

The *Anselm Corpus*: Methods and Perspectives of a Parallel Aligned Corpus

Stefanie Dipper

Department of Linguistics
Ruhr University Bochum
dipper@linguistics.rub.de

Simone Schultz-Balluff

German Department
Ruhr University Bochum
simone.schultz-balluff@rub.de

Abstract

This paper presents ongoing work in the *Anselm project* at Ruhr-University Bochum. We first present our corpus, which consists of about 50 versions of the medieval text *Interrogatio Sancti Anselmi de Passione Domini* ('Questions by Saint Anselm about the Lord's Passion'), written up in different dialects from Early New High German and Middle Low German. The versions were transcribed in a diplomatic way, and are currently being normalized and annotated with lemma and part of speech. In addition, the versions are being aligned at different levels of granularity (paragraph, sentence, word). We describe two use cases that profit from the annotations, one use case from the historical-cultural domain, the other from the linguistic domain. We finally sketch further application scenarios from other fields of Digital Humanities.

1 Introduction

This paper deals with the *Anselm Corpus*, a corpus consisting of more than 50 texts of the medieval tract *Interrogatio Sancti Anselmi de Passione Domini* ('Questions by Saint Anselm about the Lord's Passion'). The corpus is being created and annotated in the context of two cooperating projects at Ruhr-University Bochum.

The different texts are not just copies from some source but show considerable variation and, at least in parts, seem to be independent creations. As a consequence, we treat all texts equally, in contrast to most other historical text editions. One

of the project goals is a digital edition which gives access to all texts of the corpus. Users of the edition will be able to search for important concepts, such as the Last Supper, and compare the different terms used for this concept in the different texts. The edition will also support linguistically-motivated queries, e.g. for investigating the position of verbal arguments or the relative order of auxiliary-verb sequences.

In this paper, we describe use cases of the digital edition. Based on a small passage, we illustrate research questions from two different areas. One research question concerns historical semantics (the vocabulary), and investigates the different terms used for the concept 'Last Supper', and their temporal and regional distributions. The second research question concerns historical linguistics (syntax), and deals with the position of complements and adjuncts.

The paper is structured as follows. In Sec. 2, we describe the corpus and its annotations. Sec. 3 and 4 present the two use cases in detail, followed by sketches of further application scenarios from other fields of Digital Humanities in Sec. 5. Sec. 6 presents an outlook.

2 The corpus

Interrogatio Sancti Anselmi de Passione Domini represents a tract of the passion. It is the only German version in form of a dialogue. St. Anselm fasts and prays and implores Virgin Mary to reveal the events of the passion. She finally appears to him and grants his wish. He then starts asking questions, which she answers, about the Passion of Christ, beginning with the Last Supper and

ending with the entombment.

Versions The tract has been preserved in different versions, which put emphasis on different aspects of the narration. Based on the different foci and other general properties, the 50 texts can be grouped in 3 different versions: (i) verse versions (“V”); (ii) short prose versions (“PS”); (ii) long prose versions (“PL”). They differ with regard to content and distribution:

- Verse versions (V) focus on Christ’s sufferings whereas prose versions (P) focus on the sorrows of Mary.
- Since verses are written in rhyme, V-versions are rather homogeneous. In contrast, prose as a less formal textform promotes extending the basic content in various ways.
- The opening of the V-versions is very detailed and emphasizes Anselm’s scariness and emotions at the moment of Mary’s appearance. It includes a justification by Anselm for invoking Mary. PS-versions only contain the basic content; details such as Mount Sion or Golden Gate are mentioned only in PL-versions. PL-versions also often address practical issues worth knowing, such as: What exactly are the “Ismaelitic pennies”? Why can’t Mary be alone in the streets after nightfall? How big were the nails used to crucify Jesus?
- V-versions have been preserved from the north and center of Germany, PS-versions from the north and south-east of the German-speaking countries (including Austria and Switzerland). PL-versions stem from the central and southern parts. Only eastern regions produced texts of different versions.

The texts have been written (and printed) in different German dialects, in Upper German, Middle German and Low German, from the 14th to 16th century. They have not yet been investigated in research to any mentionable extent.

Anselm’s first question Both use cases presented in the next sections focus on Anselm’s first actual question, which has been preserved in

44 German and 3 Dutch versions. In this question, Anselm asks Mary to describe the beginning of Jesus’ martyrdom. Mary starts by describing the Last Supper and the betrayal of Judas. Depending on the respective version (V, PS, PL), the answer can provide further details, such as elaborate explanations of the “Ismaelitic pennies” (which Judas receives for his betrayal), or it contains supplementary elements, such as the footwashing by Jesus.

Table 1 shows Anselm’s first question and the beginning of Mary’s answer in a verse version, and short and long prose versions.¹

Annotations Currently, the texts are being annotated semi-automatically with a normalized word form, lemma, and part of speech. Moreover, corresponding passages are being aligned across the texts, in particular related questions asked by Anselm, and corresponding sentences, phrases or words. For instance, the M9-phrase *zu den iuden piſchhofen* ‘to the Jews’ bishops’ and the W1-phrase *zw den furſten der Juden* ‘to the princes of the Jews’ (see the last lines in Table 1) would be aligned to each other.

The alignments represent the core annotation of our corpus. They support comparative investigations of the various texts and versions that are part of our corpus, as is illustrated in the next sections.

3 Use case I: the term and concept *Last Supper*

This section investigates the terms used for the ‘Last Supper’, by looking at temporal, regional, and type-related distributions and variance (cf. (Besch, 1967); in contrast to Besch’s studies, which focus on High German dialects, the

¹The versions are:

- Oldenburg (“O1”): verse version written in Low German, 2nd half of the 14th century; Landesbibliothek Oldenburg, Cim I 74.
- München (“M9”): short prose version in East Upper German, 15th century; Staatsbibliothek München, Cgm 4701.
- Wien (“W1”): long prose version in East Upper German, early 15th century; Österreichische Nationalbibliothek Wien, Cod. 2969.

In the context of the Anselm corpus, we defined new text sigla O1, M9, W1, etc., that we use throughout this corpus.

Oldenburg (“O1”, V)	München (“M9”, PS)	Wien (“W1”, PL)
Maria erft wil ik di vraghen ik bidde dattu mi willeft faghen Wu quam it erft to den pranghen dat din fone wart ge vanghen	Do fragt anzhelm[us] vnd fprach O aller liebftew fraw wie hüß fich an dez erften deins liebē chindes marter	Sand Anfhelm was von herczn vrö und fprach fag mir liebe fraw wie was der anfankh der marter dynes libn chindes
Ancelme hore dat ik di faghe Dat fchude an dem guden donerfdaghe Dat he mit finen jungheren faat Lepliken dat he mit on aat He gaf on fin vleifch vnd ok fin blöt Dat he vedder vor vns göt. [...] ^u Maria do fe de rede dreuen Wur was judas do ge bleuen Judas de leip alto hant Dar he de Vorften der jodden vant	Do fprach ma[r]ia Do mein chind an dem antloz tag daz leczt ezzen het mit feinē iung[er]n vnd von dem tifch gie Do gie iudas zu den iuden pifchofen	vnfer fraw fprach da mein libs chind het geeffenn mit feinen Jungern vor feiner marter daz left mal und da fy von tifch auf ftunden da gieng Judas fcarioth zw den furften der Juden

Table 1: Anselm’s first question and the beginning of Mary’s answer. Fragments from a verse version (Oldenburg, left column), a short prose version (München, central column), and a long prose version (Wien, right column).

Anselm corpus also contains texts from Low German, including Dutch).

The term *Abendmahl* ‘Last Supper’, denoting the last meal of Jesus and his disciples, is established in church language only in the early 16th century, heavily influenced by Martin Luther. Prior to that, different, but unambiguous terms had been in use to denote Jesus’ farewell dinner. In the following, the different verbalizations of this concept in selected versions of the Anselm text are analyzed, focussing on their temporal and spatial distribution in relation to the three versions V, PS, and PL. The distributions of the variants are displayed in Table 2. The table displays the regions according to their actual locations, starting with Alemanic (1a) and Bavarian (1b) at the bottom (= in the South), and ending with North Low German (5) on the top (= the North).

The table shows that in almost all **verse versions** (V, in the North: 3a–5), the “Last Supper” is simply described as the fact that Jesus *mit en at* ‘ate with them’, i.e. with his disciples.

In the P-versions, rather fixed (but different) phrases are used:

German **short prose versions** (PS, East and North: 1b–4b, 5) mainly use two terms: combinations of *lezt* ‘last’ plus *essen* ‘meal’ in regions 1b (Bavarian) and 2b (East Franconian), and *abent* ‘evening’ plus *essen* ‘meal’ in region 2b.

Similarly, the short versions from Dutch (4a) all use combinations of *abent* ‘evening’ plus *mal(tijt)* ‘meal(time)’. Interestingly, the compound *abentmal* is already spelled in one word in the 14th century version.

In the **long prose versions** (PL, South and center: 1–2), the combination *iungst* ‘youngest’ plus *mal/maz* ‘meal’ is predominant, occurring in regions 1a, 1b, and 2a.

All other occurrences are singular, e.g. *nacht mal* ‘night meal’ (1a, PL), *abent spise* ‘evening dish’ (4b, PS), or *osterlamp* ‘paschal lamb’ (2b, PL).

The data shows that term selection depends on the region *and* the type of version in combination. Beyond the dominance of certain terms that we mentioned above, no continuity of terms spanning larger regions or time periods can be observed.

The variance that we observe across versions but also within the prose versions seem to suggest that at that time, no general term had yet been established. Terms used already in the 14th century continue to be used in the 15th and 15th centuries; besides them, new forms and combinations were coined.

In one PL version (B2 from region 2a), the term used for the Last Supper is explicitly addressed, see Ex. (1).

(1) Da myn kint hatte gefzen mit fynen iungern daz

5. North Low German

- 14 –
- 15 V mit en at (Kh1), myt en at (Arnd1494)
- 16 V mit en ath (Arnd1521)

4a. Middle Dutch

- 14 PS auontmael (Am1)
- 15 PS auont mael (Le1)
- 16 PS auont maeltijt (Berntsz1523)

3a. Ripuarian

- 14 –
- 15 V mit yn as (KoeldÄ1492), mit yn as (KoeldJ1499)
- 16 V mit yn as (Neuss1509), mit yn as (Neuss1514),
myt yn as (Neuss1514/17)

2a. Rhenish Franconian

- 14 –
- 15 PL iungfte mafze/abend efzen (B2), Jungfte was
(St1)
- 16 –

1a. Alemanic

- 14 –
- 15 PL Iung mafz (Be1), ivngefte maz (Ka1), Iungft maff
(Stu1), iungft mal (N4), iüngft mafz (sa1), iungft
<...> (Sa1), iüngst male (Schau1496/97)
- 16 PL nacht mal (SG1)

4b. West/Eastphalian

- 14 V mit em at (D1), mit em at (D2), mit on aat (O1)
PS auent fpife (Wo1)
- 15 V meth em ath (f1)
- 16 –

3b. East Central German

- 14 –
- 15 V abint effin (D3)
PS obent brot (B1)
PL nachtmal (H1)
- 16 –

2b. East Franconian

- 14 –
- 15 PS abent effen (M7), das abent effen (M8), lecztiz
obñt ezfen (N1), leczft effē (N3)
PL iüngft effen (M4), das ofterlamp (We1)
- 16 –

1b. Bavarian

- 14 PL iungift mal (M1)
- 15 PS left effen (Me1), letz effen (M5), het geffen (M6),
lecz ezzen (M9), lecz ezzen (M10)
PL jungfte mal (M2), iwngift was (M3), abent effñ
(Sb1), left mal (W1)
- 16 PL des iungften mal (Hk1)

Table 2: Terms and phrases denoting the “Last Supper”, used in different regions, time spans, and version types.

lungfte mafze daz da heifzet daz abend efzen
 ‘As my child had eaten with his disciples the
 youngest meal which is called the evening meal’

lungfte mafze ‘youngest meal’ is probably a general term, whereas *abend efzen* ‘evening meal’ seems to be a more special term, highlighted by the author. However, *abend efzen* is not a fixed term as can be seen from the variance observed in region 2b in the 15th century.

The unsteadiness of the terms is also reflected by the fact that most instantiations are spelled in two words, and only few “real” compounds can be observed: *auontmal* (Am1, 4a) and *nachtmal* (H1, 3b), both from the 14th century. Moreover, these compounds reoccur later, but spelled in two words (in Le1, 4a, and SG1, 1a).

It is remarkable that the term that has finally been established in standard German, is the term from Middle Dutch.

The concept “Last Supper”: strategies of conceptualization (cf., e.g., Busse et al. (1994))

As we have seen, the prose versions, short and long, seem to struggle for verbalizing the concept “Last Supper” but do not arrive at a common, “standardized” term. The verse versions follow another strategy: they use the unspecific phrase *dat he mit on aat* ‘that he at with them’ but add specific temporal information when this happened: *an dem guden donerſdaghe* ‘on the good Thursday’ (O1).²

A similar specification strategy is also followed by some of the prose versions. Six PS versions³ add *an dem antlaz tag* ‘on the indulgence day’. “Antlaz tag” in general means “day of release/indulgence”, and it can be used to refer to Holy Thursday in particular. Two PL versions (H1, SG1), which use the term *nachtmal* ‘night meal’, add the specifications *am heiligen gründornſtage* (H1), *am hailgen grünen donstag* (SG1) ‘on the Holy Thursday’.

This data shows that the fact that there is not yet a mandatory agreed-upon term is compensated

²D3 (written in East Central German) represents a special case: It uses rhymes but otherwise shows characteristics of the prose versions. Especially its vocabulary deviates from the other verse versions. This suggests that D3 should be considered separately from the verse versions, and in connection with the prose versions.

³Texts N3, M5, M7, M8, M9, M10, Me1.

by specification strategies. We propose that the specific strategy used in a subset of the versions can be used as a defining criterion for the version “verse”.

4 Use case II: constituents in situ and extraposed

In the second use case, we select the first sentence of Mary’s first answer and compare its different syntactic realizations in all PL versions of our corpus (20 versions in total). In particular, we investigate the positions of verb arguments and adjuncts.

In (modern) German, the “right verbal bracket” indicates the boundary between nominal and prepositional arguments and adjuncts that occur “in situ” (preceding the right verbal bracket) or “extraposed” (following the verbal bracket).

In subordinate clauses, the right verbal bracket is filled by verbal components (verbs and auxiliaries), see Ex. (2a). In main clauses, the finite verb or auxiliary takes the second position after some other constituent, filling the “left verbal bracket” (this construction is called “verb-second”). Further verbal components, such as infinite verb forms, verb particles, can occupy the right verbal bracket, see Ex. (2b).⁴ The left and right verbal brackets are underlined in the examples. Constituents occurring “in situ” are marked by “INS”, extraposed constituents by “EX”. The examples illustrate that in modern German, arguments such as the subject and object occur in situ, whereas adjuncts can be extraposed (optionally).

- (2) a. *als* [INS *Jesus*] [INS *das Abendmahl*]
 as Jesus the Last_Supper
gegessen hatte [EX *mit seinen Jüngern*]
 eaten had with his disciples
 ‘as Jesus had eaten the Last Supper with his disciples’
- b. *Jesus hatte* [INS *das Abendmahl*]
 Jesus had the Last_Supper
gegessen [EX *mit seinen Jüngern*]
 eaten with his disciples
 ‘Jesus had eaten the Last Supper with his disciples’

The verb-second pattern can already been observed in Old High German, in addition to verb-

⁴For a description of the German sentence structure, e.g. see Höhle (1986).

first, verb-third, and verb-final patterns. In Middle High German, the verb-second pattern has been established as the common structure of main clauses. Verb-final patterns in subordinate clauses are predominant from the earliest stages on. However, as can be seen from Ex. (2b), extraposed constituents can occur after the final verb.

It is well known that arguments and adjuncts occurred in extraposed positions much more frequently in older language stages than nowadays.

Based on data from Gothic, Old English, and different stages from German, Behaghel (1932) shows that short constituents, consisting of one word, predominantly occur in situ, whereas long, “heavy” constituents, e.g. constituents involving coordination, tend to be extraposed. Ebert (1986) examines two texts from the 14th century and finds that around 20% of subordinate clauses contain extraposed constituents, predominantly PPs, but also NP complements.⁵

In the 17th century, the sentence-final position of the verb in subordinate clauses has been established in standard language (Behaghel, 1932, p. 133). That is, since that time, extraposition is limited to clausal arguments and PP adjuncts.

Ex. (3), taken from Behaghel (1932, p. 132), shows an example from Martin Luther with an extraposed object. This construction would be highly marked in modern standard German.

- (3) *wenn du erkennstest* [*EX die Gabe Gottes und wer der ist, der zu dir sagt, gib mir trinken*]
 ‘if you knew the gift of God and who it is that asks you for a drink’ (John 4:10)

The presentation above shows that in Early New High German, extraposition is still applied to a range of arguments. Hence, it is interesting to investigate the amount of extraposition and the type of arguments that are extraposed in the different Anselm texts. To do this, we analyse the

⁵The terms “in situ” and “extraposed” suggest that one of the positions is the “original”, unmarked one, while the other is a secondary position, derived from the first, e.g., by a relation called “extraposition”. For modern German, the unmarked positions of NP and PP constituents are clearly in front of the verbal bracket, and positions behind the verbal bracket are exceptional. In former stages of German, however, the situation is not as clear. Hence, the reader is asked to interpret the terms “in situ” and “extraposed” as referring to pre- and postverbal positions, without implications about the actual analysis.

first sentence of Mary’s first answer in detail, see in Table 3. The table displays the W1-text in the first column, organized by constituents, a translation in the second column, each constituent’s function in the third column, and its position in the fourth column. The sentence consists of two subordinate clauses, followed by the main clause.

As can be seen from the table, the subjects occur in situ. The remaining constituents of the first subordinate clause are extraposed, in contrast to the constituents of the second subordinate clause. The positions of the main clause constituents cannot be determined in this example because the right verbal bracket is not filled (but see below).

The distribution of the constituents, as realized in this text, is in fact the “default” distribution, which shows up in 11 of the 20 PL texts.⁶ In three texts,⁷ all constituents occur in situ. Interestingly, these texts share another unique feature: the verbal components of the first clause (line 3) show the modern order *verb participle* > *finite auxiliary*, e.g. *geffen hett* ‘eaten had’, in contrast to all other texts.

In 14 texts, the NP-object of the first clause (line 6) occurs after the PP-adjuncts (lines 4 and 5). In five texts,⁸ the NP-object occurs in front of the PP-adjuncts, and in one text,⁹ it occurs between both PPs.

In three texts,¹⁰ the right verbal bracket of the main clause is filled by a verb particle. In these cases, the locative PP, which denotes the goal of the movement (line 15), is extraposed, see Ex. (4).

- (4) *Do giench* [*INS ivdas fcarioth*] *vz.*
 then went Judas Iscariot out
 [*EX zv den fvrften d^h ivden*].
 to the princes of_the Jews
 ‘Then Judas went out to the high priests of the Jews’

Further differences between the texts include: absence of the second subordinating conjunction (line 8); absence of the subject in the second subordinate clause (line 9).

⁶From region 1a: N4, Stu1, Schau1496/97; region 1b: M1, M2, M3, W1, Hk1; region 2a: B2, St1; region 2b: We1.

⁷From region 1a: SG1; region 1b: Sb1; region 3b: H1.

⁸Region 1a: Be1, Ka1, sa1, SG1; region 3b: H1.

⁹Region 1a: Sa1.

¹⁰All from region 1a: Be1, Ka1, Sa1

Wien (“W1”, PL)			Function	Position	Clause
1	da	as	subord		} Subord 1
2	mein libs chind	my dear child	NP-subj	INS	
3	het geeffenn	has eaten	verb	right VB	
4	mit feinen Jungern	with his disciples	PP-adjunct	EX	
5	vor feiner marter	before his martyrdom	PP-adjunct	EX	
6	daz left mal	the Last Supper	NP-obj	EX	
7	und	and	coord		
8	da	as	subord		} Subord 2
9	fy	they	NP-subj	INS	
10	von tifch	off table	PP-pobj	INS	
11	auf ftunden	up stood	verb	right VB	
12	da	then	adverb		} Main
13	gieng	went	verb	left VB	
14	Judas fcarioth	Judas Iscariot	NP-subj	?	
15	zw den furften der Juden	to the princes of_the Jews	PP-goal	?	

Table 3: The beginning of Mary’s first answer: ‘As my dear child had eaten the Last Supper with his disciples before his martyrdom, and as they left the table, Judas Iscariot went to the high priests of the Jews’. INS: in situ, EX: extraposed, VB: verbal bracket,

To sum up the findings of this small comparison, we have seen that extraposition of the object NP seems to be the unmarked case, in contrast to modern German. Ignoring the case of subject NPs (which seem to be extraposed only rarely) and unclear positions, the numbers of constituents in situ vs. extraposed are almost equal in the default order: two adjunct PPs and one object NP occur extraposed, two argument PPs are in situ.

5 Further application scenarios

In Sec. 3, we focused on the different terms used for the Last Supper. In this section, we have a look at the way Mary and Jesus are addressed and referenced in the different texts. In contrast to the case of the Last Supper, the fact that different forms of address are used is meaningful and indicates different intentions and relations to these persons.

The forms of address for **Mary** vary considerably between the different texts.

In the prose versions, Anselm addresses Mary by forms indicating devotion: *leue vrowe* ‘dear woman’ (Wo1), or *aller liebftew fraw* ‘most dearest woman’ (M9). When talking about Jesus, Anselm emphasizes Mary’s role of the mother: *deins liben Kindes* ‘your dear child’ (N3). This culminates in the description of Mary as the

mother of all humans: *liebe mutt[er]* ‘dear mother’ (used by Anselm). Finally, the narrators of the prose texts involve the recipient, e.g., by phrases such as *vnser fraw* ‘our woman’ (e.g. W1) or *vnfe liebe frauwe* ‘our dear women’ (e.g. B2).

In the verse versions, the relation to Mary remains more reserved. She is addressed exclusively by her name. Still, her mother role is present in that Jesus is referenced by *din fon* ‘your son’.

When talking about **Jesus**, Mary refers to him by *min kint* ‘my child’ (Be1), *mein libs kint* ‘my dear child’ (W1) or even *min alre lieffte kint* ‘my most dearest child’ (Am1). In the verse versions, Jesus is predominantly referred to by the personal pronoun.

In short, the prose versions emphasize the relation mother–son, described from the point of view of the sorrowing mother, and establish a mother relation between Mary and the recipient of the text. The personal relationship is intensified by elaborate passages of lamentation. The verse versions remain more distant in general. The idea of compassion and the role of Mary as Mater Dolorosa plays a stronger role in the prose than in the verse versions.

The forms of address, rhetoric, and style allow us to draw conclusions with regard to the context

of use of individual texts or versions, and to the image of Mary and the intended recipients.

6 Conclusion

In this paper, we presented a parallel corpus of texts from Early New High German. We argued that alignments at different levels (question–answer pairs, sentences, phrases, words) can support comparative investigations in different areas. This was illustrated by different use cases from historical semantics (comparing terms used for the Last Supper), historical syntax (comparing the distribution of constituents), and from a historio-cultural perspective (comparing the ways Mary and Jesus are addressed and referenced).

We plan to create a digital edition of the entire corpus. Users can select texts from the collection and search for specific word forms, parts of speech etc. The query results will be presented in the form of a synopsis, which places aligned passages next to each other.

We think that the alignments can also support semi-automatic creation of a critical apparatus used in a print edition. The variance observed between the three versions (verse, short prose, long prose) suggests that all three versions would be edited. The variance could also lead to considerations whether we actually deal with one or three texts.

References

- Otto Behaghel. 1932. *Deutsche Syntax. Eine geschichtliche Darstellung. Band IV: Wortstellung. Periodenbau*. Heidelberg: Winter.
- Werner Besch. 1967. *Sprachlandschaften und Sprachausgleich im 15. Jahrhundert. Studien zur Erforschung der spätmittelhochdeutschen Schreibdialekte und zur Entstehung der neuhochdeutschen Schriftsprache*. München: Francke.
- Dietrich Busse, Fritz Hermanns, and Wolfgang Teubert, editors. 1994. *Begriffsgeschichte und Diskursgeschichte. Methodenfragen und Forschungsergebnisse der historischen Semantik*. Opladen: Westdeutscher Verlag.
- Robert Peter Ebert. 1986. *Historische Syntax des Deutschen II: 1300-1750*. Frankfurt: Peter Lang.
- Tilman Höhle. 1986. Der Begriff ‘Mittelfeld’. Anmerkungen zur Theorie der topologischen Felder. In A. Schöne, editor, *Kontroversen, neue und*

alte. Akten des 7. Internationalen Germanistenkongresses Göttingen 1985, pages 329–340. Tübingen: Niemeyer.