

Sentence Segmentation Guidelines for Early New High German

Zarah Weiß and Gohar Schnelle

June 23rd, 2016, Version 2.0

1 Introduction

The following guidelines were designed to allow for a consistent sentence segmentation of Early New High German t-unit (ENHG) texts. As these texts include only partial or ambiguous punctuation, a mainly graphematic sentence definition is not applicable. Hence, these guidelines focus on a primarily syntactic approach with the purpose of facilitating further natural language processing. Therefore, the guidelines are designed to produce segments, that are reasonable with respect to the contemporary research discussion, yet also suited as input for natural language processing tools. Accordingly, the guidelines are influenced by linguistic as well as pragmatic considerations.

These lead to a variant of the classical t-units, which was modified to suit the special needs of ENHG in general as well as the corpora annotated (Ridges). In the following, this variant is referred to as Early New High German t-unit (ENHG-TU). T-units were originally introduced by **hunt1965**, who defined them as "‘shortest grammatically allowable sentences into which (writing can be split) or minimally terminable **unit** [bold font by Z.W.]". They are well known in linguistic complexity and discourse analysis of both, spoken and written language (??).

In the following, first a list of all guidelines used to define ENHG-TUs is presented in section 2. Then, a detailed discussion of the separate rules is given in section 3, featuring example annotations and problematic cases.

2 Overview

1. **Base definition:** A ENHG-TU consists of a phrasal head and all of its overt arguments and adjuncts and nothing else.
2. **Independence of the head:** The head of a ENHG-TU may not be the argument or the adjunct of another head itself, i.e. ENHG-TU do not govern each other.
3. **Uniqueness:** ENHG-TU may not overlap, i.e. no phrase is part of more than one ENHG-TU.
4. **Exhaustivity:** A text has to be partitioned exhaustively into ENHG-TUs.
5. **Finiteness:** A ENHG-TU includes preferably a finite verb. However, this is not mandatory. Therefore, if an ENHG-TU is ambiguous with respect to whether it contains a finite verb or not, the analysis including at least a single finite verb is to be preferred.
6. **Minimal length I:** If
 - a. the head of a potential ENHG-TU is structurally ambiguous with respect to its own status as argument or adjunct of another head, and
 - b. it is not possible to disambiguate the structure based on textual coherence,the potential sentential unit is annotated as an ENHG-TU.
7. **Minimal length II:** If
 - a. a phrase is structurally ambiguous with respect to its attachment to two ENHG-TUs, and
 - b. it is not possible to disambiguate the structure based on textual coherence,the phrase is considered to be attached to the shorter ENHG-TU. The length of an ENHG-TU is defined in terms of tokens. If both ENHG-TUs in question contain the same amount of tokens, the phrase is attached to its preceding ENHG-TU.
8. **Continuity:** ENHG-TUs are continuous strings of tokens. Discontinuous ENHG-TU are not possible, except if some meta text was inserted into a sentence.
9. **Sentence ending punctuation:** Sentence ending punctuation has to be located at the outermost right periphery of a ENHG-TU.

3 Description

3.1 Base definition

Rule A ENHG-TU consists of a phrasal head and all of its overt arguments and adjuncts and nothing else.

Description This rule forms the shortest, grammatical sentence, if possible. In this case, the finite verb of the main clause is the phrasal head and all overtly realised arguments and adjuncts belonging to the verbs maximal projection are included. Arguments and adjuncts themselves may be smaller phrases or subordinated clauses. Example 1 shows an ENHG-TU, which head is a finite verb (*bedeutet*). It has a nominal phrase as its subject and a subordinate clause as its direct object.

- (1) a. **Correct:** { auch **bedeutet** die rote Butte [...] / wie sehr zunahm die christliche Kirche auf dieser Erde von Christus gehalten lieb und wert }
- b. **Wrong:** { auch bedeutet die rote Butte / [...] } { wie sehr zunahm die christliche Kirche auf dieser Erde von Christus gehalten lieb und wert }¹

Because both phrases are in fact arguments of the finite verb, they have to belong to the same ENHG-TU.

Coordinated main clauses, though, form separate ENHG-TUs. This is illustrated in example 6 illustriert, which contains several coordinated main clauses, that have to be analysed as separate ENHG-TUs.

- (2) a. **Correct:** { Galienus spricht dass das Kraut gut sei zu essen mit Lactuken, } { wann es sänftigt der Lactuken Kälte } { und sein Samen ist gut wider die Wassersucht, } { wann es erhitzt die Leber und es reinigt sie ¶ }
- b. **Wrong:** { Galienus spricht dass das Kraut gut sei zu essen mit Lactuken, wann es sänftigt der Lactuken Kälte und sein Samen ist gut wider die Wassersucht, wann es erhitzt die Leber und es reinigt sie ¶ }²

So the base definition mainly corresponds to the classical definition of t-units. However, it does not use the term "sentence", which is to be considered a benefit when it comes to defining sentence-like units. Since the category of the phrasal head is not constrained to verbs, nominal and prepositional phrases, too, may form ENHG-TUs. This is a necessary divergence from the original definition of t-units, as – for example – the Ridges corpus contains noun phrases such as headlines, interjections and appellations which are not incorporated in their surrounding sentences. Interjections are prototypically not considered to be t-units (**young1995**). However, it is not desirable to include them in the segmentation, if they are not integrated. This is illustrated for interjections in example 4 and for headlines in example 3. Also, please see the subparagraph on appositions and parentheses in the trouble shooting paragraph of the rule of continuity in section 3.8.

¹From ???, Paradeiszaertlein_1588_Rosbach, Z. 3777.

²From ???, BuchDerNatur_1482_vonMegenberg.xlsx, l. 1566ff.

- (3) a. **Correct:** { Leibliche Nutzen und Wirkung. } { Dies Kräutlein Abbiss wird genannt [...] }
 b. **Wrong:** { Leibliche Nutzen und Wirkung. Dies Kräutlein Abbiss wird genannt [...] } ³
- (4) a. **Correct:** { Ach und aber ach } { wie lang hat mich die Welt in die Finstere gezogen und lockt mich nach ihr }
 b. **Wrong:** { Ach und aber ach wie lang hat mich die Welt in die Finstere gezogen und lockt mich nach ihr } ⁴

Trouble shooting

Coordination vs. subordination It is to be noted that coordinating connectives in contemporary German may be used as subordinating connectives in ENHG and vice versa. This is shown in example 5, where *doch* – a subordination in contemporary German – is used to coordinate two main clauses. Accordingly, it is analysed as the beginning of a new ENHG-TU. It also contains *denn*, which is a coordination in contemporary German, but used as a subordination in the example. Hence, the entire clause introduced by *denn* is part of the preceding ENHG-TU.

- (5) { **doch** mögen wir Deutschen wohl und recht dafür gebrauchen das Kraut so man Welsamen nennt / **denn** es der Kraft / auch zum Teil der Gestalt nach dem rechten Seriphium ganz gleich ist } ⁵

Shared arguments If two coordinated clauses share an argument, they are considered to be in the same ENHG-TU, if the structure can be analysed as elision. This is illustrated in example ??, where the first two clauses are in the same ENHG-TU, because they share the subject *das Kraut*. The following clause is in a separate ENHG-TU, because it contains its own subject *es*.

- (6) a. **Correct:** { Das Kraut entlöst Blähung in dem Leib und ~~das Kraut~~ öffnet das Verstopfen des Leibes } { und darum macht es schwitzen }
 b. **Wrong:** { Das Kraut entlöst Blähung in dem Leib } { und PRO öffnet das Verstopfen des Leibes } { und darum macht es schwitzen } ⁶

If it is not plausible to assume elision, a null argument is assumed. This leads to a separation of both clauses into two ENHG-TUs. For more details on this case, please see the subparagraph **elision vs. null arguments** in the trouble shooting paragraph of the first rule of minimal length in section 3.6.

3.2 Independence of the head

Rule The head of a ENHG-TU may not be the argument or the adjunct of another head itself, i.e. ENHG-TU do not govern each other.

³From ???, Paradeiszaertlein_1588.xlsx, l. 50ff.

⁴From ???, BuchDerNatur_1482_vonMegenberg.xlsx, l. 4905ff.)

⁵From ???, NewKreuterbuch_1543_Fuchs, Z. 304.

⁶From ???, BuchDerNatur_1482_vonMegenberg.xlsx, l. 1506ff.

Description It is necessary to impose this additional restriction, to prevent overgeneralisations based on the wide applicability of the base definition. Without further restrictions, all maximal projections of phrasal heads and subordinate clauses would qualify as ENHG-TUs. However, this is not intended. How the restriction applies to the segmentation process is illustrated in example 20.

- (7) a. **Correct:** { auch **bedeutet** die rote Butte / darin der Samen liegt / wie sehr zunahm die christliche Kirche auf dieser Erde von Christus gehalten lieb und wert }
- b. **Wrong:** { { auch } bedeutet die rote Butte / { darin der Samen liegt / } { wie sehr zunahm die christliche Kirche auf dieser Erde von Christus gehalten lieb und wert } } ⁷

Troubleshooting Currently, no issues regarding this rule are known.

3.3 Uniqueness

Rule ENHG-TU may not overlap, i.e. no phrase is part of more than one ENHG-TU.

Description This rule is consistent with the typical definitions of sentences and t-units. It mainly serves the purpose of giving a complete definition of ENHG-TUs.

Troubleshooting Currently, no issues regarding this rule are known.

3.4 Exhaustivity

Rule A text has to be partitioned exhaustively into ENHG-TUs.

Description This rule is consistent with the typical definitions of sentences and t-units. It mainly serves the purpose of giving a complete definition of ENHG-TUs.

Troubleshooting Currently, no issues regarding this rule are known.

3.5 Finiteness

Rule A ENHG-TU includes preferably a finite verb. However, this is not mandatory. Therefore, if an ENHG-TU is ambiguous with respect to whether it contains a finite verb or not, the analysis including at least a single finite verb is to be preferred.

⁷From ???, Paradeiszaertlein_1588_Rosbach, Z. 3777.

Description This rule is intended to prefer ENHG-TUs with finite verbs over ENHG-TUs without finite verbs. This preference is necessary to state, because the rules of minimal length do allow for an interpretation that would lead to an inflationary annotation of, for example, *solitaire noun phrases* as separate ENHG-TU.

Applying this rule it is important to notice, that ENHG knows constructions featuring a covert finite verb and constructions that are actually without a finite verb. For example there are so called *afinite constructions*, where the finite auxiliary verb does not have to be overtly realised, if it is positioned in the right sentence bracket in the topological field model. This is illustrated in example 8. Also, elliptical copula constructions as in example 9 are especially common. In both examples, the covert finite verb was added but crossed out for illustrational purposes.

- (8) { Unter denen erstlich M. Cato Censorius, von dem L.Columella meldet / dass er der erste gewesen ~~ist~~ / so den Feldbau die lateinische Sprache gelehrt ~~hat~~ }
- (9) { [Die Blättlein] sind ein wenig rauher und ringsumher gekerbt / die Stängel ~~sind~~ purpurrot } ⁸

Because in these cases the covert finite verb is highly salient, its overt absence qualitatively differs from actual constructions without finite verbs, such as headlines, interjections, etc., as seen in the base definition.

Troubleshooting Currently, no issues regarding this rule are known.

3.6 Minimal length I

Rule If

- a. the head of a potential ENHG-TU is structurally ambiguous with respect to its own status as argument or adjunct of another head, and
- b. it is not possible to disambiguate the structure based on textual coherence, the potential sentential unit is annotated as an ENHG-TU.

Description This rule leads to more ENHG-TUs of shorter length. These are preferred for practical reasons, since Natural Language Processing (NLP) tools such as parsers work more efficient and less error prone with shorter sentential units. Also, ENHG is known to feature particularly long sentences, which exceed the average sentence length in contemporary German by far. Example 10 illustrates how the rule is applied.

- (10) a. **Correct:**
- b. **Wrong:**

The restriction of the first rule of minimal length in 3.6 b. prevents overgeneralisations based on purely syntactic considerations. This is shown in example 11, where several direct objects are coordinated. Clearly, it is desirable to analyse those as part of the same ENHG-TU, albeit the object valency of the finite verb *nimm* could be argued to be already saturated by the first object.

⁸From ???, NewKreuterbuch_1543_Fuchs.xlsx, l. 3395ff.

- (11) a. **Correct:** { **nimm** starken Weinessig zehn Lot / ein halbes Lot der grünen Wermutblätter / Salz ein Drittel eines Quäntchens }
- b. **Wrong:** { **nimm** starken Weinessig zehn Lot / } { ein halbes Lot der grünen Wermutblätter / } { Salz ein Drittel eines Quäntchens }

So without the restriction, the rule could be interpreted to lead to a segmentation as shown in example 11b, which is clearly not desirable. However, with the restriction, the context clearly disambiguates where the last two object phrases belong, because otherwise the text would not be coherent.

Troubleshooting

Demonstrative vs. relative clauses It is to be noted that this rule leads to a systematic disambiguation between demonstrative and relative clauses, favouring the demonstrative clause analysis. These are especially ambiguous in ENHG and the disambiguation process is object of the contemporary research discussion. Example 12 shows an ambiguous case, where the clause in questions starts with a typical relative pronoun, combined with V2, which is actually a sign of demonstrative clauses.¹⁰

- (12) a. **Correct:** { der gemeine Wermut ist ein Kraut mit vielen Zinken und Ästen / } { **an welchen** sind aschefarbene Blätter / vielfältig zerspalten / und goldgelbe Blumen / runder Same. }
- b. **Wrong:** { der gemeine Wermut ist ein Kraut mit vielen Zinken und Ästen / **an welchen** sind aschefarbene Blätter / vielfältig zerspalten / und goldgelbe Blumen / runder Same. }¹¹

Since demonstrative clauses are matrix clauses and relative clauses are subordinate clauses, the analysis of the given structure as a demonstrative clause is preferred due to the first rule of minimal length. However, a primarily linguistically motivated distinction between these types of clauses might lead to other conclusions.

To summarize, the first rule of minimal length allows for a systematic and consistent analysis regarding a highly problematic issue. However, this comes at the cost of not capturing the distribution of relative and demonstrative clauses in ENHG adequately. That is, research questions sensitive to that matter cannot be answered based on the information offered by ENHG-TUs, at least not without further adjustments in the search query used.

⁹From ???, NewKreuterbuch_1563_Handsch, Z. 604ff.

¹⁰It should be noted at this point, that relative pronouns in ENHG may differ from contemporary German relative pronouns. For example *so* is a very common introduction for ENHG relative clauses, yet, in contemporary German, it cannot be used like that anymore. This is illustrated in example 1.

(1) { Unter denen erstlich M. Cato Censorius, von dem L.Columella meldet / dass er der erste gewesen / **so** den Feldbau die lateinische Sprache gelehrt }

Aus ???, PflanzGartVorrede_1639_Rhagor, Z. 169ff.

¹¹From ???, NewKreuterbuch_1543_Fuchs.xlsx, l. ???ff.

Independent vs. dependent clauses As stated in the previous subparagraph on demonstrative and relative clauses, the first rule of minimal length generally prefers matrix clause annotations over subordinate clause annotations in ambiguous cases. However, one criterion is used to identify subordinate clauses beyond any doubt, namely the position of the finite verb: if a finite verb is located in the right sentence bracket of a clause, it is considered to be a subordinate clause, i.e. it belongs to the ENHG-TU of its matrix clause. This finite verb does not necessarily have to be overtly realised, see afinite constructions discussed in rule 3.5.

The position of the finite verb is considered to be a disambiguating factor, because the final establishment of the verb asymmetry between dependent and independent clauses at the stage of ENHG is common sense in the contemporary research discussion, see **axel???**. Hence, finite verbs in the right sentence bracket disambiguate the status of a clause, hence the first rule of minimal length does not apply in these cases. Example 13 shows such a case.

- (13) a. **Correct:** { in griechischer und lateinischer Sprache wird es genannt Absinthium / welcher Name bis auf den heutigen Tag in den Apotheken geblieben **ist**. }
 b. **Wrong:** { in griechischer und lateinischer Sprache wird es genannt Absinthium / } { welcher Name bis auf den heutigen Tag in den Apotheken geblieben **ist**. } ¹²

It shows a relative clause, which is introduced with *welcher Name*. From the point of view of contemporary German, this is unexpected, because the proper relative pronoun referring to *Absinthium* would be *was*. *welcher Name* would be more indicating of a demonstrative clause. However, the position of the finite verb outranks these concerns.

When talking about the position of the finite verb, it is important to acknowledge the highly productive occupation of the postfield in ENHG. While in contemporary German, the postfield mainly contains subordinate clauses, in ENHG all elements from the middlefield could be postponed. This is shown in example 15.

- (14) { [...] eine jede Blume ein Schote zwei / drei / herbringt / darin der Samen frei verborgen **liegt** eine lange Zeit } ¹³

The term *verb-final*, which is also commonly used to describe the position of the finite verb in subordinate clauses, is therefore misleading. This also holds for the order of the verbs in the right sentence bracket: in contemporary German, the finite verb is – except for a few rare special cases – located at the rightmost periphery of the right sentence bracket, where several verbs might be clustered. However, in ENHG this order is less fixed. Hence, it is to be expected, that the finite verb is in fact located in the right sentence bracket, but not at its right periphery, as in example ??.

- (15) { Apul. Und Orpheus lernen folgende Pillen zu machen / derer man alle morgens und abends zwei Skrupel dem Schwindsüchtigen **soll** geben / und gleich darauf **heißen** trinken laues Wasser. } ¹⁴

¹²From ???, NewKreuterbuch_1543_Fuchs, Z. 40ff.

¹³From ???, Paradeiszaertlein_1588_Rosbach, Z. 2060ff.

¹⁴From ???, WieSichMeniglich_1557_von Bodenstein.xlsx, l. 630ff.

Although neither constructions with occupied postfields nor constructions with a finite verb in the non-final position in the right sentence bracket look like verb last sentences, they in fact indicate subordinate clauses, because they are not positioned in the left sentence bracket.

Elision vs. null arguments One important question imposed by the first rule of minimal length is how to decide between elision and null arguments in cases, where two potential ENHG-TUs share an argument. Since the base definition only refers to overtly realised arguments, the rules could be interpreted and ranked such that in these cases two ENHG-TUs would be annotated. However, this would assume one of the ENHG-TUs to have a null argument. This is illustrated in example 16, where two potential ENHG-TUs share a subject: The annotation in 16a assumes, that there is in fact only a single ENHG-TU, in which the redundant subject was elided (marked by crossing the subject out). The annotation in 16b assumes three separate ENHG-TUs, of which two have a null subject (marked as PRO).

- (16) a. **Correct:** { Von dieser Wurzel getrunken räumt die Brust, ~~von dieser Wurzel getrunken~~ heilt die versehrte Lunge, ~~von dieser Wurzel getrunken~~ treibt aus den Kot }
 b. **Wrong:** { Von dieser Wurzel getrunken räumt die Brust, } { PRO heilt die versehrte Lunge, } { PRO treibt aus den Kot }¹⁵

It could speak in favour of the second analysis, that null subjects are possible in ENHG, see (**axel???**). Nevertheless, such an interpretation of the rules is problematic (and, therefore, to be rejected) from both, a practical and a theoretical point of view: First, the rules of minimal length serve the purpose of facilitating the automatic analysis with NLP tools. However, these do not handle null arguments well, which makes the analysis in **??** pointless. Second, the assumption of null arguments is quite expensive from a linguistic point of view and should only be made, if there is concrete linguistic evidence for it. It is not suited as a default annotation. Third, if the assumption of elliptical constructions is plausible, it is the easier analysis from a linguistic point of view, and therefore to be preferred, even if the construction is in fact ambiguous. Therefore, example 16a shows the desired annotation.

However, there are cases where the analysis of a construction as elliptical is not plausible, such as in example 17.

- (17) a. **Correct:** { das andere so Hühnerserb genannt wird / hat Blättlein } { PRO sind ein wenig rauher und ringsumher gekerbt / [...] }
 b. **Wrong:** { das andere so Hühnerserb genannt wird / hat Blättlein ~~Blättlein~~ sind ein wenig rauher und ringsumher gekerbt [...] }¹⁶

If the referent (*Blättlein*) is adjacent to the position where it has to be repeated, it would commonly be replaced by a pronoun (*die*), not by itself. However, elision only works for the same linguistic material. That is, it is not plausible to assume, that the original structure is *Blättlein Blättlein*, it would rather be *Blättlein die*, which then could not be elided. Hence, the analysis as a null subject is more plausible in this special case.

¹⁵From ???, *WieSichMeniglich_1557_vonBodenstein*, Z. 1593ff.

¹⁶From ???, *NewKreuterbuch_1543_Fuchs*, Z. 3395ff.

Conditionals with *Spitzenstellung* The restriction that the first rule of minimal length only applies to cases, where the status of a clause is not disambiguated by text coherence, especially applies for conditional clauses. This is illustrated in example 18.

- (18) a. **Correct:** { willst du diese Arznei kräftiger haben / nimm starken Weinessig zehn Lot / [...] }
 b. **Wrong:** { willst du diese Arznei kräftiger haben / } { nimm starken Weinessig zehn Lot / [...] } ¹⁷

This is a V1-conditional clause with a so called integrated *Spitzenstellung*, i.e. the protasis (*willst du diese Arznei kräftiger haben*) is located in the prefield of the adoposis (*nimm starken Weinessig zehn Lot [...]*) (koenig1988). This construction is also well known in contemporary German. In this case, the protasis is dependant on the adoposis, as it is equivalent to the subordinate clause variant of a protasis with *wenn*. Therefore, all portases, whether the conditional clause is integrated or resumptive (adoposis with *dann*) are analysed as part of the ENHG-TU containing the adoposis.

This clearly differs from how other V1-clauses are treated, such as imperatives, which typically form separate ENHG-TUs:

- (19) a. **Correct:** { willst du diese Arznei kräftiger haben / nimm starken Weinessig zehn Lot / ein halbes Lot der grünen Wermutblätter / Salz ein Drittel eines Quäntchens / } { mische diese Stücke wohl zusammen } { und trink es warm }
 b. **Wrong:** { willst du diese Arznei kräftiger haben / nimm starken Weinessig zehn Lot / ein halbes Lot der grünen Wermutblätter / Salz ein Drittel eines Quäntchens / mische diese Stücke wohl zusammen und trink es warm } ¹⁸

3.7 Minimal length II

Rule If

- a. a phrase is structurally ambiguous with respect to its attachment to two ENHG-TUs, and
- b. it is not possible to disambiguate the structure based on textual coherence,

the phrase is considered to be attached to the shorter ENHG-TU. The length of an ENHG-TU is defined in terms of tokens. If both ENHG-TUs in question contain the same amount of tokens, the phrase is attached to its preceding ENHG-TU.

Description This rule leads to more ENHG-TUs of shorter length. These are preferred for practical reasons, since NLP tools such as parsers work more efficient and less error prone with shorter sentential units. Also, ENHG is known to feature particularly long sentences, which exceed the average sentence length in contemporary German by far. Example 20 shows, how the rule is applied.

¹⁷From ???, NewKreuerbuch_1563_Handsch, Z. 597ff.

¹⁸From ???, NewKreuerbuch_1563_Handsch, Z. 597ff.

- (20) a. **Correct:**
 b. **Wrong:**

Troubleshooting

Shared arguments There are cases, where a phrase is the argument of two, otherwise separated potential ENHG-TUs, as shown in example 22. In these cases, the second rule of minimal length might conflict with the rules of uniqueness and the base definition, as the shared argument cannot be argument to two different ENHG-TUs.

- (21) a. **Correct:** { das andere so Hühnerserb genannt wird / hat Blättlein } { PRO sind ein wenig rauher und ringsumher gekerbt / [...] }
 b. **Wrong:** { das andere so Hühnerserb genannt wird / hat Blättlein Blättlein sind ein wenig rauher und ringsumher gekerbt [...] } ¹⁹

In this specific example case, the issue is solved by assuming a null subject. However, more common are cases, where the missing argument is interpreted as elided, leading to a segmentation of both potential ENHG-TUs into a single ENHG-TU, as illustrated in example ??.

- (22) a. **Correct:**
 b. **Wrong:**

This is the preferred option. A null argument is only assumed, if for some reason the elision analysis is not plausible. For details, see the troubleshooting subparagraph on elision and null arguments under the first rule of minimal length, in section 3.6.

3.8 Continuity

Rule ENHG-TUs are continuous strings of tokens. Discontinuous ENHG-TU are not possible, except if some meta text was inserted into a sentence.

Description This rule is primarily motivated by pragmatic considerations: most NLP tools work with continuous units. From a linguistic perspective, most sentences are in fact continuous, but they do not necessarily have to be so, as illustrated in example 23.

- (23) Star Wars Episode vier – das habe ich schon immer gesagt – ist und bleibt der beste Star Wars Film.

Depending on the given sentence definition, one might argue that in fact two sentences are present in this example, because the parenthesis is an entire main clause on its own. However, based on this guidelines, the example shows only one ENHG-TU.

Troubleshooting

¹⁹From ???, NewKreuterbuch_1543_Fuchs, Z. 3395ff.

Metatextual insertions Meta texts such as glosses are not part of the text in a strict sense. Therefore, it is important to analyse the elements of different text layers independently of each other. This is especially important in texts originating from the Ridges corpus, because in the XLSX format they are inserted within the main text and only marked by according annotations at the NOTE layer.

Appositions & parentheses TODO

Unlike meta textual insertions, appositions and parentheses are part of the main text and, therefore, have to be analysed on the same level. Depending on the type of insertion, appositions and parentheses may qualify as ENHG-TU. The procedure is as follows:

If

1. the insertion does not qualify as ENHG-TU, it is analysed as a part of the surrounding or adjacent ENHG-TU.
2. the insertion does qualify as ENHG-TU , and if
 - (a) the insertion is located at the left- or rightmost periphery of the ENHG-TU, it is to be analysed as a separate ENHG-TU.
 - (b) the insertion is located within another ENHG-TU, it is to be analysed as part of this ENHG-TU.

Close appositions are, therefore, usually part of their surrounding ENHG-TU, while wide appositions and parentheses are analysed based on their position and internal structure. Example 24 shows a ENHG-TU-internal apposition, which actually qualifies as a ENHG-TU, as it may be considered as a complete grammatical sentence, except that it is surrounded by another ENHG-TU.

- (24) a. **Correct:** { Wiewohl ich weiß dass nicht allein eben daran gelegen und genügsam ist / wann du dir Kräuter bringen lasset / und ihre Tugend in Büchern leset / sondern dass auch / hoch notwendig zu wissen in welcher Zeit und an was Orte sie gewachsen sind / wann und wie ihre Einsammlung geschehen / in was Masse Proportion dieselbigen zum Brauch gegeben und dargereicht müssen werden / und welche in ihrem Leben / **das ist / wann sie noch grün und saftig sind** ihre Tugend bald erzeugen / ja auch welche nach ihrem Sterben / wann sie gedörret dennoch von Gott Gnade empfangen kräftig ihre Eigenschaft zu erzeugen. }
- b. **Wrong:** { Wiewohl ich weiß dass nicht allein eben daran gelegen und genügsam ist / wann du dir Kräuter bringen lasset / und ihre Tugend in Büchern leset / sondern dass auch / hoch notwendig zu wissen in welcher Zeit und an was Orte sie gewachsen sind / wann und wie ihre Einsammlung geschehen / in was Masse Proportion dieselbigen zum Brauch gegeben und dargereicht müssen werden / und welche in ihrem Leben / { **das ist / wann sie noch grün und saftig sind** } ihre Tugend bald erzeugen / ja auch welche nach ihrem Sterben / wann sie gedörret dennoch von Gott Gnade empfangen kräftig ihre Eigenschaft zu erzeugen. } ²⁰

²⁰From ???, WieSichMeniglich_1557_von Bodenstein, Z. 11ff.

Example 26 shows an appositive nominal phrase at the outermost right sentence periphery of another ENHG-TU, which is analysed as a separate ENHG-TU.

- (25) a. **Correct:**
b. **Wrong:**

In contrast to this is example ??, in which the address is inside of another ENHG-TU and, therefore, analysed as part of this ENHG-TU

- (26) a. **Correct:** { nach langer Trübsal / **Christenleute** / in ewigem Leben/ auch bedeutet die rote Butte / darin der Samen liegt / wie sehr zunahm die christliche Kirche auf dieser Erde von Christus gehalten lieb und wert }
- b. **Wrong:** { nach langer Trübsal / } { **Christenleute** / } { in ewigem Leben/ auch bedeutet die rote Butte / darin der Samen liegt / wie sehr zunahm die christliche Kirche auf dieser Erde von Christus gehalten lieb und wert } ²¹

3.9 Sentence ending punctuation

Rule Sentence ending punctuation has to be located at the outermost right periphery of a ENHG-TU.

Description This rule is the only acknowledgement to a graphematic sentence definition. Even with a consistent punctuation, which is not to be expected in ENHG, it is consistent with the classical definition of t-units to produce smaller units than graphematic sentences, for example in the case of coordinated matrix clauses. However, neither regular t-units nor ENHG-TUs are supposed to cross unambiguously sentence ending punctuation.

Also, this rule regulates the position of the punctuation between ENHG-TUs: they are located at the right periphery of the preceding ENHG-TU, not at the left periphery of the following ENHG-TU. Even if the punctuation initially is ambiguous between within and between ENHG-TU punctuation, after assigning ENHG-TU boundaries the punctuation becomes clearly sentence final and, therefore, belongs to the preceding ENHG-TU. This is illustrated in example 27.

- (27) a. **Correct:** { Das heimisch Eppich ist wohlschmeckend / } { aber es ist dem Haupt böse und erweckt den wallenden Siechtum }
- b. **Wrong:** { Das heimisch Eppich ist wohlschmeckend } { / aber es ist dem Haupt böse und erweckt den wallenden Siechtum } ²²

Troubleshooting

Unexpected ambiguities Dots are not unambiguously sentence ending in ENHG. Unlike in contemporary German, dots may end a sentence, but they can also occur within sentences, like commas, see ????. Example 28 shows such a non sentence ending dot.

²¹From ???, Paradeiszaertlein_1588_Rosbach, Z. 3767ff.

²²Aus ..., BuchDerNatur_1482_vonMegenberg.xlsx, l. 1525ff.

- (28) a. **Correct:** { das ist mehr für gelehrte Leute / als die sich mit dem Werk zu belustigen begehre. In welchem er aus Plinius sehr viel genommen }
- b. **Wrong:** { das ist mehr für gelehrte Leute / als die sich mit dem Werk zu belustigen begehre. } { In welchem er aus Plinius sehr viel genommen } ²³

²³Aus Rhagor, Daniel (1639): Pflantz-Gart. Bern. Ben Stephan Schmid. 1-10., PflantzGart-Vorrede_1639.xlsx, l. 985ff.