitting (‘*a-mmanfas*’).

#### 4.3. Edits to fidal text

If you need to delete, correct, or insert a word, right-click on the relevant (or closest) graphic unit in the transliteration window. From the drop-down menu, select ‘Delete GraphicUnit’ if you want to completely remove a graphic unit, ‘Replace GraphicUnit’ if you want to correct an existing word, or ‘Insert GraphicUnit Before’ or ‘Insert GraphicUnit After’ if you want to add a missing word(s) (Figure 26).

If you attempt any of these operations, a warning pop-up window shall appear, asking confirmation. Click ‘Yes’ only if you are sure. Note that if the unit you want to delete or replace had been tokenized and/or annotated before, you will have to re-tokenize and re-annotate it once it has been reinserted (the transliteration must also be manually readjusted). If the unit you want to delete or replace is associated with a Named Entity, or there is a line or page break attached to the unit, you will not be able to remove it before you had deleted the relevant links. A dialogue box saying ‘Graphic Units are part of a NE or contain PBs or LBs. Please check first the NE, PBs, and LBs and delete them’ shall appear. In your workflow, it is therefore always advisable to edit the *fidal* text before introducing annotations and, in particular, links.

Note that the delete and insert actions imply the re-indexing of the entire file in the background which may take up to several minutes depending on the length and complexity of your file.

NB If you are replacing or inserting a word you must provide the *fidal* version for it – this means you must activate the *fidal* keyboard on your computer in order to type in the text in *fidal*. You can insert special symbols employed by the GeTa tool (□ used for single omitted letters, e.g. in inscriptions, and ... used for ellipsis), which are absent from the *fidal* keyboard, by clicking directly on the corresponding symbols in the Insert box.

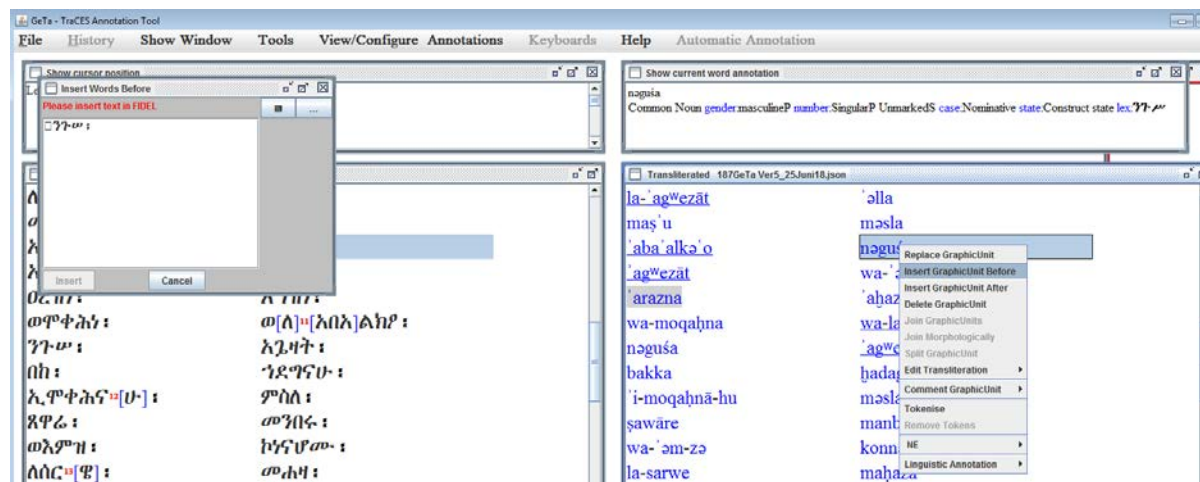


Figure 26

If a word in your edition is actually appearing in the source (manuscript or edition) written in two words and you would like to maintain the faithfulness of your *fidal* transcription, you can insert the separator character : after any letter. Right-click the relevant graphic unit in the *fidal* frame and select 'Insert Modify Fidal Separator' in the drop-down menu, the letter after which the symbol should be inserted, and type the symbol in the free text field (Figure 27). This change does not modify the transliteration or annotation.

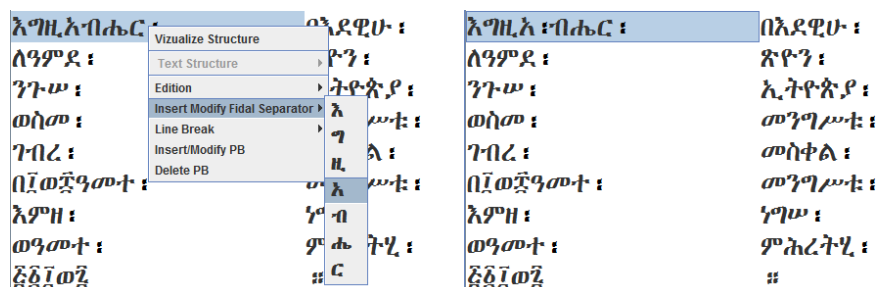


Figure 27

#### 4.4. Errors

Some changes, in particular those implying editing transcription, *fidal*, and introducing tokenization, may occasionally lead to index conflicts. These could make further editing of a graphic unit or a token impossible. To look for errors, users can use the 'Verify Index' in the Help menu; besides, the tool automatically checks for inconsistencies at loading, in doubt, close your document and reopen. If there are consistency errors they will be shown in the 'Error' frame (Figure 28).

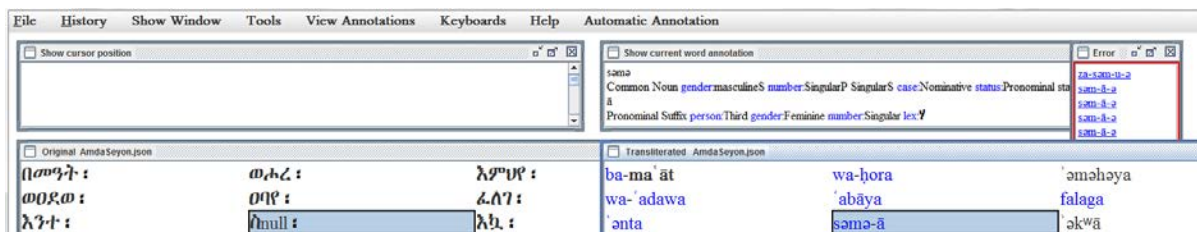


Figure 28

Click on the error to be taken directly to the relevant position in text and try to correct the error.

In some cases, the easiest way to correct an error is to replace the *fidal* graphic unit (as described in §4.3) and retokenize/reannotate it completely.

Besides index-relevant errors, some symbol combinations may sometimes produce conflicts that shall be experienced as an error by the user.

If all attempts at correcting (including replacing the graphic unit) failed, please contact the project team at [traces@uni-hamburg.de](mailto:traces@uni-hamburg.de) for assistance.

#### 4.5. Comments

At any moment of work, you can introduce comments, either by highlighting a graphic unit in a colour of your choice, or by typing in your comment, or preferably both, to draw attention to the commented unit (Figure 29).



Figure 29

You can enter free text in the Comments in the comment field – whether for the graphic unit, or for text division (see § 7 below), or for Named Entity (see § 8 below), yet make sure that the tool does not accept certain special characters, in particular ‘\*’. Note that Comments for single tokens should be provided in the ‘Deep Annotation’ frame during the ‘Linguistic Annotation’ process (see § 6.4 below and § 9.3 for an example of commenting).

#### 4.6. Editorial mark-up

It is possible in the GeTa tool to introduce minimal editorial mark-up, such as marking reconstructed or missing letters or words, and introducing the beginnings of pages and lines. Editorial mark-up is carried out on the original *fidal* version of the text, in the left tool frame.

Supplied passages should be marked by introducing square brackets around them. Select the first/last word of the relevant passage and right-click for the drop-down menu. Select ‘Edition’ and click on the first/last letter of the word (Figure 30). In the pop-up window, choose ‘[’ among the symbols to insert before the letter (or, respectively, ‘]’ to insert after the letter). You can also comment your choice in the free text field.

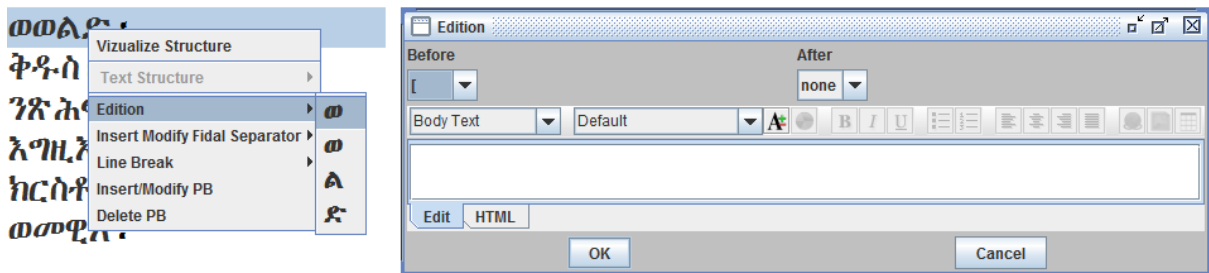


Figure 30

Other symbols that can be used include round brackets ( ), braces { }, and pointy (angle) brackets < >.

Use round brackets for abbreviation in the text, expanded by the editor, as in a(bc).

Use pointy brackets <abc> for letters erroneously omitted by the text, not restored by the editor.

Use braces {abc} for letters considered erroneous and superfluous by the editor.

Use square brackets [abc] for letters reconstructed by the editor and square brackets with ellipsis [...] for intelligible letters not reconstructed by the editor. See § 4.3 above on inserting ellipsis among *fidal* signs.

#### 4.7. Line breaks and page breaks

Line and page breaks can be inserted at the beginning of the document and after each symbol.

##### 4.7.1. Line breaks

Line breaks are numbered automatically beginning with 1; if you would like to change the number of the first line you can use the 'Set LB start number' option from 'View / Configure Annotations' menu (see Figure 31).

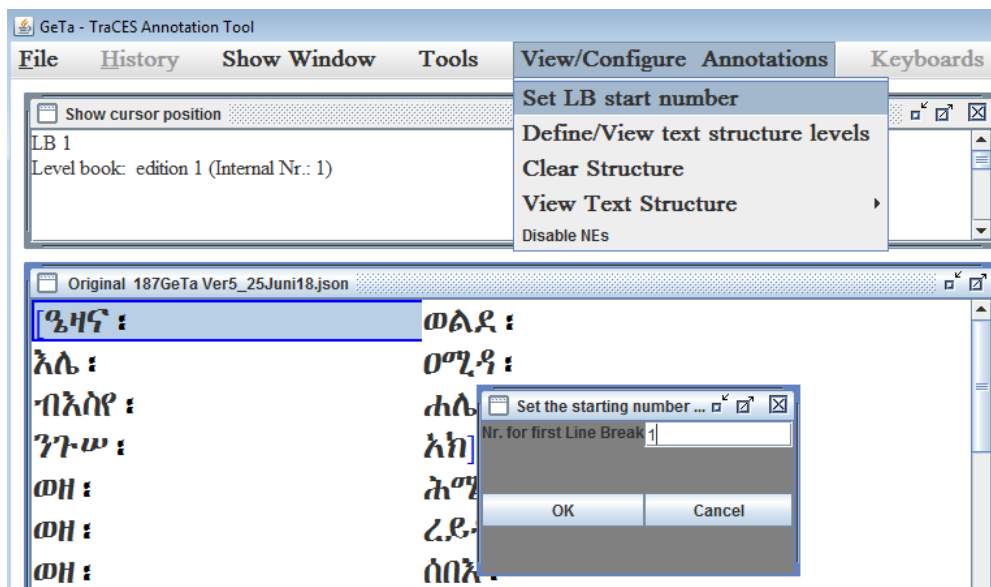


Figure 31

To insert the first line break, point at the first graphic unit in the fidal text edition and right-click for the drop-down menu (Figure 32). Select 'Insert LB Before'. You can remove the first line break by clicking the 'Remove LB Before' in the same menu.

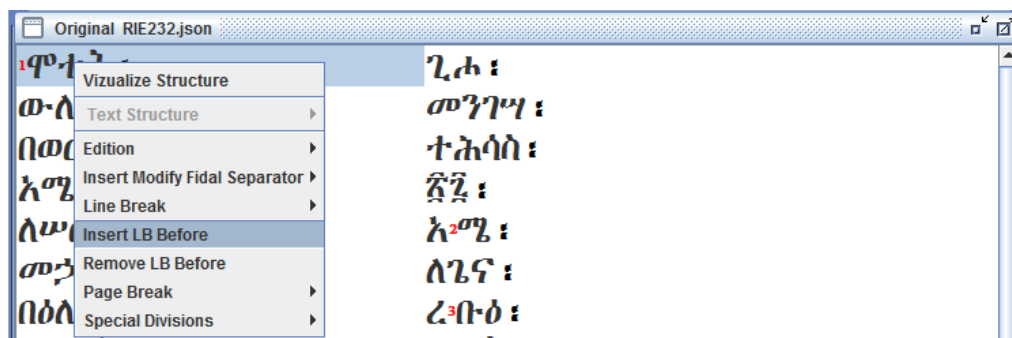


Figure 32

To insert subsequent line breaks, select the relevant graphic unit and click through the ‘Line Break’ – ‘Insert LB’ menu to select the character after which the line break must be inserted (Figure 33). You can also insert a line break before the separator by selecting the appropriate menu box. You can remove the line break(s) by selecting ‘Delete LB’ in the same drop-down menu.

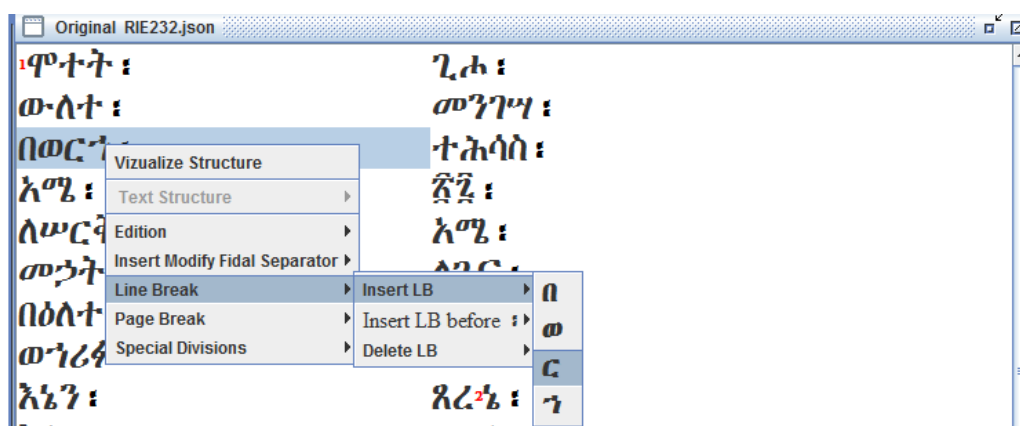


Figure 33

The line number is visualized in the “Show cursor position” frame above the text.

When you delete a line break, the following line breaks are renumbered automatically.

**Note that it is not possible at the moment to insert a new line break between the already assigned line breaks.**

#### 4.7.2. Page breaks

To introduce the page number, use the ‘Page Break’ submenu (Figure 34). For the first page, select the first word and check the ‘Before’ box. You can enter the page number in the free text field. For the subsequent breaks, choose the last word preceding the break, and insert the break after by checking the ‘After’ box.



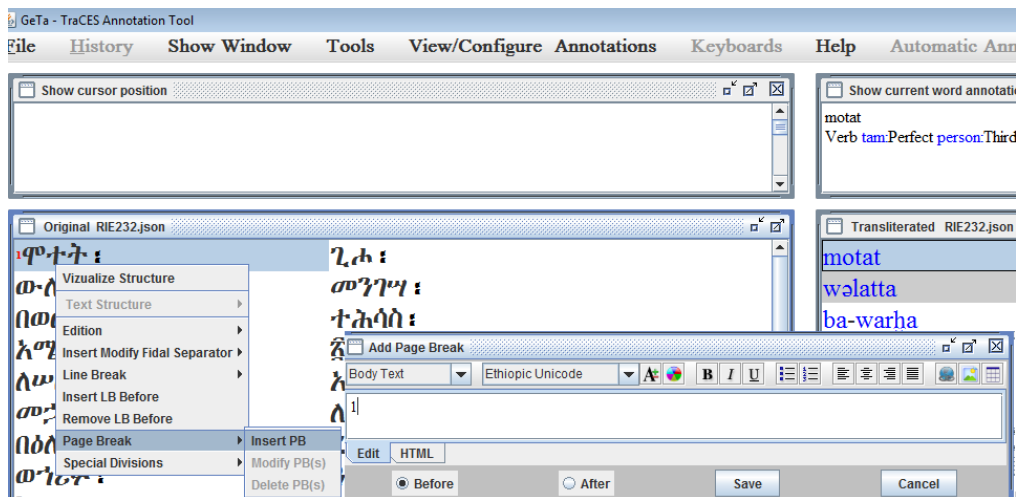


Figure 34

Pages breaks are always numbered manually, there is no automatic control of consistency, and additional care is required from the annotator.

If you insert page breaks according to several editions or manuscripts, make sure to consistently disambiguate the sequences by preceding the number by a version-specific keyword (e.g. Dillmann 1 / Dillmann 2 / Dillmann 3 for page numbering in an edition by Dillmann, Bezold 1 / Bezold 2 for page numbering in an edition by Bezold etc.).

You can modify and delete page breaks using the same Page Break submenu.

In case of several source editions it is possible to introduce several page breaks at the same point in texts. The presence of several page breaks is visualized in the text by the upper case number, corresponding to the number of breaks (see Figure 35). Pointing at the page break number in the text will show which page number annotations have been assigned in the “Show cursor position” frame.

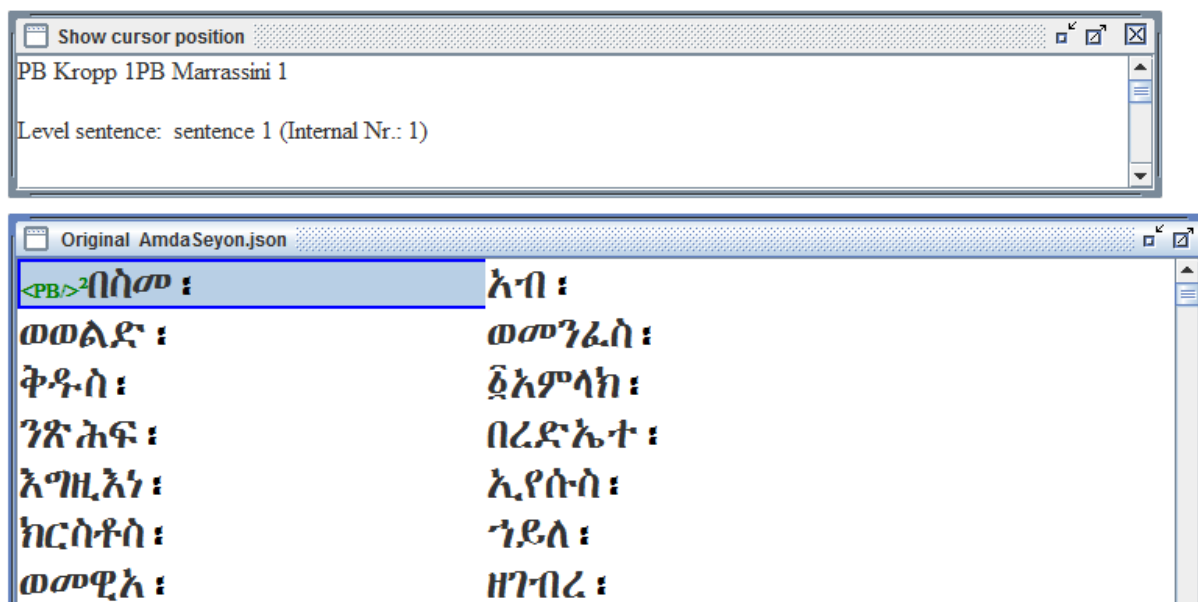


Figure 35

When deleting a page break, in case multiple page breaks are attached to the same graphic unit, you can select which page break to delete by ticking off the relevant box in the Delete Page Break(s) frame (Figure 36). Note that if you want to delete two or more page breaks you should click “Save” as many times, as many are the page breaks that are being deleted.



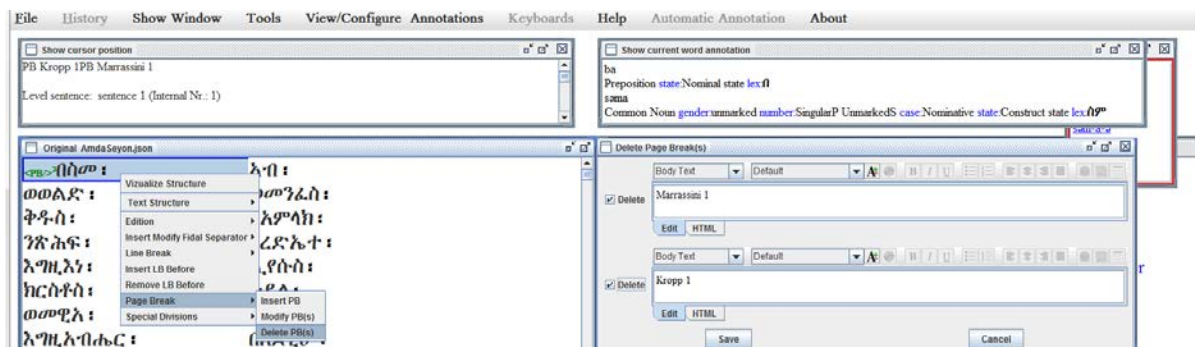


Figure 36

When modifying, you can simply modify the free text box and “Save”.

## 5. Working with GeTa: Help functions

Several add-ons help the annotators in their work.

### 5.1. Show window

Use the ‘Show Window’ option in the top menu to open the frames showing the linguistic annotation of the highlighted word, the text structure (divisions) annotation, and the index errors (Figure 37). (All the three frames open automatically when the tool is loaded; the function is useful if any of them gets closed or hidden by mistake.)

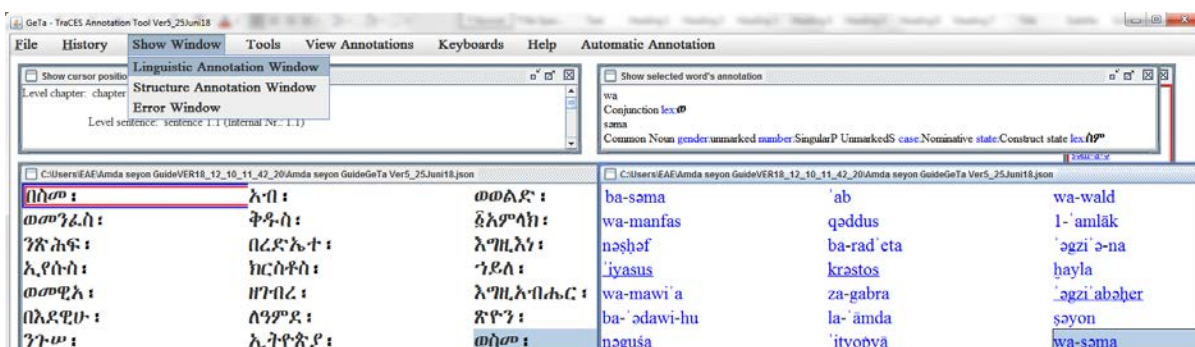


Figure 37

### 5.2. Tools: Search

At any moment of work, the Search function (under the ‘Tools’ menu) can be used. This way you can e.g. see whether a certain string – a graphic unit, token, or a sequence of units – is present in the text, and/or check how the graphic unit you are working on has been tokenized or annotated elsewhere (Figure 38).

You can search by entering the word (or part thereof) in the ‘Search word’ field both in *fidal* (except for when you look for an exact token match, see below) and in transcription; the results shall be shown in transcription.

In the results field, you can click on each result to jump to the relevant place in your text.

There are a number of search parametres that can be specified.

Note that the longer your text is and the more matches it contains, the longer the search may take, so it is always advisable to limit the search by selecting the appropriate search parametres.

### (1) Context

You can define how many words (graphic units) before and after the graphic unit containing the search string should be displayed. If you do not enter any number, the default context of 10 graphic units before and 10 graphic units after shall be applied.

### (2) Scope of search

In a drop-down menu, you can select whether the search should be carried out in the entire document (choose 'global') or only from the cursor position downward (choose 'From selected index').

### (3) String matching

In a drop-down menu, you can select whether the text string you enter in the 'Search word' field must correspond to an entire graphic unit (select 'identical'), to an entire token (select 'is token'; note that in this case you must search in transliteration), whether it must appear at the beginning (select 'at the beginning') or the end of a graphic unit (select 'at the end'), or anywhere (select 'contained in').

### (4) Annotation

If the linguistic annotation is irrelevant for your search, whether because it is inexistent or incomplete, or you want to find the string with all the possible annotations, in the annotation drop-down menu select 'no PoS annotation'.

If you want to look for your string only when it has been associated to a specific Part of Speech, select 'PoS annotated with' and then choose the appropriate PoS tag in the 'POS Gə'əz' field.

If you want to only find cases where no linguistic annotation has been provided for your search string, select 'not annotated'.

The figure below shows the search results for the string ስመ, with the context limited to five graphic units before and after, in the entire text ('global'), that can occur at any position within the graphic unit ('contained in') and that has been annotated as 'Common Noun'.

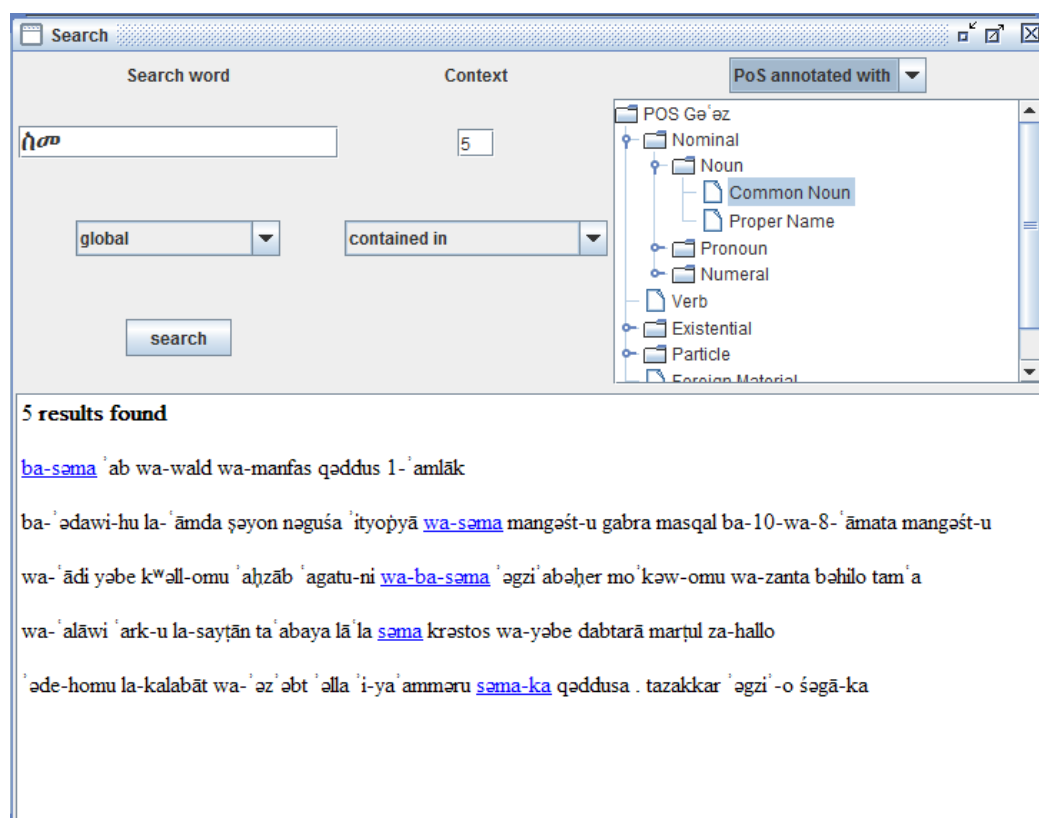


Figure 38. Search window

### 5.3. Tools: Undo function

The 'Undo' function is currently under development.

### 5.4. Tools: Statistics

You can use the 'Statistics' function in the 'Tools' menu to see where you are at currently as far as the tokenization and the linguistic annotation are concerned (Figure 39). An additional graph is generated to show the share of graphic units consisting of more than one token.

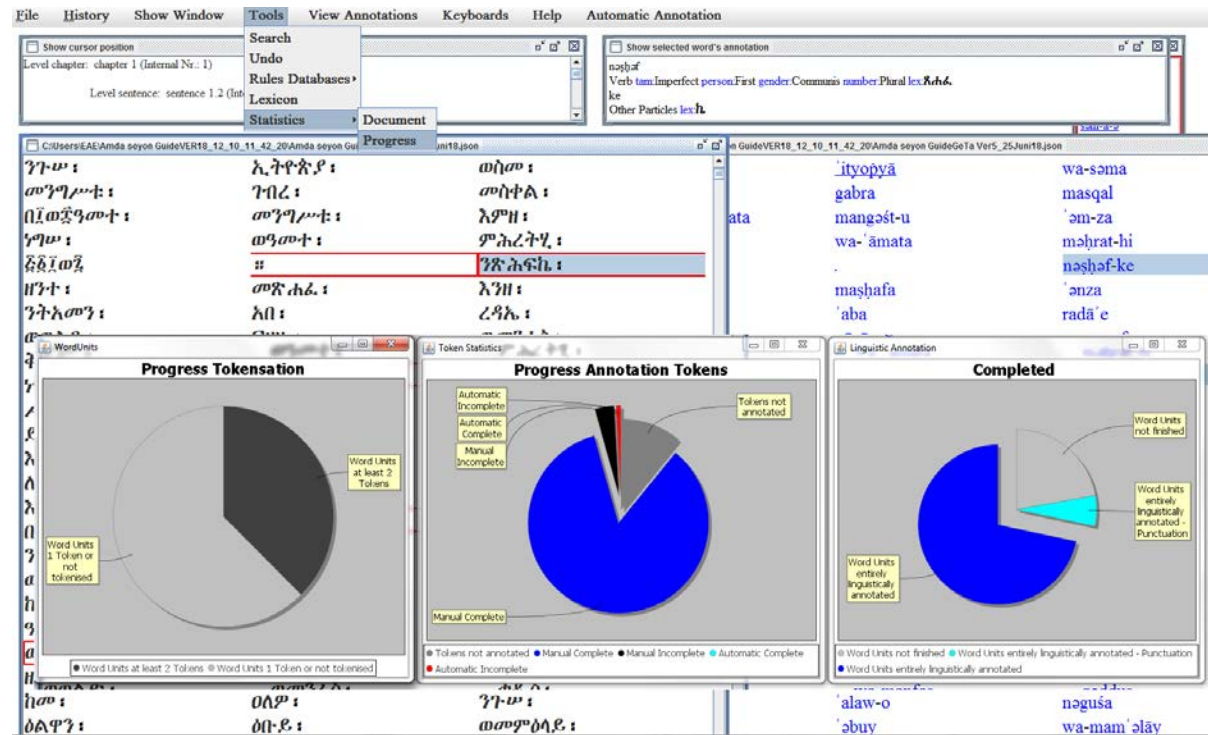


Figure 39

### 5.5. Help: Linguistic Annotation Colours

Use the 'Linguistic Annotation Colors' button of the top 'Help' menu to visualize the colour coding overview, as summarized in Table 3. Colour and format coding in the GeTa tool.

Table 3. Colour and format coding in the GeTa tool

Font colour / format in transliteration	Meaning
TNR black	Transliterated text
TNR black italic	Globally (automatically) tokenized unit
TNR black bold	Manual linguistic annotation provided
TNR red bold	Global (automatic) linguistic annotation
TNR red	Global (automatic) linguistic annotation complete
TNR blue	Manual linguistic annotation complete (final)
TNR underlined	Named entity
TNR green bold	Nouns (after 'Verify Nouns' check)

### 5.6. Help: Verify Index

Use the 'Verify Index' function if you have carried out index-relevant corrections (deleting or inserting word unit, tokenizing, undoing) to make sure that the file is error-free. (May take several minutes, depending on the length of your file).

### 5.7. Help: Correct Line Break Nr

Use this function if you want to increase or decrease all line numbers in your text, *apart from the very first line number* (see Figure 40).

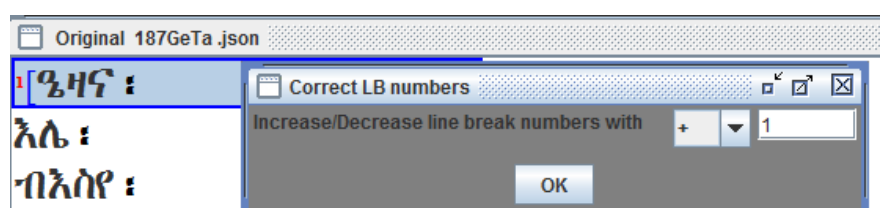


Figure 40

## 6. Working with GeTa: Linguistic annotation

The linguistic annotation is in the focus of the work with the GeTa tool. It is best carried out once your text has been corrected and tokenized (as then the 'global' function can best be applied, speeding up the annotation process), but naturally the user can also choose to normalize/tokenize/annotate graphic unit by graphic unit.

In order to insert the linguistic annotation, right-click the graphic unit in the transliteration text frame and select 'Linguistic Annotation' in the drop-down menu. If the unit consists of several tokens, you will now be able to choose the token you would like to annotate (see Figure 41).

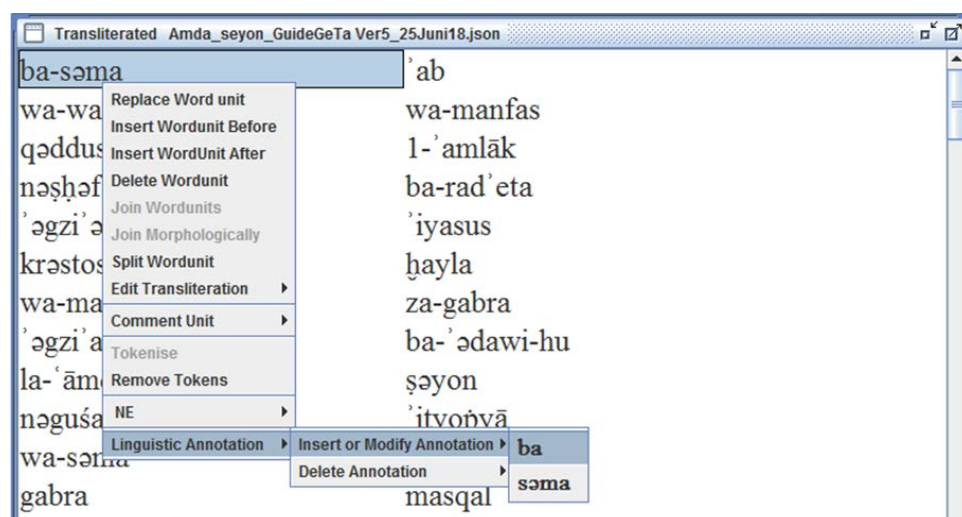


Figure 41

### 6.1. Basic annotation: POS

In the Linguistic Annotation frame, you can now select the relevant PoS (Figure 42) – preferably by going down to the lowest possible branch of the POS tree.

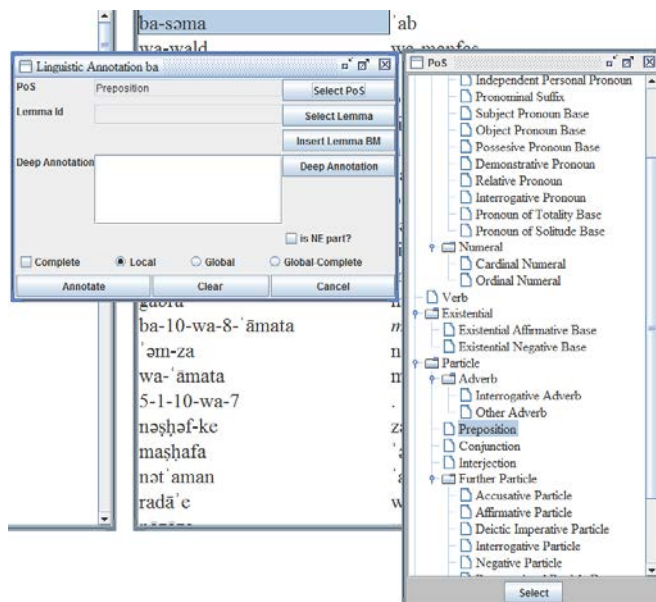


Figure 42

You do not necessarily have to fill out the ‘Deep Annotation’ fields immediately – for some purposes it might be sufficient to simply assign all PoS. The ‘Deep Annotation’ features are not available if you have selected an upper-level category from the PoS tree where lower levels are available.

You can select whether the annotation you selected should be applied only locally (‘Local’ is checked as default) or globally (check ‘Global’ at the bottom of the frame). Global annotation shall be applied to all identical tokens throughout the text. It is best applied only if you are sure that the same token can only appear with the same or very similar function. After the action, all globally annotated units shall appear in red (Figure 43), reminding you that you have to confirm the correctness of the annotation and eventually modify the features (see also Table 3. Colour and format coding in the GeTa tool).

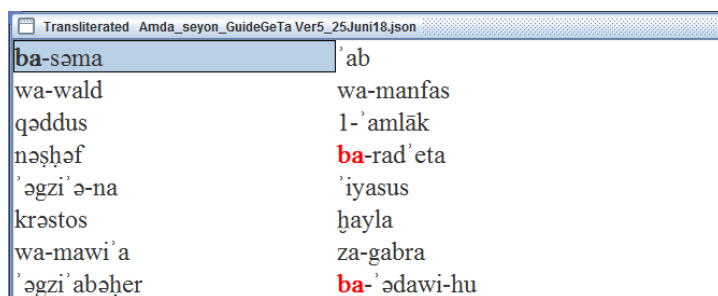


Figure 43

If you want to correct or continue your annotation at a later point, repeat the first step as shown in Figure 41, this will take you back to the Linguistic annotation frame.

## 6.2. Linking to dictionary

In the ‘Linguistic annotation’ frame, you can also provide the link to the dictionary file, to uniquely define the lexical value of each token. Again, for some purposes this may be the only step you will choose to carry out.

(1) If the lemma is already present in the dictionary database, the easiest way to link it is through the ‘Select lemma’ option. You can scroll through the alphabetical list, or start typing (use Ethiopic keyboard) your lemma in order to get to the correct place in the alphabet quickly. If there are several homographs, select the fitting lemma in the alphabetical list, check if the translation summary



(originating mostly from the *Lexicon linguae Aethiopicae* by Dillmann, and thus in Latin) fits your needs, and finally click ‘Assign’. Now you will see the Lemma ID in the relevant field of the annotation window (Figure 44).

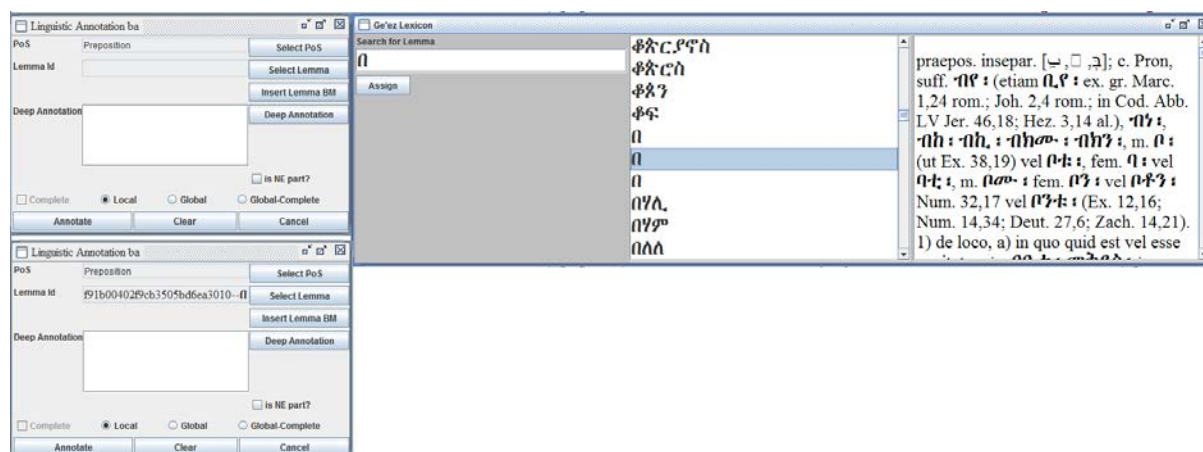


Figure 44

(2) Alternatively, or if you fail to find the lemma through the ‘Select lemma’ option (a live API connection is envisaged but is not yet in function), you can search for the lemma directly in the lexicon web app, <http://betamasaheft.eu/Dillmann/> (Figure 45).

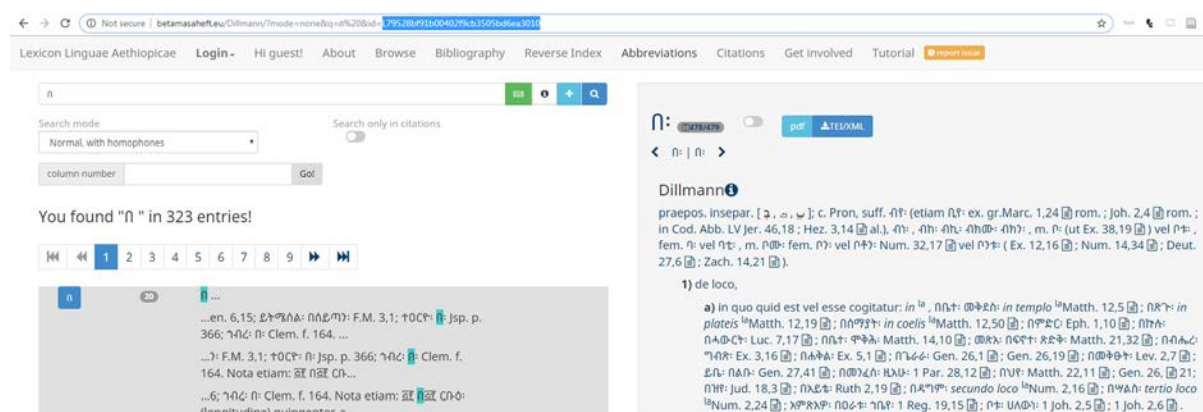


Figure 45

Subsequently you can link by providing the ID (highlighted in Figure 45) in the ‘Insert Lemma BM’ field (Figure 46): enter the lemma ID in the ‘Search for ID’ field, click ‘Search Lemma’, and, if you are happy with the search result, click ‘Assign’. This will bring you to the same result as in (1).

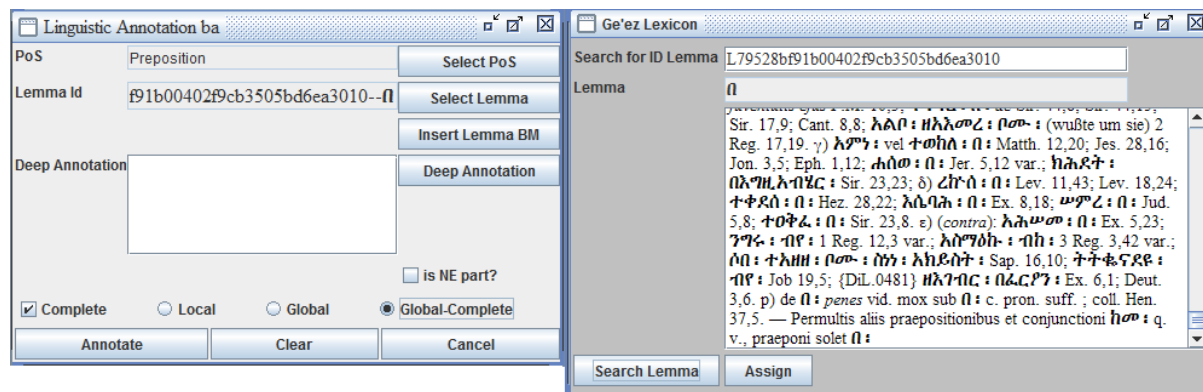


Figure 46

(3) If you cannot find the lemma in the online application either, you can create a new dictionary entry if you have the access rights, or contact the project team with the request by choosing the ‘report issue’ button in the top menu of the web app (see Figure 45) or emailing the editors.

### 6.3. Completing annotation

Once you believe you have entered all the data expected (e.g. for the conjunction *w*, no additional features are requested, therefore even if you aim at a detailed morphological ‘deep’ annotation, nothing must be entered in the ‘deep annotation’ field) you can tick off the ‘Complete’ box to signal that the annotation is final to your opinion. You can also choose the ‘Global-Complete’ option (see Figure 46) if you are sure that this token must be annotated in the same way throughout. Note that this makes all occurrences appear in blue as if they were manually checked, so only use the ‘Global-Complete’ option if you are absolutely sure (see Figure 53).

You can always see the linguistic annotation in the ‘Show current word annotation’ frame above the text (Figure 47).

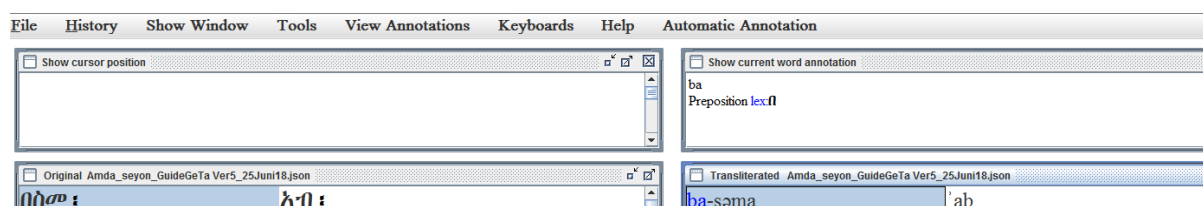


Figure 47

If you have realized that the annotation was erroneous, and it is too late for an Undo, you can always modify or even delete annotation by selecting the relevant option from the drop-down menu (see Figure 48). In the delete frame, you can also choose, once again, whether your changes are to be applied locally or globally.

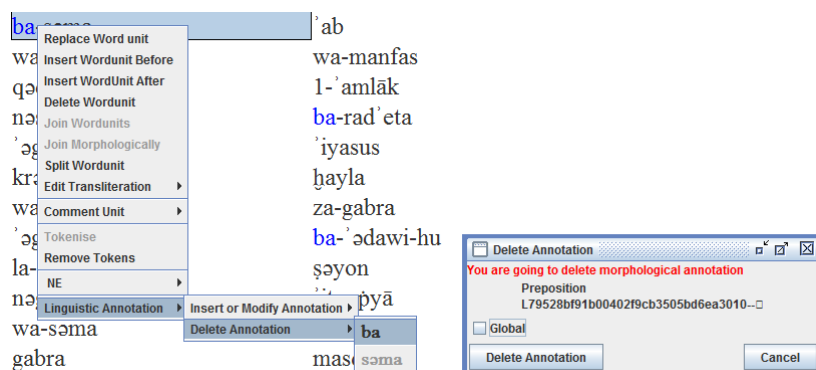


Figure 48

### 6.4. ‘Deep’ annotation

In the GeTa tool, detailed morphological annotation is referred to as ‘deep annotation’. Naturally, for some PoS (conjunctions or particles) no additional features are requested, thus the POS annotation as described above is the maximum level of detail available. For other POS, such as nouns or verbs, a varying number of features is available and can be assigned with the help of the tool (see Table 1 for the PoS and respective features).

After assigning the part of speech, in the ‘Linguistic Annotation’ frame (see Figure 44), select the ‘Deep Annotation’ button. You will see the list of available features, depending on the part of speech selected (see Figure 49 for noun, Figure 50 for verb). Features that are not applicable are (or become, depending on your choices in the annotation) deactivated and, if at all shown, appear in grey.

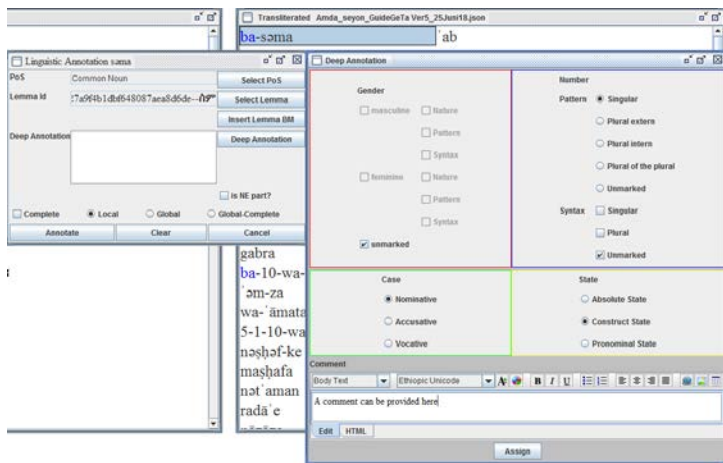


Figure 49

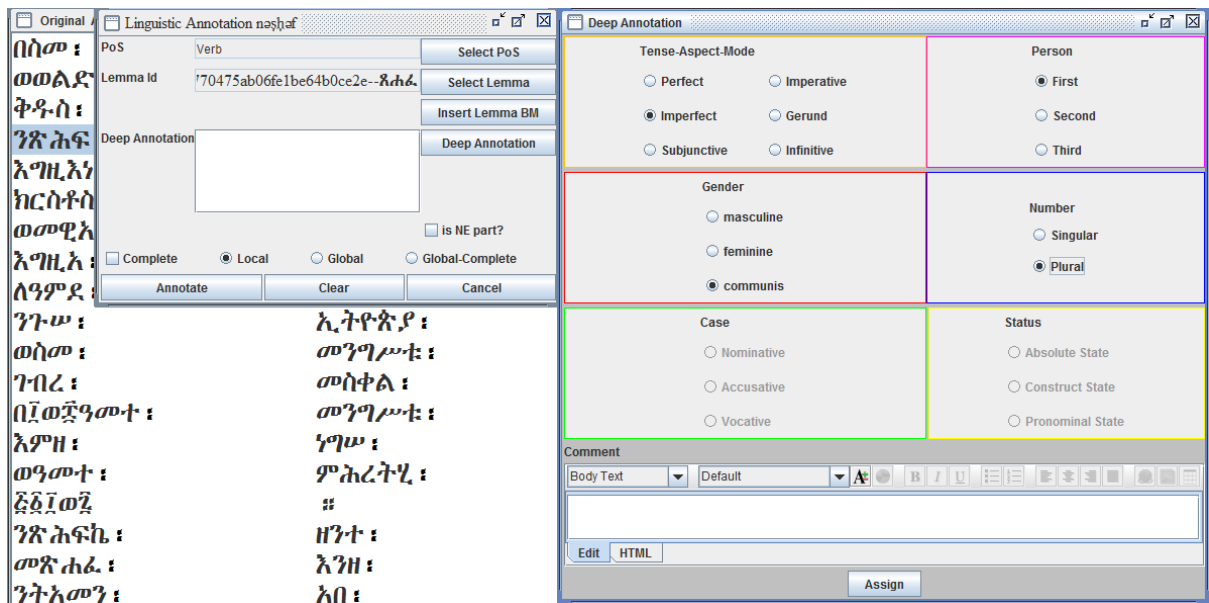


Figure 50

For numerals, you can additionally provide the Arabic number in the 'Logogram' field (Figure 51).

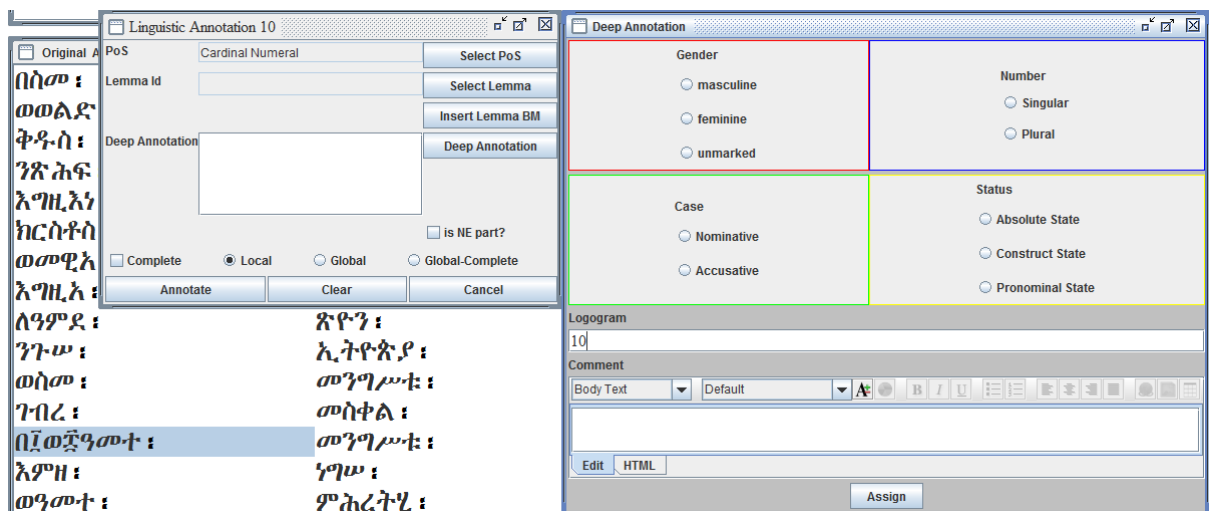


Figure 51



Select the features that apply in each specific case. The most widespread options are marked as default. Click ‘Assign’ in the ‘Deep Annotation’ frame to save the features, and finally ‘Annotate’ in the ‘Linguistic Annotation’ frame.

Note that if a noun or a verb are used as a name or part of name you can tick ‘is NE part’. This will allow you to eventually search for all words (whether common nouns or verbs) that have been used in names in Ethiopic (Figure 52). This feature is distinct from linking the named entities to the authority files.

Figure 52

## 6.5. Local and global changes

Do not forget to select whether you are annotating this single instance (by selecting ‘Local’) or all instances of this token (by selecting ‘Global’). In the former case, the instance shall be highlighted in bold, to show that the linguistic annotation is present but possibly not yet complete. In the latter case, all automatically annotated instances will be additionally highlighted in bold red font (see Table 3. Colour and format coding in the GeTa tool), to remind you that the annotation should be checked manually.

Use ‘Complete’ (or ‘Global-Complete’) if you are sure you provided the whole set of data and you will not be changing anything in the annotation. In such case the instance (or, respectively, all instances) of the token shall be highlighted in blue.

The ‘Complete’ marking (evidenced as blue colour) does not deactivate the local annotation function and if necessary, the annotation can still be modified or completed. Yet, the ‘Complete’ marking does prevent the tool from applying global changes, except for the transliteration (globally applied corrections to transliteration will also be applied for the tokens marked as ‘Complete’).

You can always apply ‘Global Complete’ annotation to, e.g.,

- particles without features, such as  
 $ssa$  = Other Particle **lex:ሰ** or  
 $nu$  = Interrogative Particle **lex:ከ**
- conjunctions  
 $wa$  = Conjunction **lex:ወ**
- most verbs in the perfect form, e.g.  
 $kona$  = Verb **tam:Perfect** **person:Third**  
**gender:Masculine** **number:Singular** **lex:ከከ**  
 $yabe$  = Verb **tam:Perfect** **person:Third**  
**gender:Masculine** **number:Singular** **lex:ብህላ**

NB: there are exceptions here, e.g. the string ንብረ may correspond to either a Verb (**tam:Perfect** **person:Third** **gender:Masculine** **number:Singular** **lex:ንብረ** ID L0eb8ac47c3fe4e9db66fdff97a061550) or a Common Noun (**gender:masculineN** **number:SingularP** UnmarkedS **case:Accusative** **state:Absolute** **lex:ንብረ** ID La1dde44e977e4102ad0073db0a0588c4): in this and other such cases no ‘Global’ changes are advisable

Figure 53

## 7. Working with GeTa: Text structure

One of the annotation levels possible with the GeTa tool is marking up the text structure. This step can precede the linguistic annotation, or also be the only action carried out with the tool, if requested.

### 7.1. Introducing levels

First, the structure levels must be defined. Up to four hierarchical levels are possible. They can be specified by selecting 'Define/View text structure level' in the upper menu (Figure 54).

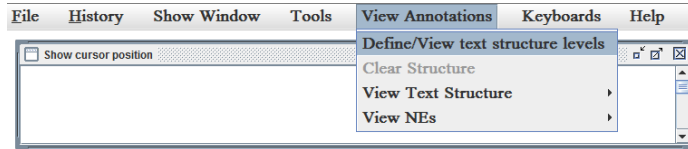


Figure 54

Up to four levels can be defined. The user can select the division subtype from a menu (book / chapter / part / section / poetic verse line / sentence / segment), choose a general name the division shall be saved with (it can but does not have to coincide with the subtype) and set the first value of the numeric sequence (Figure 55).

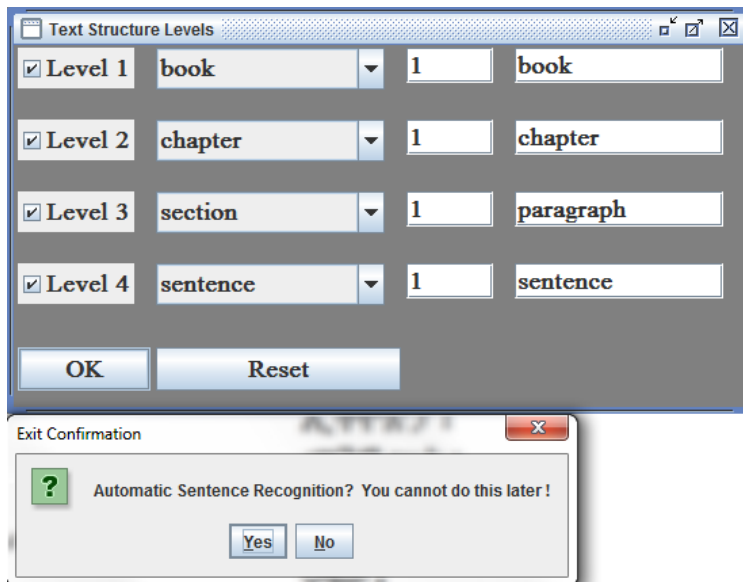


Figure 55

The default value is '1', but if you are annotating a piece of text beginning with a higher number you can specify it here.

Note that once the levels have been defined and assigned, you cannot easily modify them any more. E.g. if you initially only select 2 levels, book and chapter, assign the spans, and later want to additionally introduce the levels for sentence or verse you cannot do it without removing the existing divisions.

Once you have selected all levels you can click 'OK'. The tool will then offer you to carry out the automatic sentence recognition (Figure 55). If you select 'yes', all sentences limited by the punctuation sign ':' shall be automatically recognized and numbered, using the lowest level category in your selected text structure. Thus, if you want your sentences numbered, the lowest level should be defined as 'sentence'. If you accept the automatic sentence numbering and later discover that one sentence should be split into two, or merged with another, you can manually reassign the spans as described below. If sentence numbering should not be carried out in your text, select 'No' to save your text structure levels. The automatic sentence recognition can only be carried out once, at the very

beginning of the text division mark up. If you select ‘No’ you cannot carry out the automatic sentence recognition any more.

## 7.2. Assigning spans

Once the text structure levels have been defined, you can assign them to the text.

If you have opted ‘Yes’ for automatic sentence recognition, your sentence-spans have already been assigned, which you can see by the colour brackets highlighting of the first and last unit of the span. The ‘show cursor position’ frame informs you, in which sentence you currently are (Figure 56).

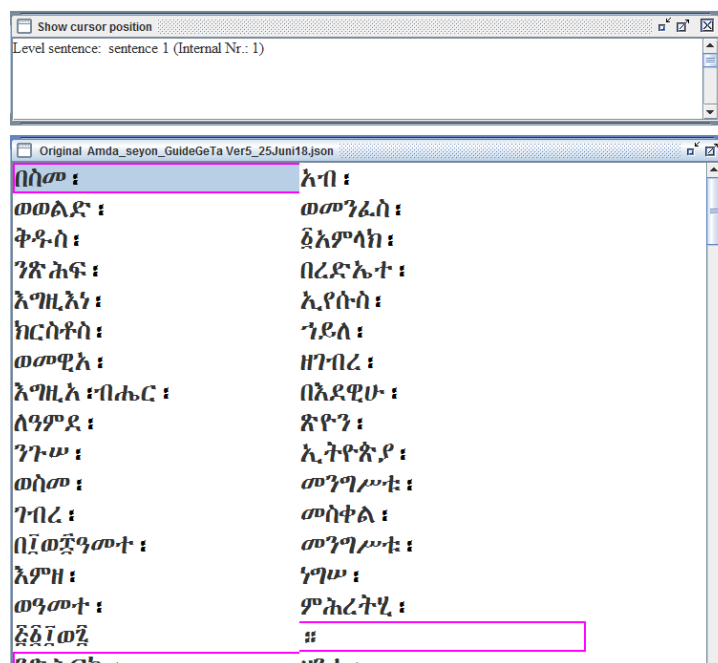


Figure 56

To introduce new spans, select the first word of the span in the *fidal* text frame and right-click to open the drop-down menu. Choose the ‘Insert Division’ option from the menu (Figure 57).

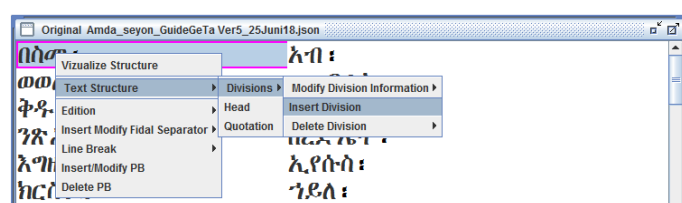


Figure 57

Now fill the fields in the ‘Insert Structure Annotation’ frame. The first graphic unit should be filled automatically (this is the word you selected). Scroll to the last graphic unit (in most cases, sentence punctuation signs) of the division you would like to select and click ‘Last GraphicUnit’ to fill in the corresponding box in your frame.

Select the division subtype (from the types you defined) and the name you would like to be associated with the division (free text, can also be ‘Prologue’, ‘Introduction’, etc.). The Internal number is the one assigned automatically, yet you can also provide a free text number which can be different. In the ‘Style’ field you may select the genre keyword best fitting with the text (from a drop-down list). In the ‘Creator’ field you have the choice between ‘Edition’, meaning that the structural unit is already present in the edition your work is based upon, and ‘Annotator’, if you are the one who have introduced the structure level. As always, you are free to provide Comments in the free text field (Figure 58).

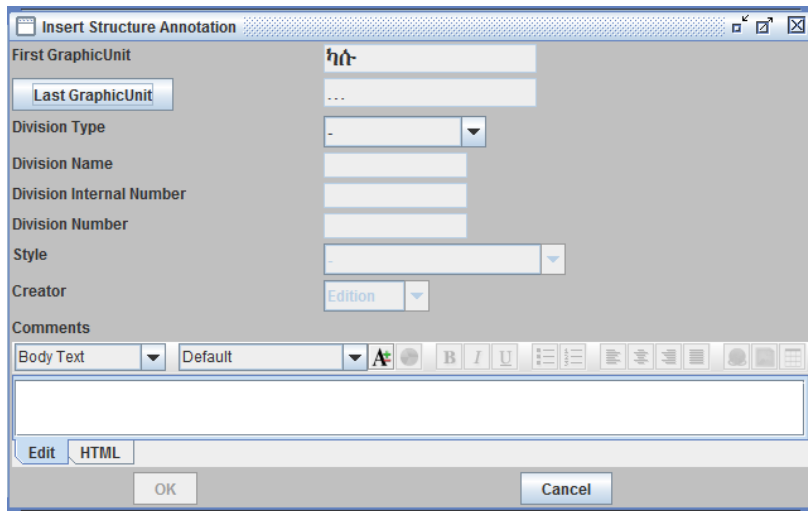


Figure 58

The Division Name, Division Number, Style, and Creator fields, as well as Comments, can also be modified later by selecting ‘Modify Division Information’ (see Figure 57).

Make sure you select all levels you want to be visualized under ‘View Text Structure’ (Figure 59).

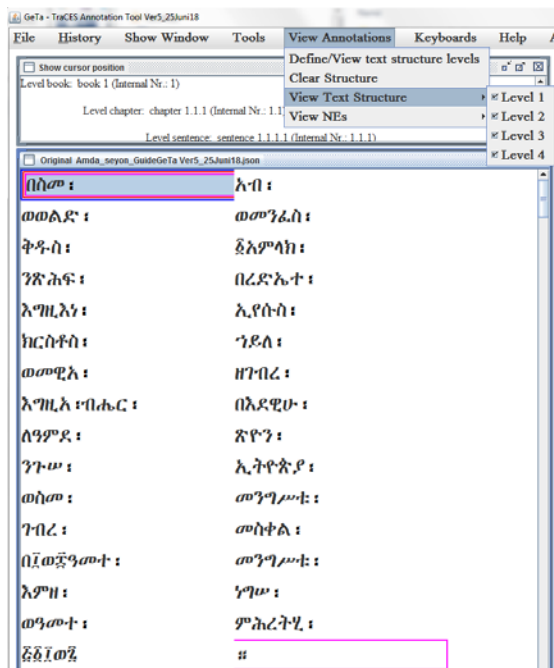


Figure 59

Continue the steps as described above to introduce further divisions. Note that the spans may not overlap; if you attempt to introduce an overlapping span you shall receive a warning message from the tool.

If you need to introduce a custom overlapping span, use the Insert Special Division option in the drop-down menu (See Figure 60).

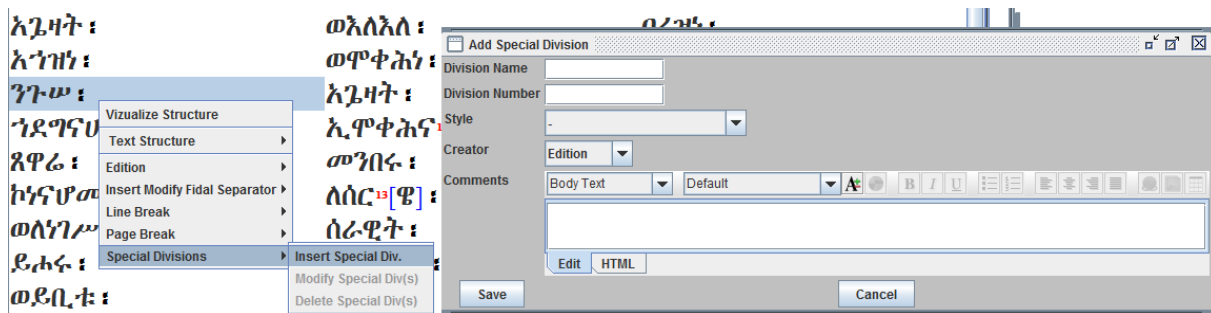


Figure 60

### 7.3. Editing spans

The easily editable information is, as shown above, Division Name, Division Number, Style, and Creator fields, as well as Comments. If you would like to modify the span itself, make it shorter or longer, you have to delete it by using the 'Delete Division' option and manually recreate the new spans.

You can also use the 'View Annotations' menu to clear the entire document structure ('Clear Structure' command): you will then have to redefine and reassign all text structure levels from the beginning.

## 8. Working with GeTa: Named entities

The GeTa tool allows for marking up all named entities, that is all real-world objects, such as persons, locations, titles, and dates.

### 8.1. Places, persons, works

All unique persons or places can be linked to the authority files stored on <http://betamasaheft.eu/> (BM), a database that should contain information on all persons and places relevant for Ethiopian and Eritrean manuscript culture.

In the transcription frame, select the word unit(s) that correspond to a named entity and right-click for a drop-down menu. Select NE -> Insert / Modify NE (Figure 61).



Figure 61

In the 'Insert NE' frame, tick off the tokens to be associated with the named entity, and then select the named entity type (Figure 62).

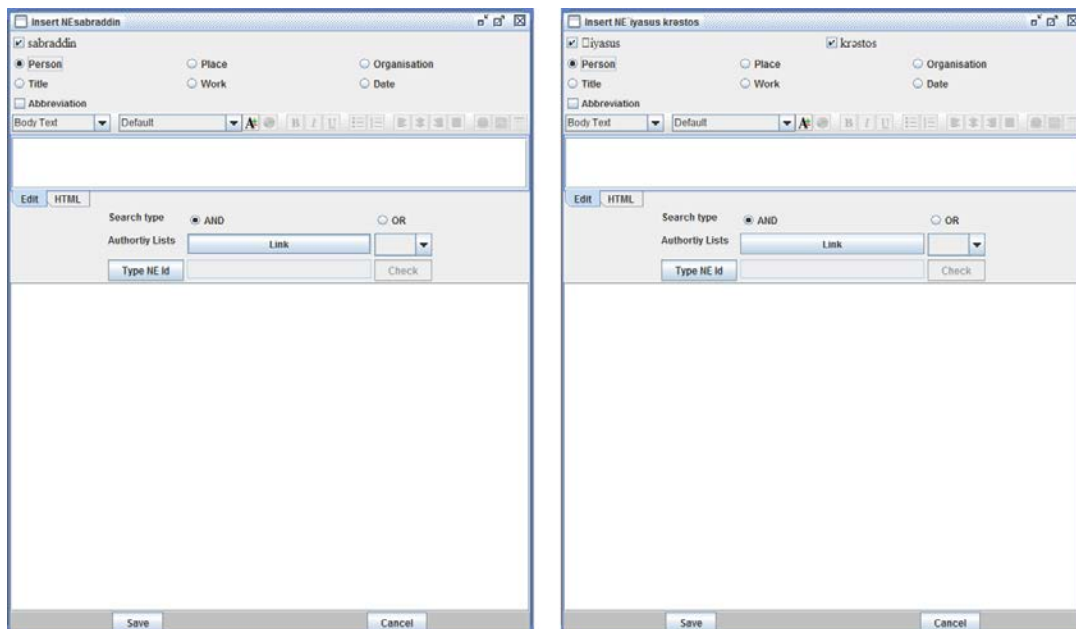


Figure 62

You can then try to search for the named entity directly in the BM database by clicking the 'Link' button. If an item (or more than one item) with identical spelling is present there, it (they) should pop up in the search result(s).

Click on the results box to check whether there is a list of options available (Figure 63).

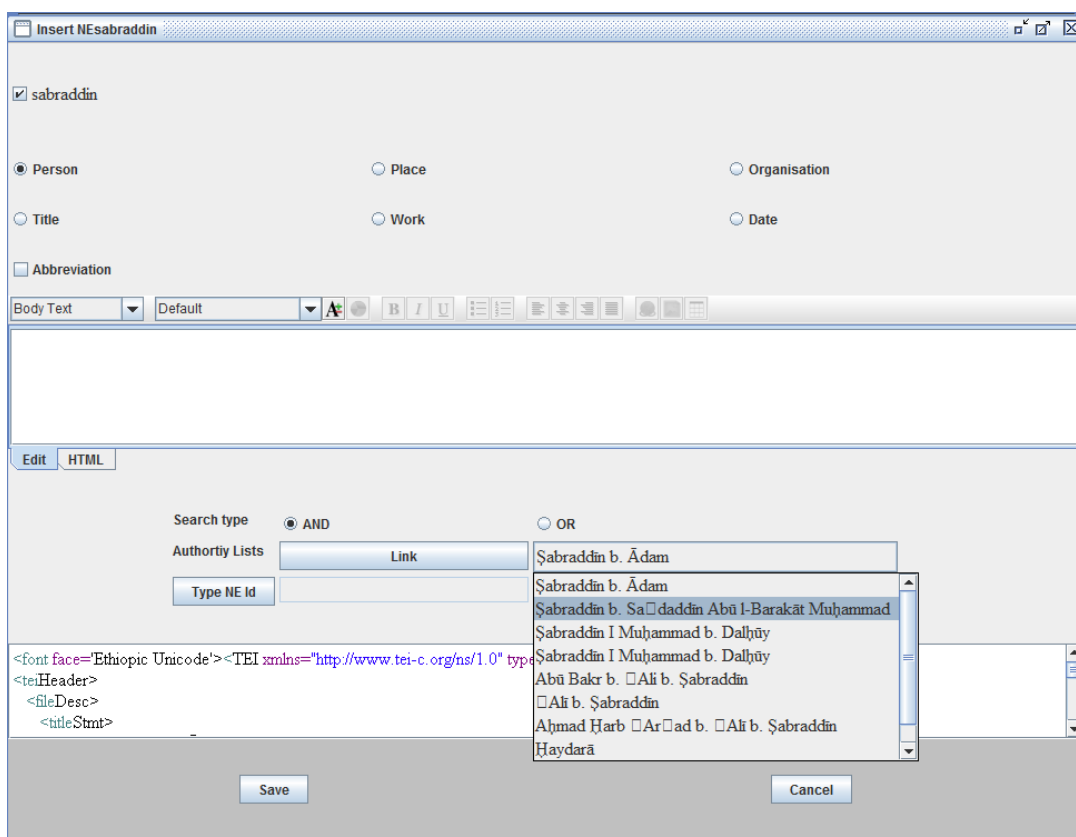


Figure 63

If the correct option is available, select it from the list and check the associated metadata in the preview field below that the selection. If you are satisfied with the result, click 'Save' to associate the

named entity with the authority file. The name in the text is now underlined to show that a NE has been associated.

If the search has been unsuccessful (no results, or results do not contain the person or place you were looking for) it probably means that there is a difference in spelling. In particular, the search for multi-word named entities via the tool may come up empty (Figure 64).

Figure 64

In such case, go directly to the BM database and use the simple or (recommended) advanced search mask there. You can use the transliterated version or also the simplified transliteration without special diacritical symbols. Not all authority files have yet been supplied with the fidal name forms, therefore do not rely on a search in fidal (even if it is possible). For internationally known names you may also try the standardized English name form (Figure 65).

Figure 65

Once you have found out the ID of the authority file, you can insert it directly into the 'Insert NE' mask of the GeTa tool by clicking the 'Type NE Id' button (Figure 66). Click 'Check' to see the results in the preview field and 'Save' if you are happy with the result.

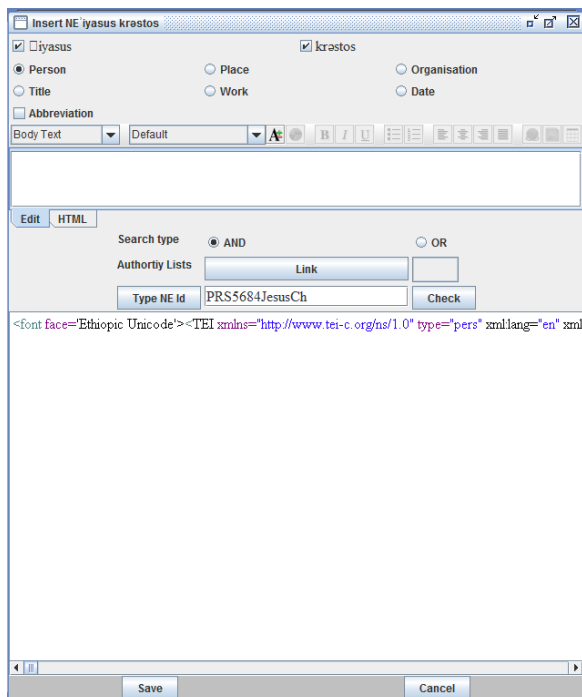


Figure 66

If the search in the BM database has not provided the expected result, please contact the project team to create a new entry (or create a new XML file directly if you have the necessary access rights).

## 8.2. Global annotation

If the same name appears throughout the text with the same meaning (always associated with the same person/place/etc.), you can now automatically associate the same authority file to all the other occurrences of the name by selecting NE->Annotate Global in the drop-down menu (Figure 67).

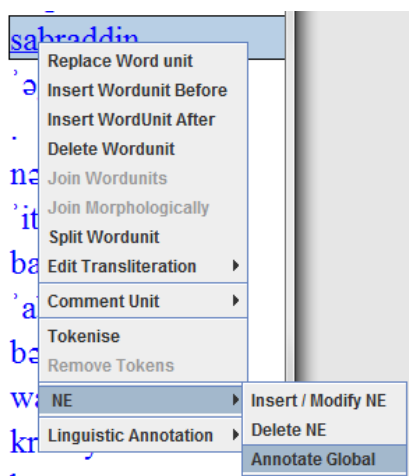


Figure 67

In the new 'Insert Global NE' frame, choose whether you want to apply your changes to the text before or after the cursor point and click 'Show'. You will now see all occurrences of the name in question in the text (Figure 68).



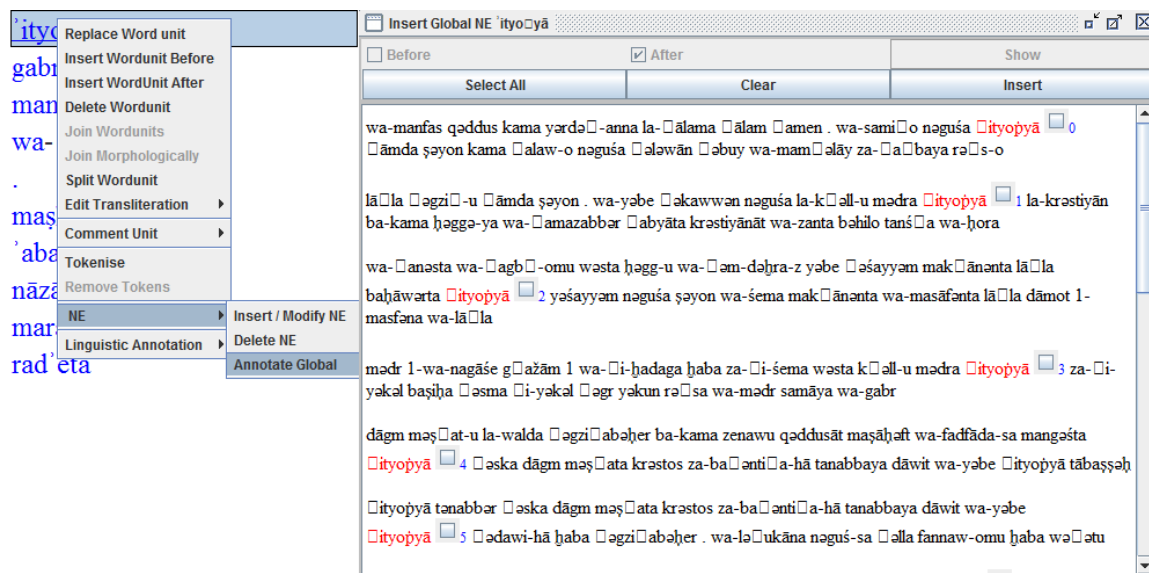


Figure 68

Tick off the occurrences you want to associate with the authority file (or use the ‘Select All’ button if all occurrences stand for the same named entity) and click ‘Insert’.

### 8.3. Editing NE annotation

Single links to the authority file can at any point be edited by returning to the Insert / Modify NE option and providing a new ID, or unselecting/selecting tokens. You can also remove the link to the authority file completely by choosing ‘Delete NE’ in the drop-down menu.

### 8.4. Titles, dates, abbreviations

Named entities of the type ‘titles’ (offices, terms of address, etc.) or dates (absolute or relative) do not have to be associated with an authority file; you can use the free text field to provide additional information if needed (Figure 69).

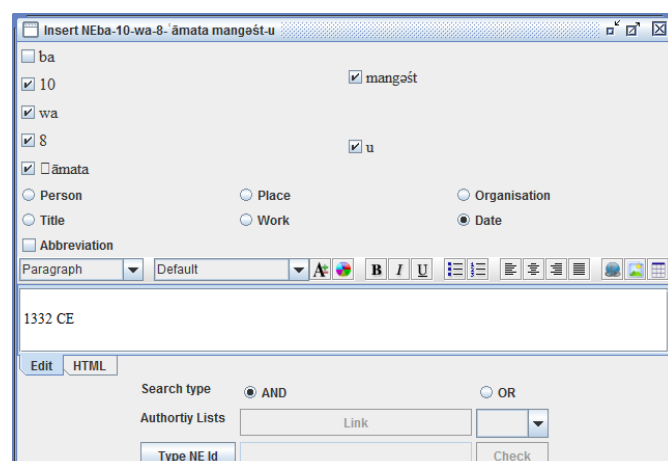


Figure 69

You can also use the NE frame to write out the abbreviations: tick the ‘Abbreviation’ box and provide the full text in the free text field.

### 8.5. Hiding NE annotation

In some cases, too much highlighting and underlining may disturb the annotators and/or users. If you do not want the named entities appear underlined you can hide the annotation by selecting ‘Disable NEs’ in the View/Configure Annotations top menu (Figure 70).

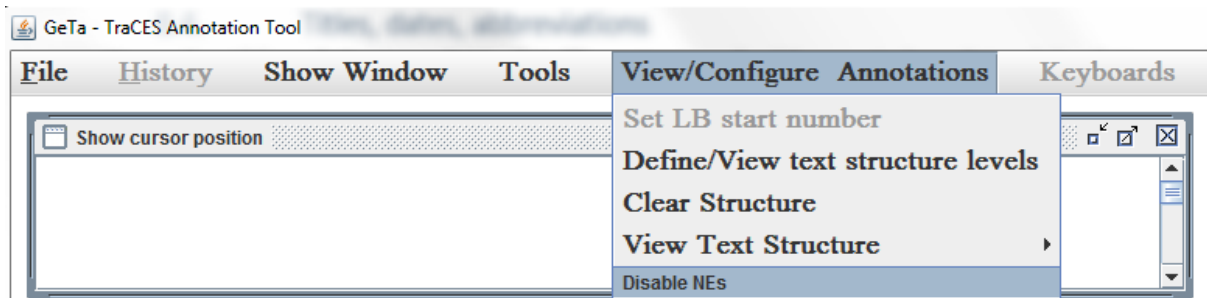


Figure 70

## 9. Working with GeTa: particular examples of annotation

### 9.1. Disambiguation

GeTa tool offers many possibilities of linguistic disambiguation. Apart from the lexical disambiguation by linking to the correct lemma in the dictionary (see e.g. Figure 44), the users might need to provide morphological disambiguation – assigning different parts of speech or grammatical forms to homonymous tokens.

Thus,  $\Lambda$  can be a conjunction (if it is followed by a verb, see Figure 71) or as a preposition (if it is followed by a noun, see Figure 72).

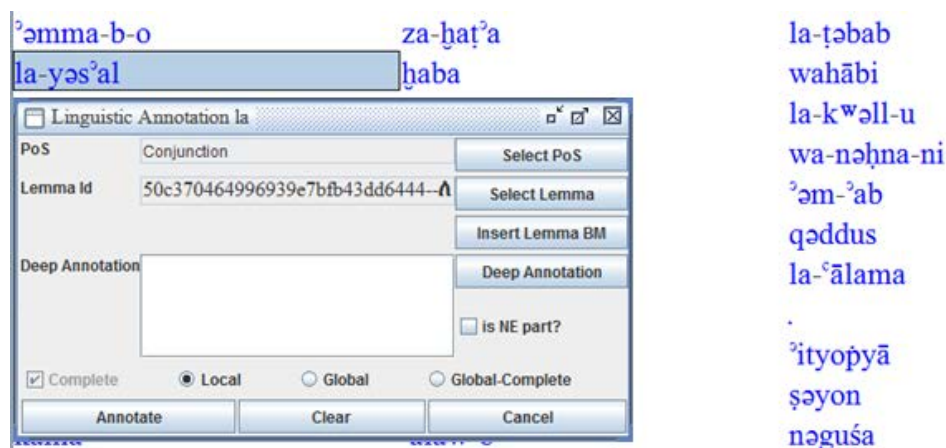


Figure 71

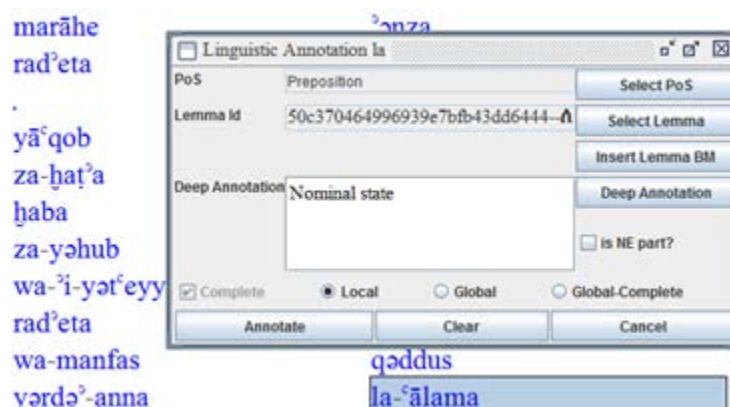


Figure 72

In some cases, morphological disambiguation can and should be reflected not only at the annotation but also at the transcription level. For example, a noun with the last consonant in the sixth order (e.g. ሰራዊት, ሐግ) should be transcribed without the final ə-vowel when it is in the absolute state (i.e. the

automatically produced final ə-vowel should be deleted to produce *sarāwit*, *həgg*) but with the vowel when the noun is in pronominal state and is followed by a pronominal suffix (i.e., e.g., *la-sarāwitə-ya*, *həggə-ka*, see also Figure 73).

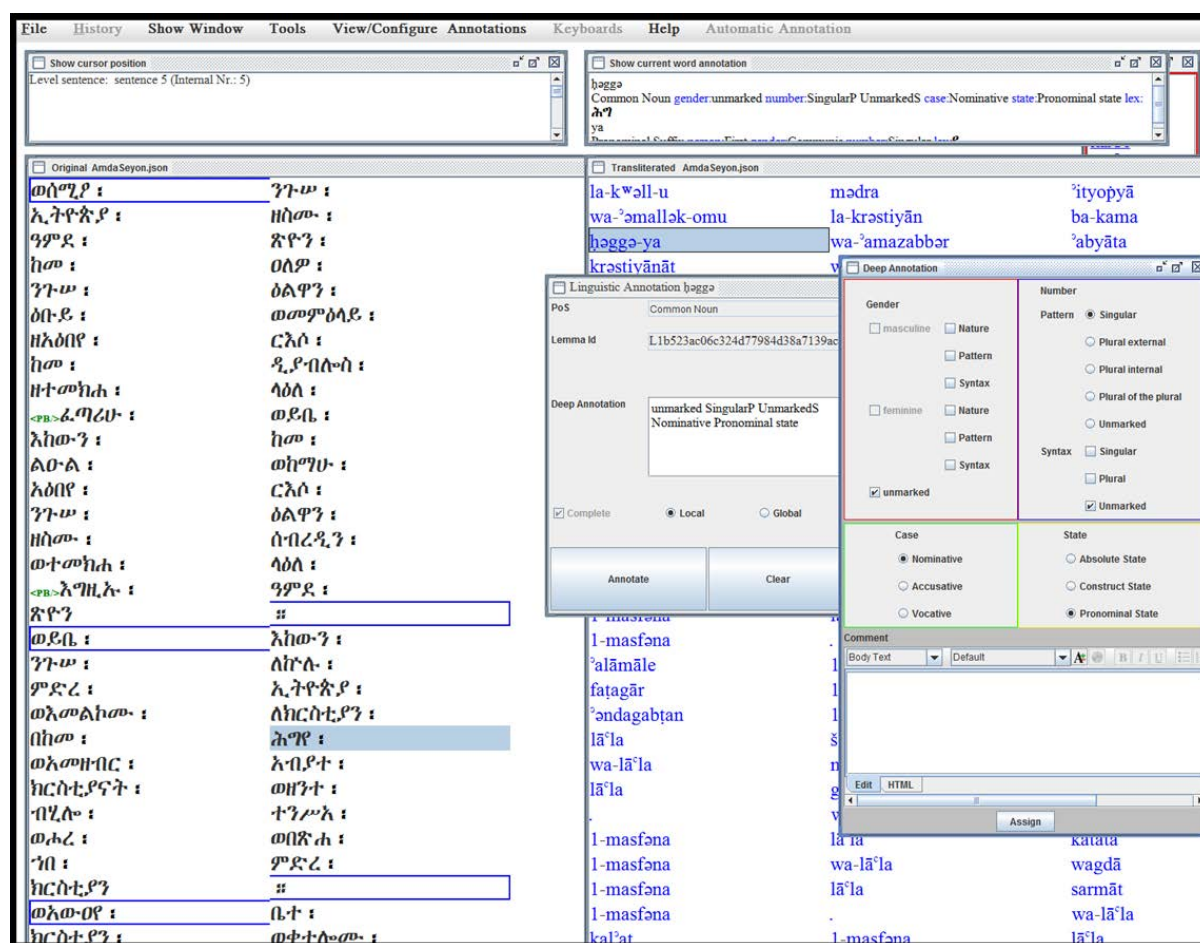


Figure 73

## 9.2. Defining the gender and number in nouns

Another particular example of deep linguistic annotation produced with the GeTa tool is assigning of gender in nouns (see Table 1. Parts of speech and possible features in the GeTa tool for a full list of possible features).

Wherever the gender is marked, the annotator is invited to specify on which basis the gender attribution took place – physical nature of the object (Nature), the morphological pattern (Pattern), or the syntactic agreement. In some cases, the gender may be marked multiply, and the physical, morphological, and syntactic gender do not necessarily coincide.

For the number, the annotators are invited to specify whether the number is marked or unmarked morphologically, i.e. by pattern, and/or syntactically, i.e. by agreement. If a noun is plural by pattern, the annotator can additionally choose between external and internal plural, or the plural of the plural. Again, the syntactic and the morphological number do not necessarily coincide.

E.g. the noun አዝብ: *həzb* (which can be both masculine and feminine according to the dictionary by Dillmann) when encountered in the sentence *wa-sam'u həzba tanbālāt* has been annotated as masculine by the agreement with the verb which precedes it (*sam'u*, Perfect Third Masculine Plural of the verb ሰምዖ: *sam'a* 'to hear, to listen') and as plural by the syntactic agreement, while being singular in pattern (see Figure 74).

ሰነድ ፡ ወግዳሽ wa-bəṛura

# ☐ Linguistic Annotation ስጋጃ

ወሀር፡ PoS Common Noun

ወውብ፡ Lemma Id 597784fcdbebe9c9fad02c333--አሽብ

ሸዋ፡

ባሕሪ፡ Deep Annotation masculineS SingularP PluralS

☐ Deep Annotation

☐ is NE part?

Gender

☒ masculine ☐ Nature

☐ Pattern

☒ Syntax

☐ feminine ☐ Nature

☐ Pattern

☐ Syntax

☐ unmarked

Number

Pattern ☒ Singular

☐ Plural external

☐ Plural internal

☐ Plural of the plural

☐ Unmarked

Syntax ☐ Singular

☒ Plural

☐ Unmarked

Case

☒ Nominative

☐ Accusative

☐ Vocative

State

☐ Absolute State

☒ Construct State

☐ Pronominal State

Comment

wa-<sup>3</sup>albās  
kəbura  
bərtə-ssa  
kama  
wa-<sup>3</sup>əm-dəḥra-z  
wəsta  
tanbālāt  
kʷəlgora  
.  
wa-hargāya  
wa-wəsta  
šawā  
ba-<sup>3</sup>afa  
<sup>3</sup>ahgura  
.  
wa-ḍewawu  
masfart  
tanbālāt  
za-wə<sup>3</sup>ətomu  
kama  
yəṣbə<sup>3</sup>u

qattant  
za-<sup>3</sup>alb-o  
wa-ḥašin  
<sup>3</sup>əbən  
fannawkəw-omu  
kʷəll-u  
<sup>3</sup>əlla  
wa-gəḍāya  
wa-fadse  
wa-bəqul  
kʷəll-u  
wa-ṣab<sup>3</sup>əw-omu  
ḥašin  
<sup>3</sup>abayta  
wa-māhəraku  
sab<sup>3</sup>a  
wa-sam<sup>3</sup>u  
wa-sab<sup>3</sup>a  
nolota  
fannawkəw-omu  
kāl<sup>3</sup>a

Figure 74

### 9.3. Using the comment function

Comments should always be left when the annotators are not sure about their choices or if these choices may not appear obvious to other users.

For example, the verb form *yəmaṣṣə*, which by pattern seems to be Imperfect 3<sup>rd</sup> person masculine singular was annotated as Imperfect 1<sup>st</sup> person communis singular when appearing as part of the graphic unit *ኢይመጽሕኦ*: 'i-yəmaṣṣə'-a: the preceding token, the negative particle *ኢ* 'i-, assimilated with the prefix of the Imperfect 1<sup>st</sup> person communis singular *እ* 'ə-, becoming the initial *ይ*. This explanation was provided in the Comment field during the Deep annotation process (see Figure 75).

rəgum  
la-saytān  
za-yəṭqārran  
wa-rəḥuq  
<sup>3</sup>əm-səbhata  
<sup>3</sup>əm-tərsita  
la<sup>3</sup>aka  
<sup>3</sup>ənza

ሕላጃ  
ሃይባል  
ነጻህ

*i-yəmaṣṣə*-a

☐ Show current word annotation

ሰነድ  
Negative Particle lex:ኢ  
yəmaṣṣə<sup>3</sup>  
Verb tam:Imperfect person:First gender:Communis number:Singular  
Comment:  
This verb, at the Imperfect, has been annotated as First Communis Singular, because there is the assimilation between the Negative Particle before , i.e. ኢ 'i-, and the prefix of the Imperfect First Person Singular እ 'ə-, which after the Negative Particle becomes ይ ሃጃ.  
lex:መጽሕኦ

Figure 75

Another example is the necessity to justify a correction or a normalization as compared to the source text of the edition (Vorlage, see Figure 76 : here an error of the source edition has been corrected).

እግዚአብሔር	wa-ʔaʔələfa
ሰሚያ :	.
ውእቱ :	ʔəm-lāʔkāna
ወልዱ :	rəgum
ለጽድቅ :	la-saytān
ክርስቶስ :	za-yəṭqārran
ወሊሉይ :	wa-rəḥuq
ወፍሉጥ :	ʔəm-səbhata
ቅዱስ :	ʔəm-tərsita
ንጉሥ :	laʔaka
ኢይመጽእኦ :	ʔenza
ወኢይቀውምኦ	ḥabe-ka-ʔa
ትመጽእኦ :	qədmə-kaʔ-a
እስመ :	ḥabe-ya-ʔa
እለ :	bə-ya
እለ :	yəbazzəḥu

ba-hayla	ʔəgziʔababer
<div> <div>Show current word annotation</div> <div> <div>qəddus</div> <div>Common Noun gender:masculineP number:SingularP UnmarkedS case:Nominative state:Absolute state</div> <div>Comment:</div> <div>Typing mistake in Marrassini's edition where the noun is written as ቅዱስ : qəddus. Cp. Marrassini 1993, p. 58: 5 (ed.).</div> <div>lex:ቅዱስ</div> </div> </div>	
wald	wa-fəluṭ
manfas	qəddus
ḥaba	nəguś
yəbəl	ʔi-yəmaşşəʔ-ʔa
.	wa-ʔi-yəqawwəmʔ-a
wa-soba	təmaşşəʔ-ʔa
ʔi-yəfarrəh-ʔa	ʔəsmə
sarāwit-ʔa	ʔəlla
ʔəm-sarāwitə-ka	ʔəlla

Figure 76



## Glossary of terms

Abbreviation	Shortening used in the edited text; can be resolved by using the Insert/Modify NE function
Accusative Particle	The affixed particle +ሃ (suffixed only to proper names) (in linguistic annotation)
Adverb	Unchangeable part of speech primarily modifying verbs and clauses; includes interrogative adverbs (አይቲ፡, ማእከ፡, እፎ፡, ሚመጠን፡) and other adverbs (location: ህየ፡, አፍአ፡, ዝየ፡, ፀጋም፡ time: ቀዳሙ፡, ካዕበ፡, ዮም፡, ይእከ፡, ጌሰም፡ and others: መጠነ፡, ቀዳሚ፡, ባሕቱ፡, አማን፡, ግሙራ፡) (in linguistic annotation)
Affirmative Particle	The particle እወ፡ and the affixed particle +ኣ <i>yes indeed!, certainly!</i> (in linguistic annotation)
ANNIS	ANNOtation of Information Structure - A web browser-based search and visualization architecture for complex multilayer linguistic corpora - <a href="http://corpus-tools.org/annis/">http://corpus-tools.org/annis/</a>
Annotation	The process of providing meta-data at various levels, see also linguistic annotation, deep annotation, global annotation, Named Entities, structure annotation
Aspect	Verbal grammatical category that expresses how an action, event, or state, denoted by a verb, extends over time (in linguistic annotation)
Basic annotation	Linguistic annotation where only PoS labels are assigned to the tokens (without deep annotation)
Cardinal Numeral	A nominal part of speech representing quantity (in linguistic annotation)
Case	A grammatical category of nominals (nouns, pronouns, numerals, infinitive verbs, etc.); Nominative, Accusative, and Vocative can be marked (in linguistic annotation)
Case Accusative	The case of a direct object (in linguistic annotation)
Case Accusative Ø	The case of a direct object with zero flexion (in linguistic annotation)
Case Nominative	The subject case (in linguistic annotation)
Case Vocative	The case of the addressee (in linguistic annotation)
Common Noun	A part of speech (noun) denoting a general item (as opposed to Proper Name) (in linguistic annotation)
Conjunction	A part of speech used to connect (coordinate) words or clauses (in linguistic annotation)
Context	Strings / words / text passages surrounding the focal unit (before and after)
Date	Time reference used in the edited text; can be explained by using the Insert/Modify NE function (in linguistic annotation)
Deep annotation	Detailed linguistic annotation of all features identifiable for each part of speech in the edited text
Deictic Imperative Particle	The particles m.sg. ነዓ፡ (ንዓ፡), f.sg. ንዒ፡, m.pl. ንዑ፡, f.pl. ነዓ፡ (in linguistic annotation)
Demonstrative Pronoun	Pronouns that point to specific things such as. ዝ+ / ዘ+ , ዝንቲ፡ / ዘንተ፡, ዛ+ or ዛቲ፡ / ዛተ፡, እሉ፡ or እሎንቲ፡ / እሎንተ፡, እላ፡ (እሎን፡) or እላንቲ፡ (እላንተ፡) / እላንተ፡ for near objects; ዝከ፡ / ዘከ፡ or

ዝኹቱ፡/ዝኹተ፡, እንትኩ፡ /እንትኩ፡ or እንታኩቲ፡/ እንታኩተ፡, እልኩ፡ or እልኹቱ፡ (እልኩቱ፡)/እልኹተ፡ (እልኩተ፡) or እላኩቱ፡/እላኩተ፡ for far objects (in linguistic annotation)

Editorial annotation	Marking up editorial choices such as reconstructed, conjunctured, or rejected/expunged letters, words, etc.
Ethiopic	See Gə'əz
Existentials	Part of speech corresponding to the English existential clause 'there is / there is not'
Existentials Affirmative Base	ቦ+Pronominal Suffix <i>there is</i> (ብየ፡, ብከ፡, ብኪ፡, ቦ፡ or ቦቱ፡, ባ፡ or ባቲ፡ ...) (in linguistic annotation)
Existentials Negative Base	አልቦ+Pronominal Suffix: (አልብየ፡, አልብከ፡, አልብኪ፡, አልቦ፡ or አልቦቱ፡, አልባ፡ or አልባቲ፡ ...) (in linguistic annotation)
Extension	A string in the file name with the time stamp, version number, and other variables
Fidal	Gə'əz script
Fidal separator	See Word separator
Foreign Material	In linguistic annotation: borrowed elements with no lexicalization in Gə'əz, e.g. ሰገድ <i>sagad</i> (element in proper names such as 'Aṣnāf Sagad)
Gə'əz	Classical Ethiopic language
Vocalized Gə'əz	Gə'əz script (fidal) with vowel orders
Unvocalized Gə'əz	Gə'əz script (fidal) with all consonants appearing in the first order (as in early inscriptions)
Gender	A specific form of noun class system in which the division of noun classes forms an agreement system with another aspect of the language, such as adjectives, articles, pronouns, or verbs. We distinguish masculine, feminine, and unmarked. Can be expressed by pattern, by syntactic agreement, or implied by the physical nature of the object (in linguistic annotation)
Gender communis	In verbs, relative pronouns, pronominal suffixes: common gender (in linguistic annotation)
Gender feminine	To be provided for nominals and verbs (in linguistic annotation)
Gender masculine	To be provided for nominals and verbs (in linguistic annotation)
Gender nature	Gender implied by the physical nature of the object - feature existing in nouns only (in linguistic annotation)
Gender pattern	Gender expressed by the morphological root pattern; marked in nouns (common and proper) only (in linguistic annotation)
Gender syntax	Gender expressed by syntactic agreement - in nominals (and infinitives) only (in linguistic annotation)
Gender unmarked	In nominals only (in linguistic annotation)
Gerund(ive)	A particular nonfinite nominal verb form denoting a completed and anterior action; also known as converb (in linguistic annotation)
GeTa	Gə'əz Toolkit for Annotation
Global annotation	Applying the global annotation lets the user automatically assign the same meta-data (tokenization, features, links, comments, etc.) to all occurrences of the same unit in the text



Graphic unit	Sequence of symbols written together, can contain more than one annotatable token
Imperative	A grammatical mood (in verbs) that forms a command/request (in linguistic annotation)
Imperfect	A verb form combining past tense and imperfective aspect (in linguistic annotation)
Independent Personal Pronoun	Pronoun associated with a particular grammatical person (1st, 2nd, 3rd): አኒ፣ አንተ፣ አንቲ፣ ወእቱ፣/ወእተ፣ ይእቲ፣/ይእተ፣ (in linguistic annotation)
Infinitive	Verbal noun, as distinguished by the pattern and nominal grammatical features (in linguistic annotation)
Index	Computer-produced register of all units in a given text
Interjection	Particles used to express pity, surprise, joy, etc., e.g. አሌ፣ (+ለ+PSuff: አሌልየ፣ or አሌ፣ ሊተ፣ ), ወይ፣; አ+ , አህ፣, እንቋዕ፣, ዩ፣ or ዮ፣; ጸጥ፣; ባሕ፣ (in linguistic annotation)
Interrogative Adverb	Question word such as አይቴ፣ , ማእዜ፣ , እፎ፣ , ሚመጠን፣ (in linguistic annotation)
Interrogative Particle	Particle converting the sentence into a yes/no question, e.g. +ሁ, +ኑ (in linguistic annotation)
Interrogative Pronoun	Question word such as መኑ፣/መነ፣ who?, ምንት፣/ምንተ፣ what?, sg. አይ፣/አየ፣ : which?, pl. አዖት፣/አዖተ፣ (in linguistic annotation)
JSON	JavaScript Object Notation - a lightweight data-interchange format
Lemma	The canonical dictionary form of a word, corresponding to an entry in the dictionary (with specific grammatical definition, translation, etc.)
Lexical unit	A single word (can be a graphic unit or a part of a graphic unit) that forms the basic element of the language lexicon and corresponds to a dictionary lemma
Lexical annotation	Linking each lexical unit to the corresponding lemma in the dictionary (in linguistic annotation)
Lines and line breaks	Features of layout of the source text (poetry, inscriptions) that can be reproduced in the GeTa tool (as part of editorial mark-up)
Linguistic annotation	Mark-up that provides meta-data concerning the lexical and grammatical categories and features of the tokens in the selected text
Logogram	A sign representing a word, such as the symbols standing for the numerals (in linguistic annotation)
Matching	As in string matching: searching for identical strings within a given text
Meta-data	Data about data: any additional information about a text, a graphic unit, a token, etc., that can be provided at any level
Mood	A verb category indicating whether a verb expresses a fact (indicative, default), a possibility (subjunctive), or a command (imperative mood) (in linguistic annotation)
Named Entity	Real-world objects or categories that words (proper names, numerals, logograms, etc.) can be related to; include Persons, Places, Works, Organisations (with a link to an authority file), Titles, Dates, and Abbreviations; often indicated by proper names, but not necessarily (in Named Entities annotation)
Negative Particle (Base)	Particles indicating negation, e.g. ኢ+, አልቦ፣, እንቢ+PSuff: እንቢየ፣, አኮ፣ (in linguistic annotation)

Nominal	Part of speech category including nouns, pronouns, and numerals (in linguistic annotation)
Noun	Part of speech used to refer to animate or inanimate objects, ideas, or events; can be common or proper (in linguistic annotation)
Number	Grammatical category of nominals (and nominal verb forms) that expresses count distinctions; can be singular, plural, or unmarked, and be expressed externally (by an affix) or internally (infix/pattern) or both (plural of the plural in nouns) or marked by syntactic agreement (in linguistic annotation)
Number pattern	Number marked by the root pattern (in linguistic annotation)
Number plural	Category corresponding to ‘more than one’ object count; can be expressed in nominals and verbs (in linguistic annotation)
Number plural external	Plural marked by an external affix (in linguistic annotation)
Number plural internal	Plural marked word-internally (infix or pattern) (in linguistic annotation)
Number plural of the plural	Number marked doubly, by pattern and external affix (in linguistic annotation)
Number singular	Category corresponding to the ‘single/one’ object count; can be expressed in nominals and verbs (in linguistic annotation)
Number syntax	Number indicated by syntactic agreement with other words in the sentence
N numeral	Part of speech (nominal) that designates a number (cardinal or ordinal)
Object Pronoun Base	Non-independent pronoun in the object function such as <i>h,ʔ</i> +Pronominal Suffix
Ordinal Numeral	Ordinal number, a nominal part of speech used to specify position in a sequence (in linguistic annotation)
Organisation	A Named Entity type, an institution of any type
Pages and page breaks	Features of layout of the source texts (editions, manuscripts) that can be reproduced in the GeTa tool (as part of editorial mark-up)
Part of speech	Abbreviated PoS - a category of lexical items with similar grammatical behaviour, includes nominals (nouns, pronouns, numerals), verbs, and particles (adverbs, prepositions, conjunctions, interjections, further particles); the GeTa tool additionally specifies existentials. Punctuation is also treated at the PoS level (in linguistic annotation)
Particle	Part of speech without inflection, includes adverbs, prepositions, conjunctions, interjections, further particles (in linguistic annotation)
Perfect	Verbal tense/aspect for completed actions
Person	Grammatical category in verbal conjugation reflecting the reference to the participant of the action, can be first, second, or third (in linguistic annotation of verbs and pronominal suffixes)
Place	A toponym indicating a specific location; a link to the authority file can be provided (during Named Entities annotation)
PoS	See Part of speech

Possessive Pronoun Base	Non-independent pronoun in the possessive function such as m.sg. <i>il.h+PSuff</i> , f.sg. <i>እንቲሕ+PSuff</i> , c.pl. <i>እሊ.አ+PSuff</i> (in linguistic annotation)
Preposition	PoS for words (without inflection) showing relationship between other words in a sentence
Presentational Particle Base	Non-independent particles such as <i>ኅ+PSuff:3m.sg</i> ( <i>ኅሁ፡</i> ), <i>ነው+PSuff:3f.sg</i> ( <i>ነዋ፡</i> ), <i>ነይ+PSuff</i> , <i>እን+PSuff</i> (in linguistic annotation)
Pronominal Suffix	Abbreviated PSuff - non-independent affix in the personal pronoun function, such as <i>+የ/ +ኒ, +ከ, +ከ., +ሁ, +ሃ</i> etc. (in linguistic annotation)
Pronoun	Part of speech for words substituting for nouns, can be personal, possessive, demonstrative, relative, interrogative
Pronoun of Solitude Base	Non-independent pronoun meaning ‘alone’, followed by pronoun suffix <i>ባሕቲት+</i> (in linguistic annotation)
Pronoun of Totality Base	Non-independent pronoun meaning ‘all, whole’, followed by pronoun suffix <i>ኩል+ /ኩለንታ+</i> (in linguistic annotation)
Proper Name	Names signifying unique objects
PSuff	See Pronominal Suffix
Punctuation	Sentence marks (the GeTa tool only supports the word separator <i>:</i> and the full stop <i>::</i> ) (in linguistic annotation)
Quotative Particle	Suffixed particle <i>+አ</i> marking direct speech (in linguistic annotation)
Relative Pronoun	Pronoun referring back to a thing or person previously mentioned in relative clauses, m.sg. <i>il+</i> , f.sg. <i>እንቲ፡</i> , c.pl. <i>እለ፡</i> (in linguistic annotation)
Sabaic	Language spoken in South Arabia and used also in the Horn of Africa in the first millennium BC with its script used in both Sabaic and Aksumite Ethiopic inscriptions
Separator	See word separator
Span	A part of a document/text that has been assigned an identifying label, can be of any length, from one character to the entire text, depending on the level of meta-data. The term is most frequently used in connection with the text structure annotation.
State	Nominal category describing, for nouns (incl. verbal nouns, or infinitives), the role of a noun in a genitive construction, and for particles, the attribution of an additional marker
State Absolute	A grammatical category of the independent (main) noun in unbound form, as opposed to construct or pronominal state (in linguistic annotation)
State Construct	Status constructus - a grammatical category of a noun bound to a following noun in a genitive relationship (in linguistic annotation)
State Construct Ø	Status constructus without grammatical marker (in linguistic annotation)
State Nominal	Prepositions and interjections in unbound form (in linguistic annotation)
State Pronominal	A grammatical category of a noun bound to a following pronoun in a genitive relationship; also in prepositions, interjections (in linguistic annotation)
String	A sequence of two or more characters

