

Tatian Corpus of Deviating Examples

T-CODEX 2.1

Corpus Description

Svetlana Petrova

Project B4 in Collaborate Research Center on Information
Structure

*in cooperation with Carolin Odebrecht, student researcher of project
B4*

Humboldt University Berlin

Contents

1 Preliminaries	3
2 Tagset Declaration	4
2.1 Layer I: Source data	4
2.2 Layer II: Morpho-syntactic annotation.....	4
2.2.1 POS – Part of Speech.....	5
2.2.2 CAT – Syntactic Category	6
2.2.3 CLAUSE STATUS.....	7
2.2.4 GF – Grammatical Function	8
2.2.5 Syllable number	9
2.3 Layer III: Information Structure	9
2.3.1 GIVENNESS.....	9
2.3.2 TOP-COMM.....	10
2.3.3 FOC-BG – Focus-Background.....	11
2.3.4 Context.....	12
2.4 Meta-Information.....	12
2.4.1 COMMENT	12
2.4.2 BIBLIOGRAPHY	12
References.....	13

1 Preliminaries

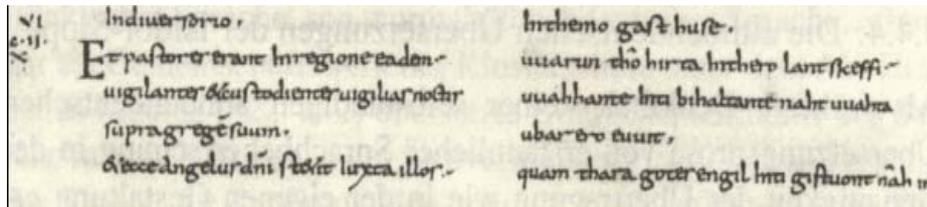
The present corpus, the Tatian Corpus of Deviating Examples T-CODEX 2.0, provides morpho-syntactic and information structural annotation of parts of the Old High German translation attested in the MS St. Gallen Cod. 56, traditionally called the OHG Tatian, one of the largest prose texts from the classical OHG period. This corpus was designed and annotated by Project B4 of Collaborative Research Center on Information Structure at Humboldt University Berlin.

The OHG Tatian text is a gospel harmony translated from Latin and written down in the middle of the 9th century by at least 6 scribes. In the MS, as Picture 1 shows, the Latin source and the OHG translation are attested as two juxtaposed columns. It has been noticed that each line in the OHG text translates exactly the same material found in the corresponding Latin line (cf. Masser 1997 a and b). A new diplomatic edition made available by Masser (1994, cf. Picture 2) reflects these major characteristics and makes it possible to compare the source and target text. We are grateful to Prof. Masser (Innsbruck) as well as to Vandenhœck & Ruprecht publishing company (Göttingen, Germany) for allowing us to publish parts of the diplomatic text edition by Masser (1994) in our corpus.

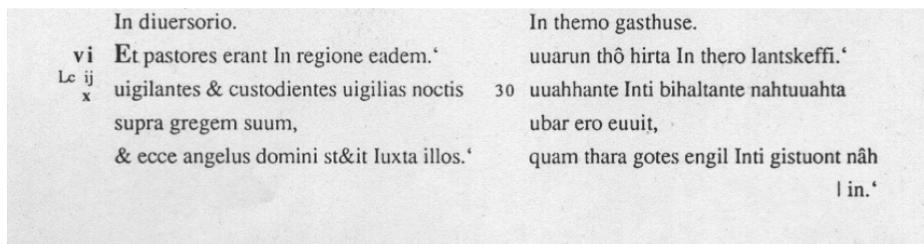
A fundamental question concerning medieval translations from Latin is whether the attested word order patterns and constructions are influenced by the syntax of the original or rather representative for the native grammar of the target language. Recent work on the syntactic analysis of OHG has established as a methodological precondition the restriction on clauses that differ from the order of the Latin original (Fleischer 2006). Such instances are called *deviating examples*. They are generally estimated as evidence for genuine OHG structures (cf. Dittmer and Dittmer 1998, Donhauser 1998). A detailed examination of such examples has shown that the deviations from the Latin occurring in the OHG Tatian part are not random but apply systematically throughout the text (cf. Fleischer, Hinterhölzl and Solf 2008).

The present corpus compiles ca. 2.000 deviating examples found in the text portions of the scribes α , β , γ and ϵ . Each clause structure represents an extra file annotated with the annotation tool EXMARaLDA and searchable via ANNIS, a general-purpose tool for the publication, visualisation and querying of linguistic data collections, developed by Project D1 of the Collaborative Research Center on Information Structure at Potsdam University.

Picture 1: The beginning of Luke 2,8 in the manuscript of the St. Gallen, Stiftsbibl. Cod. 56. Facsimile, pag. 35. In Sonderegger (2003, 130).



Picture 2: The same part of the text in the edition of Masser (1994, 85).



2 Tagset Declaration

The objectives and design principles of a pilot version of T-CODEX 1.0 are described in Petrova et al. (2009).

2.1 Layer I: Source data

In this layer, we provide a line with the Latin source and the OHG translation of the text. In a separate alignment layer, we provide information which OHG item corresponds to which item in the Latin line. Please note that only the OHG text is annotated. In case two words are spelt together in the original, without a space in-between, e.g. because of cliticisation, we treat the two words as separate items but place '+' on the clitic, cf. (1):

- (1) *niuuari* (T 47, 10)
 NEG.PRT-be.3.Sg.PAST.SBJ
 ni+ uuari

2.2 Layer II: Morpho-syntactic annotation

Below we provide a list of the tags used in T-CODEX 2.0. The tag-set declaration is organized alphabetically for each layer of annotation. In general, we aimed at sticking to the canonical definitions of the terms applied. Whenever we apply specific notions or use

common terms in a non-canonical way, we provide additional explanation in the third column.

2.2.1 POS – Part of Speech

2.2.1.1 Inflected parts of speech

A	adjective	
N	noun	
NUM	numeral	Numeral is used as a cover term for cardinal numbers, ordinal numbers and some special classes like 'both', 'lots of', 'half' (cf. Helbig and Buscha, 1987, 320 f.) Eg.: <i>zuelifiaro</i> 'twelve years', T_096_31-32
PRONDEM	demonstrative pronoun	In OHG, demonstrative pronouns are also used as definite determiners.
PRONINDEF	indefinite pronoun	This is a cover term for indefinites and quantifiers in OHG including items like 'many', 'all', 'some' etc. This classification reflects the standard interpretation of these classes of words in traditional grammar books on OHG.
PRONINTERR	interrogative pronoun	
PRONPRS	personal pronoun	
PRONPOS	possessive pronoun	According to the standard view, there is no formal class of possessive pronouns in OHG. We assign this tag to the genitive form of personal pronouns used to express possessive relations.
PRONRFL	reflexive pronoun	
PRONREL	relative pronoun	In OHG, there is no formal class of relative pronouns. We assign this tag to different demonstrative pronouns used to introduce a relative clause.
PRONREZ	reciprocal pronoun	
VAUX	auxiliary verb	This tag is assigned to the verbs <i>be</i> and <i>become</i> in combination with the past participle of the main verb when used to translate a Latin passive. The third of the modern German auxiliaries, the verb <i>have</i> as a formal constituent of the perfect construction, is rare in this function in the OHG Tatian.
VCOP	copula verb	A copula is a verb that links the subject or an object to another phrase, adjective etc. which is part of a complex predicate. The complement of a copula is annotated as PRDNOM at the level of grammatical function (GF).

VMOD	modal verb	Modals are a class of verbs that are used most often with an infinitive to express ability, disposition or obligation. Pay attention to the fact that the list of modal verbs in OHG is larger than in modern German.
V	main verb	This tag is assigned to the main lexical verb in the clause.
VN	verbal noun	According to the standard literature the category of 'verbal noun' in Old High German is a cover term that comprises non-inflected and inflected infinitives (gerund) and the two participles, the present and past participle. (Braune 2004, § 301.5, p.257) E.g.: non-inflected infinitive: <i>ziohan</i> 'to draw, to pull'; inflected infinitive after a preposition (<i>zi</i>) which selects the dative case (gerund): <i>zigebanne</i> , T_031_23; present participle: <i>sehenti</i> 'seeing/ looking', T_026_21; past participle: <i>gitruobet</i> 'aggrieved, saddened', T_026_20

2.2.1.2 Non-inflected parts of speech

ADV	adverb	This tag is assigned to adverbs of all kinds, i.e. to adverbs of time, place, incl. location and direction, and to adverbs of cause and manner.
CONJ	conjunction	This tag is assigned to conjunctions in general. We do not differentiate at this level between coordinating and subordinating conjunctions.
INTERJ	interjection	
NEG	negation word	In this category we subsume words that express sentence negation in Old High German, for example the negative particle <i>ni/ ne</i> which cliticizes on the finite verb, like in <i>Ni quam ih</i> 'I didn't come' or words like <i>nalles</i> 'and not', <i>noh</i> 'neither'.
PTC	particle	Non-inflected parts of speech with no proper denotation, but special functions which contribute to the truth value of the sentence. Eg: <i>uuârlihho</i> , <i>ía</i> , <i>enu</i>
P	preposition	

2.2.2 CAT – Syntactic Category

AP	adjective phrase
ADVP	adverbial phrase
CP	sentential complement
NP	nominal phrase
PP	prepositional phrase
VP	verbal phrase

2.2.3 CLAUSE STATUS

Apart from the traditional classification of main and subordinate clauses, we distinguish causal and relative clauses as separate classes for the reasons described below.

CAUSAL	Clauses introduced by the causal conjunctions <i>uuanta</i> or <i>bithiu uuanta</i> 'because' in Old High German are ambiguous between main and subordinate clauses. For this reason, they are classified as causal uniformly.	
MAINDECL	main declarative clause	This tag is assigned to main declarative clauses.
MAINIMPER	imperative clause	This tag is assigned to main clauses which use the imperative mood to utter requests.
MAININTERR	main interrogative	This tag is assigned to main interrogative clauses. No further distinction is made between constituent (wh-questions) and propositional questions.
MAINREQUEST	request	This tag is assigned to main clauses which use the subjunctive mood to utter requests.
MAINOPTAT	main optative sentences	This tag is assigned to clauses expressing wish and desire which contain the verb in the subjunctive mood.
REL	This tag is given to attributive relative clauses modifying a nominal argument as well as to free relatives (headless relative clauses) which provide the subject or the object of the matrix clause. Note that attributive relative clauses introduced by demonstrative pronouns in relative function in Old High German are sometimes ambiguous between main and subordinate clauses. Cf. <i>sum tuomo uuas In sumero burgi/ ther niforhta got</i> (T 200, 30-31) lat. <i>Iudex quidam erat In quadam ciuitate/ qui deum non timebat</i> 'a certain judge was in a certain town DEM NEG-feared God	
SUB	subordinate clause (no further specification)	This category comprises finite complement clauses. Adverbial clauses are further subdivided by type (except of causal clauses which are kept separately).
SUBCOMPAR	a clause of comparison	This type of subclause expresses comparison.
SUBCONC	concessive clause	This type of subclause expresses a concessive relation wrt the matrix clause. In modern German, such clauses are introduced by <i>obwohl</i> etc.
SUBCOND	conditional clause	This type of subclause expresses conditions.
SUBCONS	consecutive clause	This type of subclause expresses consequences.
SUBFINAL	purpose clause	This type of subclause expresses intentions.
SUBLOCAL	local clause	This type of subclause expresses embedded locations.
SUBMODAL	modal clause	This type of subclause expresses manner.
SUBTEMP	temporal clause	This type of subclause expresses temporal location or extension.

PARTCONSTR Participial constructions are used in various functions in Old High German, e.g. to specify the circumstances of the matrix event or to modify a nouns in the matrix clause. As they lack a finite verb form they are kept apart from finite subordinate clauses.

INFCLAUSE infinitival clause This tag is assigned to non-finite complement clauses.

2.2.4 GF – Grammatical Function

ADVCAUS	adverbial of cause	This is an adverbial of cause.
ADVCOM	comitative adverbial	This is an adverbial denoting an accompanying person.
ADVCOND	conditional adverb	This adverbial denotes a condition.
ADVINSTR	instrumental adverbial	This is an adverbial of instrument.
ADVLOC	adverbial of location	This is an adverbial of location.
ADVMOD	adverbial of manner	This is an adverbial of manner.
ADVSENT	sentential adverb	This kind of adverbial relates to the entire proposition.
ADVTEMP	temporal adverbial	This is an adverbial of time and temporal duration.
APPOS	apposition	An apposition is a repetition of an argument expression which provides further information, e.g. <i>zachariás sin fater</i> 'Zacharia, his father'.
ATTR	attribute	An attribute is a modifier to a noun.
ATTRPRDNOM	attribute to the predicate noun	This attribute modifies a predicate noun.
DIR	directional argument	This tag is assigned to the directional arguments of verbs of movement, e.g. <i>in b&shleem in santa sie in b&shleem</i> 'And sent them to Bethlehem.', T_040_02
DO	direct object	
GERUND	gerund	This tag is forced by the subject matter of the project B4, namely to investigate verb-object orders in OHG. It is of utmost importance to distinguish between tensed and untensed verb forms and to have a quick access to these categories in the annotation. (cf. Pintzuk 1999).
INF	infinitive	see above
IO	indirect object	
PARTPF	past participle	This tag is forced by the subject matter of the project B4, namely to investigate verb-object orders in OHG. It is of utmost importance to distinguish between tensed and untensed verb forms and to have a quick access to these categories in the annotation. (cf. Pintzuk 1999).

PARTPR	present participle	see above
PO	prepositional object	This is a prepositional phrase which is selected as an argument of a verb or an adjective. E.g. <i>in inan</i> in <i>giloubtun in inan thó sine iungiron</i> 'His disciples started to believe in Him.', T_056_10; <i>fora gote</i> in <i>siu uuarun rehtiu beidu fora gote</i> 'They were both righteous before God.' T_026_3
PRD	predicative attribute	This is a traditional term which expresses secondary predication. E.g. <i>blint</i> in <i>thaz her blint uuwrđi geboran</i> 'That he was born as a blind man.', T_220_17
PRDNOM	predicate noun	The predicate noun (germ. <i>Prädikatsnomen</i>) is a term used in classical philology to describe the obligatory argument of copula verbs like 'to be', 'to become', 'to call', etc. The predicate noun can be a noun, an adjective, a prepositional phrase and even a whole sentence. E.g. <i>uuituua</i> in <i>thiu uuas uuituua</i> 'She was a widow.'; <i>cephas</i> in <i>thu bist giheizzan cephas</i> 'Your name will be Cephas.'
SUBJ	subject	The nominative NP which agrees to the finite verb in the clause
VFIN	inflected verb	This tag at the level of grammatical function is forced by the subject matter of the project B4, namely to investigate verb-object orders in Old High German. It is of utmost importance to distinguish between tensed and untensed verb forms and to have a quick access to the finite verb in the annotation. (cf. Pintzuk 1999)
VOC	vocative expression	An expression referring to a person to which the utterance is addressed, e.g. <i>truhtin</i> , <i>meistar</i> or <i>fater</i> . The vocative expression typically occurs outside of the clause and not in an argument position selected by the predicate.

2.2.5 Syllable number

Here, we note the number of syllables per syntactic constituent.

2.3 Layer III: Information Structure

2.3.1 GIVENNESS

The following tags are assigned to discourse referents only. For more details on entities that have to be annotated at this level, see the Annotation Guidelines of the SFB, Part on Information Structure, (Dipper et al. 2007, Section 3.2.1.). We assign the following tags:

ACC	accessible	This tag is assigned to expressions that are accessible via contextual reference to a given entity or to the assumed world knowledge of the interlocutors.
GIV	given	This tag is assigned to expressions that refer to entities with a given antecedent in the context.
NEW	new	This tag is assigned to expressions that refer to entities with no antecedent in the context.

2.3.2 TOP-COMM

Annotation at this level applies if the sentence favours an unambiguous topic-comment reading. We decompose the notion of topicality into various features that are related to topics in the literature. Next to informational status annotated separately at the level of GIVENNESS, we especially look at the application of standard topic-tests (2.3.2.1), the use of topic-marking constructions (2.3.2.2), the use of morphological topic markers (2.3.2.3) and the semantic type of NPs (definite vs. indefinite specific vs. existential reading of NPs, cf. 2.3.2.4)

TOP	topic	If the sentence favours a topic-comment reading as described above, we mark the constituent that is put in A for topic.
COMM	comment	The rest of a clause with a topic is annotated for comment. If the entire clause is annotated for comment, this implies that a topic-comment division applies but there is no overt topic phrase (e.g., due to topic-drop etc.).

2.3.2.1 ABOUTNESS

This tag is given to phrases that are potential candidates for topics in the clause. With much current work, we distinguish between frame-setting and aboutness topics as instances of an overall notion of topicality, cf. Jacobs (2001).

FRAME	frame	This tag is assigned to frame-setting topics, i.e. to expressions that reflect identifiable time spans and locations within which the predication holds. For more details, see Annotation Guidelines of the SFB (Dipper et al. 2007), Sec. 4.2.3.
AB	aboutness	This tag is given to expressions that are possible candidates for A in some well-known topic tests like “As for X, A says that X...” or “A says about X that X...”.

2.3.2.2 POSITION: TOPIC-MARKING CONSTRUCTIONS

In this tire, we mark the occurrence of a topic-marking construction. We follow the literature in assuming that left dislocation (LD) and hanging topic (HT) are typical topic-marking strategies, though differences in the pragmatic role of both are considered. Additionally, we distinguish the clause-initial constituent of a V2 main clause (German Vorfeld = GV). This distinction only concerns root declaratives. A consistent extension over subordinate clauses will be provided in future.

LD	left dislocation	In this construction, a left dislocated element is resumed by a pronoun in the prefield of the main clause. Note that the pronoun and the left dislocated element agree in case, gender and number.
HT	hanging topic	In this construction, a dislocated element and its resuming pronoun do not necessarily agree in case, number and gender.
GV	German Vorfeld	This tag is assigned to the clause-initial constituent of a V2 main clause.

2.3.2.3 TOPIC MARKERS

Here, we put in formal markers of topicality, e.g. adverbs, particles or modifiers that licence the topical status of a constituent.

2.3.2.4 DEFINITENESS

At this level, we annotate the semantic properties of phrases with respect to the definite/indefinite distinction. In OHG, the system of definite determiners is not fully developed yet. Therefore, we annotate the semantic interpretation of the phrase despite of the presence of a definite or indefinite marker.

DEF	definite
INDEF	indefinite

2.3.3 FOC-BG – Focus-Background

On the definition of focus and background, as well as on the different semantic types of focus, see Annotation Guidelines of the SFB (Dipper et al. 2007), Sec. 5. Please note that second occurrence focus is marked by an index `cf_1` and `cf_2` in the annotation.

CF	contrastive focus	
NIF	new-information focus	
NIFCF		This combination of tags suggests that new and contrastive focus co-occurs on the same XP.
CF_01		In cases of multiple foci, we assign indices to the foci belonging to different contrastive pairs.
CF_02		see above

2.3.1 FOC-MARKER

In this layer, we annotate focus sensitive operators like additive *also*, exclusive *only*, etc., as well as *wh*-words.

FOPR	focus sensitive operator
WH	question word

2.3.4 Context

Here, we annotate the relation of the clause with respect to the preceding context, especially in terms of discourse organization. We mainly distinguish between “situation shift” (*Situationswechsel*) und “anaphoric reference” (“AR”). *Situationswechsel* involves a change of time, place, speaker/narrator and participants.

2.4 Meta-Information

2.4.1 COMMENT

This is a line for individual remarks and comments of the annotator.

2.4.2 BIBLIOGRAPHY

We provide the respective information according to manuscript page and line, scribe and concordance notes.

References

Braune, Wilhelm

- 2004 *Althochdeutsche Grammatik I - Laut- und Formenlehre*, 15. edition by Ingo Reiffenstein. Tübingen: Niemeyer.

Dipper, Stefanie, Goetze, Michael and Skopeteas, Stavros

- 2007 *Information Structure in Cross-Linguistic Corpora. Annotation Guidelines for Phonology, Morphology, Syntax, Semantics and Information Structure*. Special issue of *Working Papers of the SFB 632 (ISIS)* 7.

Dittmer, Arne and Ernst Dittmer

- 1998 *Studien zur Wortstellung – Satzgliedstellung in der althochdeutschen Tatianübersetzung*. Göttingen: Vandenhoeck & Ruprecht.

Donhauser, Karin

- 1998 Negationssyntax im Althochdeutschen. Ein Sprachhistorisches Rätsel und der Weg zu seiner Lösung. In *Deutsche Grammatik. Thema in Variationen. Festschrift für Hans-Werner Erms zum 60. Geburtstag*, eds. Karin Donhauser and Ludwig M. Eichinger, 283-298. Heidelberg: Carl Winter.

Fleischer, Jürg

- 2006 Zur Methodologie althochdeutscher Syntaxforschung. *Beiträge zur Geschichte der deutschen Sprache und Literatur* 128:25-69.

Fleischer, Jürg, Roland Hinterhölzl and Michael Solf

- 2008 Zum Quellenwert des AHD-Tatian für die Syntaxforschung: Überlegungen auf der Basis von Wortstellungsphänomenen. *Zeitschrift für germanistische Linguistik* 36: 211-240.

Helbig, Gerhard and Joachim Buscha

- 1987 *Deutsche Grammatik*. 10th edition. Leipzig: Verlag Enzyklopädie.

Jacobs, Joachim

- 2001 The dimensions of topic-comment. *Linguistics. An interdisciplinary journal of the language sciences* 39 (4): 641-681.

Masser, Achim ed.

- 1994 *Die lateinisch-althochdeutsche Tatianbilingue Stiftsbibliothek St. Gallen Cod. 56*. Göttingen: Vandenhoeck & Ruprecht.

Masser, Achim

- 1997a Syntaxprobleme im althochdeutschen Tatian. In *Semantik der syntaktischen Beziehungen. Akten des Pariser Kolloquiums zur Erforschung des Althochdeutschen 1994*, ed. Yvon Desportes, 123-140. Heidelberg: Carl Winter.

Masser, Achim

- 1997b Wege zu gesprochenem Althochdeutsch. In *Grammatica Ianua Artium. Festschrift für Rolf Bergmann zum 60. Geburtstag*, eds. Elvira Glaser and Michael Schläefer, 49-70. Heidelberg: Carl Winter.

Petrova, Svetlana, Solf, Michael, Ritz, Julia, Chiarcos, Christian, Zeldes, Amir

- 2009 Building and using a richly annotated interlinear diachronic corpus: the case of Old High German Tatian. *Traitement Automatique des Langues* 50 (2): 47-71.

Pintzuk, Susann

1999 *Phrase Structures in Competition. Variation in change in Old English word order*, New York: Garland.

Sonderegger, Stefan

2003 *Althochdeutsche Sprache und Literatur. Eine Einführung in das älteste Deutsch. Darstellung und Grammatik*. 3. durchgesehene und wesentlich erweiterte Auflage. Berlin, New York: Walter de Gruyter.