User Manual

Volume 3

Data Documentation

Sabine Lehmann, Dominique Estival, Kirsten Falkedal,
Herv Compagnion, Lorna Balkan, Frederik Fouvry,
Judith Baur, Judith Klein

Istituto Dalle Molle per gli Studii Semantici e Cognitivi

Preface

The TSNLP User Manual summarizes the results of the multi-national project TSNLP (*Test Suites for Natural Language Processing*) funded by the European Commission (DG13) as part of the Language Engineering programme (as research grant LRE-62-089). TSNLP had a duration of 29 months and started in December 1993.

The project produced a methodology and tools for test suite construction in addition to substantial test suites in three languages — English, French, and German (over 4500 distinct items in each language). The test material covers central syntactic phenomena in the three languages and aims to be relevant to any developer or user of systems with grammar components who wish to test, benchmark, or evaluate those systems. To allow for flexible access to the test suites they are mounted on a database.

The project consortium consisted of the following centres and principal researchers:[1]

- University of Essex, UK (coordinator):
  Doug Arnold, Lorna Balkan, Frederik Fouvry
- Istituto Dalle Molle per gli Studii Semantici e Cognitivi (ISSCO), Switzerland:
  Dominique Estival, Kirsten Falkedal, Sabine Lehmann, Hervé Compagnion
- Aerospatiale, France:
  Eva Dauphin, Veronika Lux, Sylvie Regnier-Prost
- Deutsches Forschungszentrum für Künstliche Intelligenz GmbH, Germany:
  Klaus Netter, Stephan Oepen, Hans Uszkoreit

Other contributors included Evangelia Kordoni, Siety Meijer, and Martin Rondell (University of Essex) and Judith Baur, Tom Fettig, Judith Klein, and Fred Oberhauser (DFKI).

The User Manual is in three volumes:

- Volume 1: Background, Methodology, Customisation and Testing
  Lorna Balkan, Frederik Fouvry, Sylvie Regnier-Prost (editors);

[1]Full postal addresses and phone numbers for contact persons of the consortium members are:

CL|MT Group	ISSCO	Aerospatiale France	DFKI GmbH
University of Essex	Université de Genève	Common Research Center	CL Department
Wivenhoe Park	54, route des Acacias	12, rue Pasteur BP 76	Stuhlsatzenhausweg 3
UK Colchester CO4 3SQ	CH 1227 Genève	F 92152 Suresnes Cedex	D 66123 Saarbrücken
+44 - 1206 - 87 20 86	+41 - 22 - 705 7870	+33 - 1 - 4697 3061	+49 - 681 - 302 52 82

- Volume 2: Core Test Suite Technology
  Stephan Oepen, Frederik Fouvry, Klaus Netter, Tom Fettig, Fred Oberhauser, John Gregory;
- Volume 2b: Test Suite Technology
  Frederik Fouvry (editor);
- Volume 3: Test Data Documentation
  Sabine Lehmann, Dominique Estival, Kirsten Falkedal, Hervé Compagnion, Lorna Balkan, Frederik Fouvry, Judith Baur, Judith Klein.

Volume 1 contains a background chapter in which some of the factors which have influenced the design of the project are sketched. The methodology chapter gives a step by step account of how one can go about writing core data, that is, data that cover central phenomena of a language and that are intended to be applicable to a wide range of applications. The customisation chapter describes how the core data can be customized to a particular application (and sketches how it could be customized to a particular domain or text type). The chapter on testing gives an example of how the test suite can be applied to a real life evaluation scenario.

Volume 2 contains a description of the annotation scheme on which the data was constructed, the construction tool tsct used to create the data, the database tsdb on which the data is mounted, and the automated import and consistency checking procedure from tsct to tsdb.

Volume 2b contains a description of and documentation on the automatic test suite generation tool (AutoTSG) and the lexical replacement tool.

Volume 3 contains the detailed documentation that accompanies the data, and which is intended to make the data more accessible to users. It also contains the category and function labels used in the English, French, and German test data with examples for each language, in addition to the vocabulary list used for the test data by the three languages.

TSNLP reports and this User Manual are available from the coordinator at the following address:

Lorna Balkan
CL|MT Group
Department of Language and Linguistics
University of Essex
Wivenhoe Park
UK Colchester, CO4 3SQ
Fax: +44 1206 872085
email: `balka@essex.ac.uk`

Additionally, all TSNLP deliverables, data, and software can be accessed through the internet from the following address:

`http://tsnlp.dfki.uni-sb.de/tsnlp/`

or via the ftp site:

`anonymous@tsnlp.dfki.uni-sb.de/tsnlp/`

Contents

Overview

This volume of the TSNLP user manual describes the TSNLP test data available through the TSNLP database. The documentation explains how the various phenomena have been treated in the three languages English, French and German. Not all the phenomena have been covered to the same extent in all languages and in some cases, they haven't been covered at all in one or more language.

The document provides an overview of the number of existing test sentences for the different phenomena. Furthermore, it includes two annexes which present:

- the category and function labels used in the English, French and German test data with examples for each language.
- a vocabulary list used for the test data by the three languages.

The complete set of data can be accessed through the TSNLP home page

`http://tsnlp.dfki.uni-sb.de/tsnlp/`

maintained at DFKI.

The following table summarizes the total number of test items per phenomenon for each of the three languages.

Phenomenon	English		French		German	
Sentence Types	**162**	(64/98)	**776**	(389/387)	**121**	(107/14)
Questions	180	(64/98)	428	(247/181)		
Imperatives			309	(126/183)		
Conditionals			39	(16/23)		
Diathesis/Passive	**281**	(124/157)	**295**	(176/119)	**295**	(147/148)
Tense-Aspect-Modality	**196**	(157/39)	**352**	(77/275)	**322**	(188/134)
Complementation	**1048**	(158/890)	**827**	(209/618)	**464**	(218/246)
C_Complementation	1011	(148/863)	755	(188/567)	464	(218/246)
NP_Complementation	37	(10/27)	40	(12/28)		
AP_Complementation			32	(9/23)		
Modification	**782**	(300/482)	**392**	(329/63)	**198**	(131/67)
C_Modification			392	(329/63)		
NP_Modification	782	(300/482)			198	(131/67)
Agreement	**1319**	(269/1050)	**1641**	(376/1265)	**2430**	(523/1907)
C_Agreement	123	(68/55)	287	(104/183)	399	(224/175)
NP_Agreement	1196	(201/995)	1354	(272/1082)	2029	(299/1732)
Word-Order			**14**	(7/7)	**329**	(103/226)
NP_Word-Order					329	(103/226)
Coordination	**333**	(147/186)	**698**	(379/319)	**534**	(105/429)
C_Coordination			165	(105/60)	534	(105/429)
NP_Coordination	333	(147/186)	371	(190/181)		
PP_Coordination			162	(84/78)		
Negation	**418**	(289/129)	**168**	(68/100)	**255**	(110/145)
S_Negation					172	(72/100)
C_Negation	418	(289/129)			58	(24/34)
NP_Negation					25	(14/11)
Parentheticals	**40**	(12/28)				
Abbreviations	**18**	(12/6)				
TOTAL	**4597** (1532/3065)		**5163** (2010/3153)		**4948** (1632/3316)	

Figure 0.1: Status of the data construction end of December 1995: relevance and breadth of individual phenomena differ for the three languages. (The figures are total numbers, followed grammatical and ungrammatical (including dubious and relatively ill-formed items)

Chapter 1

Sentence Types

1.1 English

1.1.1 Questions

The English data only contain test items for (positive) questions as main clauses. However, nearly all of the types occurring as main clause can occur as a corresponding subordinate clause. So, though the actual data will differ, more or less the same classification can be used for subordinate clauses. The same applies to the interaction with negation (negated questions).

A general property of questions is that they end with a question mark.

General Division of the Phenomenon

The following types of questions are covered in the data:

- S_Types-Questions-y/n_questions
  A y/n question can be answered with yes or no:
  Did she come ?
- S_Types-Questions-Wh_questions
  A Wh-question is a question that begins with a *wh*-word:
  When does he come ?

A third type of question, choice questions (*Did she come or didn't she ?*) cannot be answered by yes or no, but otherwise the same syntactic structure as y/n-questions, so could be treated under the same name.

The following closed class of *wh*-words was expanded: *who*, *which*, *what*, *whom*, *how*, *why*, *where* and *when*. These however are not expanded everywhere. This expansion was generated semi-automatically, using the Auto_TSG (WP5.1). Most of the English auxiliaries are included (the central auxiliaries are in any case).

One general presupposition that has not been filled in in the phenomenon file, states that the constituents that are not analysed further, are grammatically correct. The other presuppositions are that Agreement, Complementation, Diathesis and Word Order are correct. Word order and Diathesis are assumed as default presuppositions.

The standard word order assumed is:

subject verb object(s)

or

verb subject object(s)

as appropriate.

Ill-formed input only contains one error (relative to the positive item in the set). They were generated in the following way:

1. changing the punctuation ("?" or ".")
2. inverting subject and verb, or omission of the inversion.

Test sets are made to be self-contained. This means that in each test set *all* the possible variations of a well formed item are enumerated, regardless of whether this repeats an item or not.

The relatively ill-formed test items (WF = 3) were omitted in case they were always predictably ill-formed (like replacing a question mark by a full stop in a yes/no question without inversion, this always gives a relatively ill-formed item w.r.t. questions).

S_Types-Questions-Y/N_questions-Inverted

Yes/no-questions are inverted.

Has he carried it ?

S_Types-Questions-Y/N_questions-Non_inverted-Non_tagged

Y/n questions can always be asked with the word order of a declarative sentence. Tag questions can be attached, but this is not tested here (there is an interaction between the form of the verb in the actual question and the verb in the tag.)

She has carried it ?

S_Types-Questions-Wh_Subj(PRON_Wh)

The *wh*-word (a pronoun) has the function of subject.

Who comes ?

S_Types-Questions-Wh_Subj(DET_Wh+NP)

The *wh*-word is a determiner of the subject.

Which manager comes ?

S_Types-Questions-Wh_Obj(PRON_Wh)
The *wh*-word has the function of direct object.
What manager does he see ?

S_Types-Questions-Wh_Obj(DET_Wh+NP)
The *wh*-word is a determiner of the direct object.
What manager does she see ?

S_Types-Questions-Wh_Iobj(PRON_Wh)
The *wh*-word is a pronoun, and functions as indirect object.
Who does he tell ?

S_Types-Questions-Wh_Pobj(PRON_Wh)
The *wh*-word has the function of prepositional object.
Whom does she give the proposal to ?

S_Types-Questions-Wh_Pobj(DET_wh+NP)
The *wh*-word is a determiner of the prepositional object.
Which manager does he give the proposal to ?
The following cases are analogous to the previous two, but without stranding.

S_Types-Questions-Wh_Pobj(Prep+PRON_wh)
To whom does she give the proposal ?

S_Types-Questions-Wh_Pobj(Prep+DET_wh+NP)
To which manager does he give the proposal ?

S_Types-Questions-Wh_Complementizer_adv
The *wh* word is an adverb.
When does she come ?

S_Types-Questions-Wh_Mod(DET_wh+ADV)
The *wh* word is the specifier of an adverb.
How late does she come ?

S_Types-Questions-Wh_Pred(ADV_wh)
The *wh* word is the predicate (ADVP) of the subject.
How is he ?

S_Types-Questions-Wh_Pred(PRON_wh)

The *wh* word is the predicate (NP) of the subject.

Who is he ?

S_Types-Questions-Wh_Pred(DET_wh+ADJ)

The *wh* word is a determiner of the predicate (AP) of the subject.

How ill is she ?

S_Types-Questions-Wh_Pred(DET_wh+NP)

The *wh* word is the predicate (AP) of the subject.

What manager is he ?

Data

S_Types-Questions	
162	test items
64	well-formed items
98	ill-formed items
0	doubious items
0	relatively ill-formed items

1.2 French

1.2.1 Questions

According to Grevisse (1991) [Grevisse, 1991] (§381-391), questions can be classified as:

- Direct questions
- Indirect questions
- Global questions
- Partial questions

Global questions are also called Yes/No-questions, whereas partial questions correspond to WH-questions, i.e. questions introduced by a question word.

In the written language, questions can be further described as presenting the following characteristics:

- punctuation: presence of the question mark ("?");
- word order: inversion of the subject and the verb;
- duplication of the subject;
- possible presence of the interrogative particle *est-ce que*.

General Division of the Phenomenon

The division of questions we have adopted for the TSNLP test data is partly based on the approach given in Grevisse (1991), but is more fine-grained. We only give here an overview of this phenomenon and describe in more detail the different subphenomena in the next sections.

So far, the TSNLP test data does not include indirect questions, and all test items are instances of direct questions. Thus, the basic division of the data for questions is that between **Yes/No-Questions** and **WH-Questions**.

Furthermore, since the presence of a question mark is relevant for both Yes/No-Questions and WH-Questions, this first division also includes Punctuation.

- S_Question_Punctuation
- S_Question_y-n
- S_Question_wh

Further subdivisions concern the other characteristics of questions, and the TSNLP test data is classified according to whether the test item is an inverted question or whether it includes the particle *est-ce que*.[1] However, some questions do not fall in either category: for instance, WH-Questions whose subject is the WH-word itself or WH-questions which do not contain an inflected verb. Thus we have defined the following classification:

[1] The TSNLP category of this particle is PART_wh(est-ce-que).

- S_Question_y-n:
  - S_Question_y-n_inverted
  - S_Question_y-n_est-ce-que
- S_Question_wh:
  - S_Question_wh_inverted
  - S_Question_wh_est-ce-que
  - S_Question_wh_subj
  - S_Question_wh_infinitive
  - S_Question_wh_verbless

S_Question_Punctuation

As Punctuation in questions interacts with the general phenomenon of Punctuation, the supertypes of S_Question_Punctuation are both S_Question and S_Punctuation. The test data checks for the presence of the question mark:

Viens-tu ?
** Viens-tu*

The data for S_Question_Punctuation includes the four basic types of questions, namely Yes/No-Questions and WH-Questions, both with an inverted subject and with the interrogative particle *est-ce que*.

S_Question_y-n

Yes/No-Questions are questions where no particular element of the sentence is questioned, but which can be answered by *Yes* or *No*. In French, Yes/No-Questions can either present an inversion of the subject, or they can start with the interrogative particle *est-ce que*.

S_Question_y-n_inverted

A question in which the subject is not inverted and which does not contain the interrogative particle *est-ce que* is incorrect. But it is rather difficult to test for inversion, i.e. that the subject should follow the verb instead of preceding it, because this only leads to **relatively ill-formed** test items:

Viens-tu ?
( ) Tu viens ?*

The first problem with an item such as *Tu viens ?* is that it could be a well-formed declarative sentence, and in that case the error would be the presence of the question mark. A second problem is that, although this sentence does not present an inversion of the subject, it cannot be considered as totally ill-formed, because it is quite common in the spoken language. Thus, the TSNLP test data does not contain this kind of test items.

However, the inversion of the subject presents some characteristics which can be tested:

- there must be a hyphen between the subject and the verb;
- the subject must follow the inverted verb;

- if the subject starts with a vowel, a euphonic *-t* must be present;
- a 1st person singular pronoun may behave differently from other subjects:
  – not all the verbs can have an inverted 1st person singular subject;
  – the spelling may change;
- the inverted subject must be a pronoun;[2]
- the subject may be duplicated: since a full NP subject cannot be inverted, it must be repeated as a pronoun which must agree with the NP.

S_Question_y-n_inverted-PRON_Punctuation

There must be a hyphen between the inflected verb and the subject:

Viens-tu ?
** Viens tu ?*

S_Question_y-n_inverted-PRON_Wordorder

In inverted Yes/No-Questions, the subject must follow the inflected verb.[3] To test this aspect of inverted questions, the verb in the test item is used in the *passé composé*, so that the ill-formed test items can also test whether the subject follows the past participle instead of the auxiliary verb:

Suis-je venu ?
** Suis venu-je ?*

S_Question_y-n_inverted-PRON_euphonic-t

If the (subject) pronoun starts with a vowel, there must be a euphonic *-t* between the verb and the inverted subject:[4]

Viendra-t-il ?
** Viendra-t-elle ?*

S_Question_y-n_inverted-PRON-1-sg

All verbs do not accept a first person singular pronoun as inverted subject. Although this cannot be tested in an exhaustive manner (since it is a lexical matter), the most important verbs of this class are included in the test data.

Dois-je venir ?
** Sors-je ?*

---

[2]Since inverted questions involve pronouns, and these are closed class items, pronouns are exhaustively tested.

[3]As this aspect of questions interacts with the phenomenon of Word Order, the supertypes of S_Question_y-n_inverted-PRON_Wordorder are both S_Question and S_Wordorder.

[4]This concerns all 3rd person subject pronouns: *il, elle, ils* and *elles*, which are exhaustively tested. The category of the euphonic *-t* is PART_euph.

S_Question_y-n_inverted-PRON-1-sg(accent)

If the verb ends with an *-e*, it must be changed to *é* (with an accute accent).

Ouvré-je ?
** Ouvre-je ?*

S_Question_y-n_inverted-PRON_subj(NP)

If the subject is a full NP, it must be duplicated with an inverted pronoun:

Jean vient-il ?
** Vient-Jean ?*

S_Question_y-n_inverted-PRON_subj(NP)_Agreement

If the subject is a full NP, the inverted pronoun must agree with it.[5]

S_Question_y-n_est-ce-que

With the interrogative particle *est-ce que*, there is no inversion of the subject. Therefore it is difficult to test for the presence of *est-ce que*, since its deletion from a well-formed question results in a well-formed declarative sentence.

Est-ce qu'il vient ?
** Il vient ?*

The only aspect of these questions that the TSNLP data can then test for is the position of the subject: since the subject should be in the same position as in a declarative sentence, an inverted subject is added:[6]

Est-ce qu'il vient ?
** Est-ce que vient-il ?*

S_Question_wh

WH-Questions start with a question word (sometimes called WH-word) which refers to the element being questioned. As with Yes/No-Questions, a WH-question may either have subject-verb inversion (except when the element questioned is the subject, e.g. *Qui vient ?*) or it can start with the interrogative particle *est-ce que*.

S_Question_wh_inverted

The main characteristics of inverted WH-Questions (i.e. not starting with *est-ce que*) which can be tested with the TSNLP test data are: the inversion of the subject, its agreement and punctuation (presence or absence of a hyphen with the inverted subject).

For WH-Questions in which the questioned element is not the subject, two constructions are possible:[7]

---

[5] The supertypes of S_Question_y-n_inverted-PRON_subj(NP)_Agreement are S_Question and C_Agreement.

[6] The supertypes of S_Question_y-n_est-ce-que are thus both S_Question and S_Wordorder.

[7] Although we do not include interrogative adverbs such as *quand*, *comment* etc., in the closed class items, we systematically test for the most important of them; their category is ADV_wh.

- if the subject is a pronoun, it must be inverted with the verb:
  Comment va-t-il ?
- if the subject is a full NP, either:
  – the full NP must be inverted with the verb;
  Comment va Jean ?

  – or it is duplicated by an inverted pronoun:
  Comment Jean va-t-il ?

S_Question_wh_inverted_Agreement

The full NP subject and the duplicated pronoun must agree:[8]

Pourquoi Jean vient-il ?
** Pourquoi Jean vient-elle ?*

S_Question_wh_inverted_Punctuation

While a hyphen is necessary with a inverted subject pronoun, there must be no hyphen with a inverted subject NP. Thus, the TSNLP data tests for the absence of the hyphen with a full NP.[9]

** Comment va-Jean ?*

S_Question_wh_inverted_PRON_wh-attribute_Agreement

If the interrogative word is an attributive pronoun, there must be agreement between this pronoun and the subject.[10]

Quel est cet homme ?
** Quelle est cet homme ?*

S_Question_wh_inverted_obj(PRON_wh)

The form of the interrogative object pronoun depends on its semantics: *qui* can only be human and *que* non-human.

Que fait cet homme ?
** Qui fait cet homme ?*

Although TSNLP is not meant to test for semantics, since this feature is a very important characteristic of questions and a possible source of error, it has been included in the TSNLP test data.

S_Question_wh_inverted_p-obj(PRON-wh-det)_Wordorder

---

[8]This phenomenon has two supertypes: S_Question and C_Agreement.
[9]Since this aspect of questions interacts with Punctuation, its supertypes are both S_Question and S_Punctuation.
[10]Thus, this phenomenon has two supertypes: S_Question and C_Agreement.

Inversion of the subject is tested when the question starts with an interrogative pronoun such as *duquel* or *auquel*, i.e. a form which is a contraction of the preposition *à* or *de* and an interrogative pronoun.[11]

Duquel parle-t-il ?
** Duquel il parle ?*[12]

As these interrogative pronouns are closed class items, they are systematically tested. The data also include similar test items for non-contracted prepositional interrogatives with the preposition *pour*.

S_Question_wh_inverted_p-obj(DET_wh+NP)_Wordorder

Inversion of the subject is tested when the interrogative form consists of a proposition followed by an interrogative determiner and a full NP.[13]

De quel livre parle-t-il ?
** De quel livre il parle ?*[14]

S_Question_wh_inverted_p-obj(PRON_wh)_form

The form of the interrogative pronoun in prepositional objects depends on its semantics: while the animate is *qui*, the non-animate must be the tonic form *quoi* and not the non-tonic form *que*.

De qui parle-t-il ?
De quoi parle-t-il ?
** De que parle-t-il ?*

S_Question_wh_inverted_p-obj(PRON_wh)_Wordorder

Inversion of the subject is tested when the interrogative element is a prepositional object.[15]

De quoi parle-t-il ?
** De quoi il parle ?*

S_Question_wh_inverted_pourquoi

In contrast to sentences with the other interrogative adverbs, the NP subject cannot be inverted with *pourquoi*, but must be duplicated as a pronoun.[16]

Pourquoi Jean vient-il ?
** Pourquoi vient Jean ?*

[11]The supertypes of this phenomenon are S_Question and S_Wordorder.
[12]The fact that this sentence would be accepted in spoken, familiar French is mentioned in the comments on the test input. In this case, the item is considered to be well-formed and the value of the attribute REGISTER is “informal”.
[13]The supertypes of this phenomenon are S_Question and S_Wordorder.
[14]The fact that this sentence would be accepted in spoken, familiar French is mentioned in the comments on the test input. In this case, the item is considered to be well-formed and the value of the attribute REGISTER is “informal”.
[15]The supertypes of this phenomenon are S_Question and S_Wordorder.
[16]The supertypes of this phenomenon are S_Question and S_Wordorder.

S_Question_wh_est-ce-que

WH-Questions using the interrogative particle *est-ce que* present the normal subject-verb word order. Therefore, it is easier to test for the presence of the interrogative particle *est-ce que* in WH-questions than in Yes/No-Questions (see above), because when it is deleted, the resulting sentence is ill-formed:

Comment est-ce que Jean va ?
** Comment Jean va ?*[17]

S_Question_wh_est-ce-que_subj(PRON_wh)

When the subject is the questioned element, both the case and the "animacy" of the interrogative pronoun are checked:

– *est-ce qui* is nominative whereas *est-ce que* is accusative;
– *Qui* is human whereas *Que/Qu'* is non-human.

Qui est-ce qui vient ?
** Qui est-ce que vient ?*

Qu'est-ce qui se passe ?
** Qui est-ce qui se passe ?*

Although TSNLP is not meant to test for semantics, since this feature is a very important characteristic of questions and a possible source of error, it has been included in the TSNLP test data.

S_Question_wh_est-ce-que_p-obj(PRON_wh-det)

The presence of the interrogative particle *est-ce que* is tested in WH-Questions starting with a contracted prepositional interrogative pronoun, such as *duquel* or *auquel*.

Duquel est-ce qu'il parle ?
** Duquel il parle ?*

S_Question_wh_est-ce-que_p-obj(DET_wh+NP)

The presence of the interrogative particle *est-ce que* is tested in WH-Questions with an interrogative determiner.

De quel livre est-ce qu'il parle ?
** De quel livre il parle?*

S_Question_wh_est-ce que_p-obj(PRON_wh)

The presence of the interrogative particle *est-ce que* is tested in WH-Questions with an interrogative prepositional object.

[17]The fact that this sentence would be accepted in spoken, familiar French is mentioned in the comments on the test input. In this case, the item is considered to be well-formed and the value of the attribute REGISTER is "informal".

De quoi est-ce qu'il parle ?
** De quoi il parle ?*

S_Question_wh_subj

Without the interrogative particle *est-ce que*, WH-Questions in which the subject is the questioned element behave differently than other WH-Questions with respect to word order, since the subject is in the normal position, before the verb, and there is no inversion or duplication of the subject.

The TSNLP test data is divided according to the types of subject.

In contrast with the interrogative subject pronoun, which can be used in conjunction with the interrogative particle *est-ce que*, complex interrogative subject forms do not allow a variant with *est-ce que*.[18]

S_Question_wh_subj(PRON_wh)

The subject interrogative pronoun is semantically constrained, because only the human pronoun *qui* is allowed as subject.[19]

Qui vient ?
** Que vient ?*

Although TSNLP is not meant to test for semantics, since this feature is a very important characteristic of questions and a possible source of error, it has been included in the TSNLP test data.

S_Question_wh_subj(DET_wh+NP)

Subject questions with an interrogative determiner followed by an NP do not allow a variant with *est-ce que*, neither can there be subject inversion.

Quel homme vient ?
** Quel homme vient-il ?*
** Quel homme est-ce que vient ?*

S_Question_wh_subj(PRON_wh-det)

Subject questions with an interrogative determiner do not allow a variant with *est-ce que*, neither can there be subject inversion.

Lequel vient ?
** Lequel vient-il ?*
** Lequel est-ce que vient ?*

---

[18] See S_Question_wh_est-ce-que_subj(PRON_wh), above.

[19] Non-human subjects must be questioned with the interrogative particle *est-ce que*:
Qu'est-ce qui s'est passé ?
See S_Question_wh_est-ce-que_subj(PRON_wh), above.

S_Question_wh_infinitive

WH-Questions may also be in the infinitive. Infinitive questions cannot contain the interrogative particle *est-ce que* and the interrogative pronoun cannot be in the strong form (tonic) but has to in the weak form (non-tonic).

As with the other types of questions, the TSNLP test data for infinitive questions also checks for the semantic feature of animacy. Thus the following three characteristics are tested for:

Que faire ?
* *Quoi faire ?* (form of the pronoun)
* *Qui faire ?* ("animacy" of the pronoun)
* *Qu'est-ce que faire ?* (absence of *est-ce que*)

S_Question_wh_verbless

Some questions may not contain any verb at all. These questions cannot contain the interrogative particle *est-ce que* either, but unlike in infinitive questions, the interrogative pronoun must be in the strong form (tonic). Thus, the same three characteristics are checked in the TSNLP test data:

Quoi de neuf ?
* *Que de neuf ?* (form of the pronoun)
* *Qui de neuf ?* ("animacy" of the pronoun)
* *Quoi est-ce que de neuf ?* (absence of *est-ce que*)

Parameters

WF	Attribute	Value	input
S_Question_Punctuation			
1	punct	PUNCT(?)	Viens-tu ?
0	punct	PUNCT(?)	* Viens-tu
S_Question_y-n_inverted-PRON_Punctuation			
1	punct	PUNCT(-)	Viens-tu?
0	punct	PUNCT(-)	* Viens tu?

S_Question_y-n_inverted-PRON_Wordorder			
1	position_subj(PRON)	post_aux	Es-tu venu?
0	position_subj(PRON)	post_aux	* Es venu-tu?
S_Question_y-n_inverted-PRON_euphonic-t			
1	euphonic-t	present	Mange-t-il?
1	euphonic-t	absent	Vient-il?
0	euphonic-t	present	* Mange-il?
0	euphonic-t	absent	* Vient-t-il?
S_Question_y-n_inverted-PRON-1-sg			
1	position_subj(PRON_pers-1-sg)	post_verb	ai-je
1	orthography	accent_aigu	ouvré-je?
0	orthography	accent_aigu	* ouvre-je?
S_Question_y-n_inverted-PRON_subj(NP)			
1	inverted_subj	PRON	Jean vient-il?
0	inverted_subj	PRON	* Vient-Jean?
S_Question_y-n_inverted-PRON_subj(NP)_Agreement			
1	num-gen	sg-masc	Jean vient-il?
0	num-gen	sg-masc	* Jean vient-elle?

S_Question_y-n_est-ce-que_subj_Wordorder			
1	position_subj	ante_verb	Est-ce qu'il vient?
0	position_subj	ante_verb	Est-ce que vient il?

S_Question_y-n_est-ce-que_subj_Wordorder			
1	position_subj	ante_verb	Est-ce qu'il vient?
0	position_subj	ante_verb	Est-ce que vient il?

S_Question_wh_subj(DET_wh+NP)			
1	position_subj	ante_verb	Quel homme vient?
0	position_subj	ante_verb	* Quel homme vient-il?
0	position_subj	ante_verb	* Vient-il quel homme?
1	wh_particle(est-ce que)	absent	Quel homme vient?
0	wh_particle(est-ce que)	absent	* Quel homme est-ce que vient?
S_Question_wh_subj(PRON_wh)			
1	animacy(PRON_wh)	human	Qui vient?
0	animacy(PRON_wh)	human	* Quoi vient?
S_Question_wh_subj(PRON_wh-det)			
1	position_subj	ante_verb	Lequel vient?
0	position_subj	ante_verb	* Lequel vient il?
0	position_subj	ante_verb	* Vient lequel ?
1	wh_particle(est-ce que)	absent	Lequel vient?
0	wh_particle(est-ce que)	absent	* Lequel est-ce que vient?

S_Question_wh_inverted_Punctuation			
1	punct(-)	absent	Comment va Jean?
0	punct(-)	absent	Comment va-Jean?
S_Question_wh_inverted_Agreement			
1	num-gen	sg-masc	Pourquoi Jean vient-il?
0	num-gen	sg-masc	* Pourquoi Jean vient-elle?
S_Question_wh_inverted_PRON_wh-attribute_Agreement			
1	num-gen	sg-masc	Quel est cet homme?
0	num-gen	sg-masc	* Quelle est cet homme?
S_Question_wh_inverted_pourquoi			
1	inverted_subj	PRON	Pourquoi Jean vient-il?
0	inverted_subj	PRON	* Pourquoi vient-Jean?

S_Question_wh_inverted_obj(PRON_wh)			
1	case	acc	Que fait cet homme?
0	case	acc	* Qui fait cet homme?
1	form(PRON_wh)	weak	Que fait cet homme?
0	form(PRON_wh)	weak	* Quoi fait cet homme
S_Question_wh_inverted_p-obj(PRON_wh-det)_Wordorder			
1	position_subj	post_verb	Duquel parle-t-il?
0	position_subj	post_verb	* Duquel il parle?
S_Question_wh_inverted_p-obj(DET_wh+NP)_Wordorder			
1	position_subj	post_verb	De quel livre parle-t-il?
0	position_subj	post_verb	* De quel livre il parle?
S_Question_wh_inverted_p-obj(PRON_wh)_form			
1	form(PRON_wh)	strong	De quoi parle-t-il?
0	form(PRON_wh)	strong	* De que parle-t-il?
S_Question_wh_inverted_p-obj(PRON_wh)_Wordorder			
1	position_subj	post_verb	De quoi parle-t-il?
0	position_subj	post_verb	* De quoi il parle?

S_Question_wh_est-ce-que			
1	wh_particle	est-ce que	Comment est-ce que Jean va?
0	wh_particle	est-ce que	* Comment Jean va?
1	position_subj	ante_verb	Comment est-ce qu'il vient?
0	position_subj	ante_verb	* Comment est-ce que vient-il?
S_Question_wh_est-ce-que_subj(PRON_wh)			
1	case(PART_wh)	nom	Qui est-ce qui vient?
0	case(PART_wh)	nom	* Qui est-ce que vient?
1	animacy(PRON_wh)	non-human	Qu'est-ce qui se passe?
0	animacy(PRON_wh)	non-human	* Qui est-ce qui se passe?
1	animacy(PRON_wh)	human	Qui est-ce qui pleure?
0	animacy(PRON_wh)	human	* Qu'est-ce qui pleure?
S_Question_wh_est-ce-que_p-obj(PRON_wh-det)			
1	wh_particle	est-ce que	Duquel est-ce qu'il parle?
0	wh_particle	est-ce que	* Duquel il parle?
S_Question_wh_est-ce-que_p-obj(DET_wh+NP)			
1	wh_particle	est-ce que	De quel livre est-ce qu'il parle?
0	wh_particle	est-ce que	* De quel livre il parle?
S_Question_wh_est-ce-que_p-obj(PRON_wh)			
1	wh_particle	est-ce que	A qui est-ce qu'il parle?
0	wh_particle	est-ce que	* A qui il parle?

S_Question_wh_infinitive			
1	animacy(PRON_wh)	non-human	Que faire?
0	animacy(PRON_wh)	non-human	Qui faire?
1	wh_particle(est-ce que)	absent	Qui rencontrer?
0	wh_particle(est-ce que)	absent	Que rencontrer?
S_Question_wh_verbless			
1	form(PRON_wh)	strong	Quoi de neuf?
0	form(PRON_wh)	strong	* Que de neuf?
1	animacy(PRON_wh)	non-human	Quoi de neuf?
0	animacy(PRON_wh)	non-human	* Qui de neuf?
1	wh_particle(est-ce que)	absent	Quoi de neuf?
0	wh_particle(est-ce que)	absent	Quoi est-ce que de neuf?

Data

The total number of test items for S_Question is:

S_Question	
428	test items
247	well-formed items
126	ill-formed items

1.2.2 Imperatives

The imperative is one of the three sentence types (declarative, question and imperative) which characterize the **discourse** function of a sentence, (in this case, that of a command or a request).

In addition to Grevisse (1991) [Grevisse, 1991], which is the standard reference for French grammar we have used in building the French test data for TSNLP, we have also consulted McCawley (1988) [McCawley, 1988], for a more specifically syntactic characterization of the imperative construction.

There is only one domain for the imperative construction, i.e. the whole sentence, and we name the phenomenon S_Imperative.

Although traditionally, a declarative sentence ends with a full stop ("."), a question with a question mark ("?") and an imperative with an exclamation mark ("!"), since the exclamation mark is often ommitted with imperatives, we do not take it as a necessary feature of this construction and do not provide test items for it in the TSNLP French data.[20]

Grevisse (§399-400) distinguishes three main types of formally distinct imperatives, in which he also counts "optatives":

a) Subjectless imperatives: subjectless sentences where the possible forms of the finite verb are limited to 2.sg, 1.pl and 2.pl.

b) Subjunctive clauses introduced by *que*. Possible subjects are limited to 2nd person singular and plural.

c) Infinitive clauses (without subject)

A fourth class, **averbal phrases**, is described as limited to oral communication.

For each of the three main types, a number of variations and exceptions are listed.

The TSNLP data in S_Imperative covers the three types of imperatives, with most of the subclasses and exceptions listed in Grevisse. Most, but not all, subclasses cover the allowed person/number combinations extensively, others are only covered for second person.

Although, strictly speaking, this aspect of imperatives belongs to morphology, and thus is outside the scope of TSNLP, imperative forms of some highly frequent verbs like *être*, *avoir* and *savoir* have been included.

The presuppositions needed for S_Imperative are:

- C_Wordorder
- C_Agreement
- C_Complementation

General Division of the Phenomenon

As default values, the sentences are positive (i.e. not negated), and the tense is "Présent". Following Grevisse, we subdivide imperative constructions according to whether:

- the main verb is in the imperative, with no expressed syntactic subject (default value): S_Imperative;

[20] Compare with the test data for questions, where the presence of a question mark is taken as necessary.

- the main verb is in the imperative, with a vocative NP: S_Imperative_vocative;
- the main verb is in the subjunctive, with an expressed syntactic subject: S_Imperative_subjunctive;
- the main verb is in the infinitive, with no expressed syntactic subject: S_Imperative_infinitive.

For the "pure" imperatives with an imperative verb form, we further consider the following aspects:

- form of the verb
- past tense
- type of subject
- realization of pronominal objects

The last aspect is also considered for the infinite imperative.

The TSNLP data for imperatives also takes into account the interaction with **Negation**, and a similar division as that given above is used to treat negative imperatives:

- S_Imperative_Negation
- S_Imperative_Negation_infinitive

S_Imperative

On a pragmatic level, other forms than the imperative form of the verb can be used in commands, but morphologically, the imperative only concerns three persons of the verb: the 2nd person singular and the 1st and 2nd person plural. The test data check for this restriction, as well as for the absence of an expressed subject.[21]

Pars !
Partons !
Partez !
** Part !*

S_Imperative_past

Usually the imperative is in the present tense (which we take here as the default value), but it is possible to use it in the past tense. It then refers to an action which must be accomplished at some moment in the future (which may be specified by an adjunct).

[21] For the latter characteristic, an obvious way to derive an ill-formed test item is to introduce a subject. However, in most cases this creates a well-formed declarative sentence, which can only be considered a **relatively ill-formed** test item (WF=3), i.e. ill-formed w.r.t. the imperative construction.
Partez !
** Vous partez !* (WF=3)
The French TSNLP test data does not include such test items.

Here the test data only checks that the correct forms are accepted and does not include ill-formed test items. The only parameter is the tense of the verb (past). This assumes that the choice of the correct auxiliary (*avoir* or *être* has already been checked (in C_Tense-aspect-modality)).

Soyez partis demain !
Ayez fini ce travail !

S_Imperative_form

For the actual verbal form of the imperative, we are more concerned with spelling than with syntax.

1. Irregular imperatives:

   Most French verbs use the indicative present for their imperative form, but there are three verbs (*avoir*, *être* and *savoir*) whose imperative is not formed in a regular way.

   Sois gentil !   ** Es gentil !*

   These are checked exhaustively in the TSNLP data.

2. -s in 2nd pers.sg.:

   The form of the imperative of the 2nd person singular of verbs ending in *-er* does not exactly correspond to that of the present indicative (no *-s*).

   Chante !
   ** Chantes !*

   This is complicated by the fact that the *-s* must be present when the verb is followed by a clitic complement starting with a vowel.

   Vas-y !
   ** Va-y !*
   Parles-en !
   ** Parle-en !*

Except in the last case, where the presence of an object is crucial, only intransitive verbs are used in the test data for the imperative form of the verb.

S_Imperative_PRON

When the objects of the imperative verb are pronominals, they do not obey the same rules as in declarative sentences for their form and position. The test data for pronominal objects does not check for all the verbal imperative forms, but take as default the 2nd person singular.

S_Imperative_PRON_form

In a positive imperative, the pronouns must be in the strong (or tonic) form (they are in the unstressed clitic form in the negative imperative, see below).

Dis-moi !
** Dis-me !*

Since pronouns form a closed class, the test data check for the form of the pronouns in imperatives in an exhaustive way.

S_Imperative_PRON_Wordorder

In a positive imperative, the pronominal objects follow the verb instead of preceding it as in declarative sentences.

Présente-le !
* *Le présente !*

However, the ordering among the pronominals is the same as in declarative sentences.

Parlez-en-lui !
* *Parlez-lui-en !*
* *Lui en parlez !*
* *En lui parlez !*

Since pronouns form a closed class, the test data checks for the position of the pronouns in imperatives in an exhaustive way.[22]

S_Imperative_PRON_Punctuation

The clitic pronouns must be connected to the verb by a hyphen.

Présente-le !
* *Présente le !*

Since pronouns form a closed class, the test data check for the presence of the hyphen in imperatives in an exhaustive way.[23]

S_Imperative_PRON_euphonic-t

When the object pronoun is *en* and the verb ends in a vowel, some verbs require a **euphonic** *-t*.[24]

Here we test not only for the presence of this euphonic *-t*, but also for its punctuation because the euphonic *-t* used in imperatives is not followed by a hyphen (as in questions), but by an apostrophe.

Va-t'en !
* *Va-en !*
* *Va-t-en !*

S_Imperative_vocative

A vocative NP may be used to indicate the intended addressee of the imperative verb. It is then at the beginning of the sentence and must be separated by a comma. Thus, the test data check for the presence of the comma:

[22] The supertypes of S_Imperative_PRON_Wordorder are both S_Imperative and S_Wordorder.
[23] The supertypes of S_Imperative_PRON_Punctuation are both S_Imperative and S_Punctuation.
[24] Most verbs require an *-s*, e.g. *Parles-en !* in 1.2.2 above or *Manges-en !* The verbs which require *-t* have to be stated as such in the lexicon.

Nicolas, pars !
** Nicolas pars !*

Note that this sentence would probably be interpreted by an NLP system as containing an error in subject-verb agreement, and not as an ill-formed imperative.

S_Imperative_subjunctive

S_Imperative_subjunctive-que

Imperative sentences in which the intended addressee is in the 3rd person use the subjunctive form of the verb and are introduced by the complementizer *que*. The test data checks for the mood (subjunctive) and for the person (3rd) of the verb.[25]

Qu'il parte !
** Qu'il part !*
** Que tu partes !*

S_Imperative_subjunctive-bare

A few idiomatic expressions consist of an imperative sentence whose verb is in the subjunctive, but unlike in the subjunctive imperative of 1.2.2, it is not introduced by the complementizer *que*.[26] The test data checks for the mood (subjunctive).

Vive le roi !

Note that if we derive ill-formed test items by systematically replacing the subjunctive with an indicative, the result may only be a **relatively ill-formed item**.[27]

Dieu veuille me pardonner !
** Dieu veut me pardonner !* (WF=3)

S_Imperative_infinitive

Infinitive constructions used as imperatives are mostly found in special kinds of texts such as technical manuals (or recipe books), in which instructions are general and not directed to specific persons.

The test items for imperative infinitives only check that the correct forms are accepted. There are no ill-formed test items and the parameters only mention the form of the verb (infinitive). The same test set may include several positive examples.

Laisser reposer.

S_Imperative_Negation

Negative imperatives involve an interaction between two phenomena:
S_Imperative and C_Negation. We only consider negative imperatives whose verb is in the

[25] An extension of the data would be to add test items to check for the presence of the complementizer.
[26] This kind of imperative is sometimes called **optative**.
[27] There are no such ill-formed test items in the French TSNLP data.

imperative or in the infinitive, because a negative subjunctive imperative behaves exactly like a declarative clause (see C_Negation).[28]

The two negative elements *ne* and *pas* are taken as the default instantiation of the negation, i.e. there are no test items with other negative elements such as *rien*, *personne*, etc.

For the "pure" imperatives, the test data check all three persons.

S_Imperative_Negation_Wordorder

The test data check for the presence of both of the negative elements *ne* and *pas* in the present imperative and for their position in front of the verb.

Ne partez pas !
** Ne pas partez !*
** Ne partez !*
** Partez pas !*

S_Imperative_Negation_past_Wordorder

In the past imperative, the position of the negative element *pas* is after the auxiliary and in front of the main verb.

Ne soyez pas partis demain !
** Ne pas soyez partis demain !*
** Ne soyez partis pas demain !*

The test data do not check for the presence of both of the negative elements *ne* and *pas* in the past, since this is assumed to have been tested for the present imperative.

The test data contain test items for both auxiliaries *avoir* and *être*.

S_Imperative_Negation_PRON

As with the positive form of the imperative, the test data check both the form and the position of pronominal objects in a negative imperative.

S_Imperative_Negation_PRON_form

In a negative imperative, the pronominal cannot be a tonic form (as opposed to the pronominals which must be used in the affirmative, see above).

Ne me présente pas !
** Ne moi présente pas !*

S_Imperative_Negation_PRON_Wordorder

Unlike in the positive form of the imperative, pronominal objects must precede the verb in a negative imperative. The test data also check the position of the negative element *pas* which must follow both the pronoun and the verb.

[28] Most of the test items in S_Imperative_Negation check for the position of the negation. Thus, they also treat Word Order and the phenomenon of S_Imperative_Negation has three supertypes: S_Imperative, C_Negation and S_Wordorder.

Ne le présente pas !
** Ne présente le pas !*
** Ne pas le présente !*
** Ne présente pas le !*

Since pronouns form a closed class, the test data check for the position of pronominal objects in imperatives in an exhaustive way.

S_Imperative_Negation_infinitive_Wordorder

In an infinitive imperative, both negative elements must precede the verb. (as seen above, *pas* must follow a finite verb). Since negation behaves differently with infinitives than with finite verbs, the test data also check for the presence of both *ne* and *pas* in the infinitive imperative, and for their position in front of the verb.

Ne pas toucher !
** Ne toucher pas !*
** Ne toucher !*
** Pas toucher !*

S_Imperative_Negation_infinitive_PRON_Wordorder

In infinitive imperatives, pronominal objects precede the verb and must be non-tonic. The test data also check the position of the negative element *pas*, which must precede both the pronoun and the verb.

Ne pas le présenter !
** Ne le pas présenter !*
** Ne le présenter pas !*
** Ne pas présenter le !*

Parameters

PARAMETER			
WF	Attribute	Value	input
S_Imperative			
1	pers-num	2-sg	Pars !
0	pers-num	2-sg	Part !
S_Imperative_form			
1	mood	imperative	Sois gentil !
0	mood	imperative	* Es gentil !
1	orthography	verb without ‘s’	Chante !
0	orthography	verb without ‘s’	* Chantes !
S_Imperative_past			
1	tense	past	Soyez partis demain !
S_Imperative_PRON_Wordorder			
1	position_PRON	post_verb	Présente-le !
0	position_PRON	post_verb	* Le présente !
S_Imperative_PRON_Punctuation			
1	punct	PUNCT(-)	Présente-le !
0	punct	PUNCT(-)	* Présente le !
S_Imperative_PRON_euphonic-t			
1	euphonic-t	present	Va-t’en !
0	euphonic-t	present	* Va-en !
1	punct	PUNCT(’)	Va-t’en !
0	punct	PUNCT(’)	* Va-t-en !
S_Imperative_PRON_form			
1	form(PRON_pers)	tonic	Dis-moi !
0	form(PRON_pers)	tonic	* Dis-me !
S_Imperative_vocative			
1	punct	PUNCT(,)	Dominique, pars !
0	punct	PUNCT(,)	* Dominique pars !
S_Imperative_subjunctive-que			
1	mood	subjunctive	Qu’ il parte !
0	mood	subjunctive	* Qu’i l part !
1	pers	third	Qu’ il parte !
0	pers	third	* Que tu partes !
S_Imperative_subjunctive			
1	mood	subjunctive	Vive le roi !
0	mood	subjunctive	* Vit le roi !
S_Imperative_infinitive			
1	verbform	infinitive	Laisser reposer !

P A R A M E T E R			
WF	Attribute	Value	input
S_Imperative_Negation_Wordorder			
1	position_NEG(pas)	post_verb	Ne partez pas !
0	position_NEG(pas)	post_verb	* Ne pas partez !
1	negation	(ne, pas)	Ne partez pas !
0	negation	(ne, pas)	* Ne partez !
0	negation	(ne, pas)	* Partez pas !
S_Imperative_Negation_past_Wordorder			
1	position_NEG(pas)	post_aux	Ne soyez pas partis demain !
0	position_NEG(pas)	post_aux	* Ne pas soyez partis !
0	position_NEG(pas)	post_aux	* Ne soyez partis pas !
S_Imperative_Negation_infinitive_Wordorder			
1	position_NEG(pas)	ante_inf	Ne pas toucher !
0	position_NEG(pas)	ante_inf	* Ne toucher pas !
1	negation	(ne, pas)	Ne pas toucher !
0	negation	(ne, pas)	* Ne toucher !
0	negation	(ne, pas)	* Pas toucher !
S_Imperative_Negation_PRON_Wordorder			
1	position_NEG(pas)	post_verb,post_PRON	Ne le présente pas !
0	position_NEG(pas)	post_verb,post_PRON	* Ne le pas présente !
0	position_NEG(pas)	post_verb,post_PRON	* Ne pas le présente !
1	position_PRON	ante_verb	Ne le présente pas !
0	position_PRON	ante_verb	* Ne présente le pas !
S_Imperative_Negation_form			
1	form(PRON_pers)	non-tonic	Ne me présente pas !
0	form(PRON_pers)	non-tonic	* Ne moi présente pas !
S_Imperative_Negation_infinitive_PRON_Wordorder			
1	position_NEG(pas)	ante_verb,ante_PRON	Ne pas le présenter !
0	position_NEG(pas)	ante_verb,ante_PRON	* Ne le pas présenter !
0	position_NEG(pas)	ante_verb,ante_PRON	* Ne le présenter pas !
1	position_PRON	ante_verb	Ne pas le présenter !
0	position_PRON	ante_verb	* Ne pas présenter le !

Data

The total number of test items for S_Imperative is:

S_Imperative	
309	test items
126	well-formed items
183	ill-formed items

1.2.3 Conditionals

The "Conditional Construction" is a relation between two clauses, one of which gives the conditions under which the other one would hold true. It is thus a semantic relation and there are other syntactic means of expressing it. For the purpose of TSNLP, the data under this phenomenon is defined syntactically and does not contain constructions which are semantically similar, but stylistically very marked (e.g. the use of subjunctive clauses).

General Division of the Phenomenon

By definition, the conditional construction involves two clauses, the main clause (C_main) and the conditional clause (C_cond) which can be said to act as a sentential modifier. Thus, there is only one possible domain for this phenomenon, the sentence, i.e. we can only have S_Conditional.

The four most important characteristics of the conditional construction in French form the basis of the division of S_Conditional into four main subphenomena, which are discussed in turn in the following sections:

- order of the two clauses: S_Conditional_Wordorder
- conjunction: S_Conditional_conjunction
- mood and tense: S_Conditional_mood-tense
- coordination of clauses: S_Conditional_Coordination

S_Conditional_Wordorder

The ordering between the main clause and the conditional clause is an interaction with the phenomenon word order.[29]

The order of the two clauses is relatively free. However, if the conditional clause is in front of the main clause, a comma is obligatory.[30]

Si l'ingénieur vient, la réunion aura lieu.
La réunion aura lieu si l'ingénieur vient.
** Si l'ingénieur vient la réunion aura lieu.*

S_Conditional_conjunction

The existing TSNLP data mostly contains test items that check for the presence and correct use of the most usual conditional conjunction. i.e. *si*. However, besides *si*, several other conjunctions can introduce a conditional clause and we also describe here how additional TSNLP test data could be written for the other conjunctions.

[29] This means that both S_Conditional and S_Wordorder are given as the supertypes of S_Conditional_Wordorder.

[30] However, the subjunctive construction with *que* as a subordinating conjunction only accepts the order C_cond - C_main.

conjunction *si*

The most usual conditional conjunction is *si*. It requires the indicative mood in the conditional clause. In front of a vowel, *si* becomes *s'*.

Si l'ingénieur vient, la réunion aura lieu.
S'il vient, la réunion aura lieu.
* *S'l'ingénieur vient, la réunion aura lieu.*
* *Si il vient, la réunion aura lieu.*
* *Que l'ingénieur vient, la réunion aura lieu.*

The test data check that the correct conjunction is used and for the elision.

conjunction *au cas où*

au cas où behaves exactly like *si* (it does not require the subjunctive and the order of the two clauses is free) so the test data for *si* could be modified straightforwardly to systematically create similar test data for *au cas où*.

conjunction *que*

que as a conditional conjunction, is mostly literary. It requires the subjunctive mood in the C_cond, and the the order of the two clauses is fixed: C_cond must precede C_main.

Que l'ingénieur vienne, la réunion aura lieu.
Que l'ingénieur vient, la réunion aura lieu.
* *La réunion aura lieu que l'ingénieur vienne.*

conjunction *quand*

quand as a conditional conjunction is very literary and test data for this construction does not belong to the TSNLP core data..

Quand vous me haïriez, je ne me plaindrais pas. (Racine)

no conjunction

The use of a bare (with no *que* complementizer) plus-perfect subjunctive clause as a conditional clause (obligatorily in front of the main clause) is very literary and test data for this construction does not belong to the TSNLP core data.

Fussiez-vous au fond des abîmes, la main de Jupiter pourrait vous en tirer. (Racine)

S_Conditional_mood-tense

When the conjunction is *si*, there are strict rules for tense and mood concordancy between the conditional clause and the main clause. The conditional clauses is always in the indicative mood (present or past, never future), and the main clause may be either in the conditional (if the conditional clause is in the past) or in the indicative (if the conditional clause is in the present).[31]

[31] The test data could be divided into S_Conditional_mood and S_Conditional_tense, but because the choice of mood depends on that of tense, it is easier to keep them together.

Si l'ingénieur vient, la réunion aura lieu.
** Si l'ingénieur viendrait, la réunion aura lieu.*
** Si l'ingénieur viendra, la réunion aura lieu.*
** Si l'ingénieur vient, la réunion aurait lieu.*

Whether the main clause is in the conditional present or past depends on whether that clause describes a 'realis' or 'irrealis' situation. As that distinction is necessarily semantic in nature, it is not tested in TSNLP.

NOTE: *si* can also be used to introduce a temporal clause. In that case, the main clause is not in the subjunctive. Thus, the test data for past conditional clause can only be used meaningfully if the system under test gives some indication about the function of the *si* clause, either as a conditional or as a temporal adverbial.

• temporal:
Si la porte était fermée, nous passions par la fenêtre.

• conditional:
Si la porte était fermée, nous passerions par la fenêtre.

S_Conditional_emphatic

There are no test items yet for the emphatic conditional construction in the TSNLP test data.

This construction uses *si* as the conditional conjunction, and it is characterized by a fixed ordering of the two clauses and by the use of "expletive negation":

• the conditional clause must follow the main clause.
• the negative particle *ne* must be present, but not any other negative element.

La réunion n'aura lieu que si l'ingénieur vient.
** Que si l'ingénieur vient, la réunion aura lieu.*
** La réunion aura lieu que si l'ingénieur vient.*
** La réunion n'aura pas lieu que si l'ingénieur vient.*

The TSNLP test data can easily be augmented for this phenomenon by using a subset of the positive test items for the non-emphatic conditional with *si* as a basis for deriving the positive test items for the emphatic construction.

S_Conditional_Coordination

The coordination of conditional clauses is part of the general phenomenon coordination.[32]

When two conditional clauses are coordinated, the second one is introduced by *que* and not by *si*. It should then be in the subjunctive, unless the first conjunct is in the present. However, the use of the indicative is not a mistake.

Si l'ingénieur vient et qu'il est prêt, la réunion aura lieu.
Si l'ingénieur venait et qu'il soit prêt, la réunion aurait lieu.

[32]This means that both S_Conditional and S_Coordination are given as the supertypes of S_Conditional_Coordination.

Parameters

S_Conditional_Wordorder

Attribute	Value
position_C_cond	(ante_main-clause,post_main-clause)
punct	PUNCT(,)

S_Conditional_conjunction

Attribute	Value
elision	(yes,no)

S_Conditional_mood-tense

Attribute	Value
mood-tense(C_cond)	(ind-pres,ind-passecomp, ind-imperf,ind-plusqueparf,subj-pres)
mood-tense(C_main)	(ind-fut,cond-pres,cond-past)

S_Conditional_Coordination

Attribute	Value
coord-conjunctions	et(C_cpl(si),C_cpl(que))
mood-tense	(ind-pres,ind-imper, etc.)

S_Conditional_emphatic

Attribute	Value
position_C_cond	(ante_main-clause,post_main-clause)
NEG_part(ne)	(present, absent)

Data

The total number of test items for S_Conditional is:

S_Conditional	
39	test items
16	well-formed items
20	ill-formed items
3	dubious items

1.3 German

Sentences types in German are determined by several parameters, namly position of the verb, verb mood (imperative vs. indicative/conjunctive) and presence of wh-elements (interrogative pro forms. The four main types of sentences are imperatives, interrogatives with no wh-pronoun, declaratives and questions with a wh-pronoun. The wh-element can occur in different positions in the sentence, but pied piping is much more restricted than in English. Mainly it is possible with prenominal possessives (*wessen*), specifier (*welcher*) or NP complements of prepositions.

There also exists a class of independent verb final sentences, either starting with a complementizer or with a wh-pronoun:

Daß du doch schläfst.
Wo der Manager schläft?

It is also possible to construct sentences where the verb is not finite but a bare or *zu* infinitive. The bare infinitive often is used in so-called telegram style and only possible with verbs building the perfective forms with *sein* or with passive forms.

Kohl zurückgetreten.
Kohl angeklagt.

The infinitive with *zu* is used like a passive:

Hausrat zu verkaufen.

Certain types of sentences can be ambiguous between

- questions and imperatives

  Kommen Sie?
  Kommen Sie!

- declaratives with a topic ellipsis and interrogatives with optional complements

  Kennst du schon.
  Kennst du schon?

- request and question function

  Du schläfst!
  Du schläfst?

For verb-final sentences there are ambiguities beetween

- declarative and exclamatory sentences

  Daß du <u>doch</u> schläfst.
  Daß du <u>doch</u> schläfst!

- optatives and interrogatives with elliptical

  Wenn der Manager <u>doch</u> schlafen würde.
  Wenn der Manager <u>doch</u> schlafen würde?

However, the function of a sentence sometimes can be determined by certain modal particles.

Schläfst du eigentlich?
**Schläfst du eigentlich!*

1.3.1 General Division of the Phenomenon

The following four basic types of sentences are covered by the data:

- imperatives
- interrogatives
- declaratives
- conditionals

Ill-formed items result from the variation over the modal particals *denn, doch, eigentlich, ja, mal, wohl* in relevant sentences:

Schläfst du denn?
Schläfst du doch?
Schläfst du eigentlich?
**Schläfst du ja?*
Schläfst du mal?
Schläfst du wohl?

Test sets group items, that represent more than one type.

Schlaft ihr?
Schlaftt ihr!

S_Types_Imperatives

There are two imperative forms without a subject:

Geh!
Geht!

For the imperatives with subject, there is an item for every combination of number and person of the subject (i.e. second singular and all plural forms). For the second person singular there are both possible verb forms annotated.

Schlafe du!
Schlaf du!
Schlafen wir!
Schlaft ihr!
Schalfen Sie!

There are also examples for imperatives with

- third person singular as a subject

  Brat mir einer einen Storch!

- the verb in bare infinitive form

  Den Rasen nicht betreten!

- the verb in the final position

  Daß du ja schläfst!

S_Types_Interrogatives

The data includes for the questions with a wh-pronoun:

- items for most of the wh-pronouns in the following three configurations:

  Wo schläft der Manager?
  Der Manager schläft wo? (echo questions)
  Wo der Manager schläft ?

- items for the main constructions with an embedded wh-pronoun (pide piping)

  Mit wessen Auto fährt der Manager?
  Mit wieviel Autos fährt der Manager?
  Mit welchem Auto fährt der Manager?

- items for most wh-adverb phrases with the alternative question, starting with a prepositional phrase

  Wonach strebt der Manager?
  Nach was strebt der Manager?

- an example where the verb is in infinitive form

  Was tun?

For the questions with no wh-pronoun the data covers the following:

- verb first sentences, also with different modal particles and ellipsis

  Schläfst du?
  Schläfst du doch?
  **Schläfst du ja?*
  Kennst du schon?

- some verb final sentences, also with different modal particles

  Wenn der Manager doch schläft?
  Ob der Manager doch schläft?

S_Types_Declaratives

Because most of the items for other phenomena are declarative sentences, we added only a few special ones:

- sentences with topic ellipsis, which is only grammatical if the deleted constituent was not an expletive element

  Kennst du schon.
  **Regnet morgen.*

- independent verb final sentences, also with different modal particles

  Daß du doch schläfst.
  **Daß du eigentlich schläfst.*

- sentences, where the verb is not finite, but a bare or *zu* infinitive (e.g. newspaper headlines)

  Kohl zurückgetreten.
  Kohl angeklagt.
  Hausrat zu verkaufen.

S_Types_Conditionals

For conditionals the data includes only four items: A *wenn...dann* sentence with all variations of ellipses of the adverbial complementizers

Wenn der Manager kommt, dann gehe ich.
Wenn der Manager kommt, gehe ich.
Kommt der Manager, dann gehe ich.
Kommt der Manager, gehe ich.

1.3.2 Data

The total number of test items for S_Types is:

S_Types	
121	test items
107	well-formed items
13	ill-formed items
1	dubious items

Chapter 2

Diathesis

2.1 English

Diathesis is the phenomenon that describes changes in the predication structure of the verb. Strictly speaking the only possible domain is the clause (C), however, by extension noun phrases (NP) could also be treated in Diathesis. It is however only the verb that shows differences in form between passive and active.

Diathesis is a phenomenon that is heavily determined by lexical and semantic properties. TSNLP focuses on syntactic testing, and as a consequence we limit ourselves as much as possible to syntactic properties.

English has four forms of Diathesis:

- Active, e.g. *He proposed the development.*
- Passive, e.g. *This development is suggested (by the manager).*
- Middle, e.g. *They have a house.* This concerns the transitive verbs that have an active form, but no passive (although being transitive). From the syntactic viewpoint, they coincide with the active verbs. There is only no passive counterpart.
- Reflexive, e.g. *They pride themselves.* These verbs are transitive, but the category of the object is restricted to a reflexive pronoun. They also don't have a passive. Again, from the syntactic point of view, they coincide with the active verbs.

The sets of this phenomenon are constructed in another way than for the other phenomena. Elsewhere they contain one well-formed item, and all other items (well-formed or ill-formed) derived from it. Here they contain an active item, and the "corresponding" passive item(s).[1]

In this phenomenon there are some items adapted from the Hewlett-Packard test suite.

There are no reflexive constructions present in the test suite.

[1] Active and passive items belong to a different phenomenon.

2.1.1 C_Diathesis-Active

The items in this subphenomenon are taken over from C_Complementation. The classification remains exactly the same with one general change in phenomenon names: "C_Complementation" becomes "C_Diathesis-Active".

This classification is given in section 4.1.1. These items are a subset of the well-formed active C_Complementation items.

2.1.2 C_Diathesis-Passive

The ways to derive items from the well-formed active C_Complementation items are the following:

1. Each complement of the active frame in its turn is made subject of the verb (in the passive form).

   This means that nominal objects are placed on the first place in the sentence. Verbal and sentential complements produce two items: one with the C or VP on the first place, and one with the pronoun *it* and the C or VP at the end of the sentence.

2. A passive item gets an instance with and without the *by*-phrase.

Most of the test items contain the passive auxiliary *be* which is most commonly used. The auxiliary *get* is tested as a separate subphenomenon (because of its limited use).

The classification itself gives first the valency of the (passive) frame, analogous to C_Complementation and C_Diathesis-Active. The presence or absence of *by* does not play a role in this classification.

- C_Diathesis-Passive-Monovalent
- C_Diathesis-Passive-Divalent
- C_Diathesis-Passive-Trivalent
- C_Diathesis-Passive-Get

A consequence of this method is that even passives belonging to a different subphenomenon can be put in the same set. E.g. (passive related to the active sentence *They say that she will come.*):
It is being said that she will come. (C_Diathesis-Passive-Divalent)
That she will come is being said. (C_Diathesis-Passive-Monovalent)

C_Diathesis-Passive-Monovalent

Passive constructions with only a subject. Some of them have no passive at all, others do have one. If there is a *by*-phrase, it is optional in most cases.

* *It is rained.*
* *He is come.*

He is carried .
He is carried by him.

He is born.

C_Diathesis-Passive-Divalent

These verbs have in the passive one object. Most of the time there is at least one passive.

He is given the book.
He is given the book by him.
The book is given to him.
The book is given to him by him.
** The book is given him.*

C_Diathesis-Passive-Trivalent

Two objects in the passive construction.

He is bet one pound by him that she leaves.

C_Diathesis-Passive-Get

The auxiliary is *get*. A systematic variation is difficult to judge, and the use of auxiliary is lexically determined.

He got run over.
He got run over by the car.

** He got given the book.*

2.1.3 C_Diathesis-Middle

These verbs look like transitive verbs, but they have no passive.

** An office is had.*

2.1.4 Data

C_Diathesis	
281	test items
157	well-formed items
112	ill-formed items
0	relatively ill-formed items
12	dubious items

2.2 French

Diathesis concerns the possible change of arguments that predicate allows. Some of these changes are accompanied by specific morphological forms e.g. the alternation between active and passive voice, others are not morphologically marked, but have different syntactic realizations for the arguments, e.g. the reflexive and middle constructions.

Since we are not concerned in TSNLP with the lexical coverage of the systems that may be under test, we do write test items which aim at proving that specific lexical items are correctly entered in the lexicon with all their possible alternations. Rather, we want to enable a user to check that the different syntactic realizations are correctly recognized and analyzed, and that for a given predicate, the semantic relations of the predicate with its syntactic arguments are correctly interpreted.

2.2.1 General Division of the Phenomenon

By definition, diathesis is a phenomenon whose domain can only be the clause, although, once more complex subordinate structures are checked for a system, it will be worth testing that the diathesis relations are also correctly interpreted when one of the arguments is e.g. questioned or relativized.

le chien qui a été vu
** le chien qui il a été vu*
le chien qui se gratte les oreilles
** le chien qui il se gratte les oreilles*
etc.

For French, we subdivise the general phenomenon of C_Diathesis into:
- C_Diathesis_Passive,
- C_Diathesis_Reflexive,
- C_Diathesis_Middle.

For ease of use and reference, we abbreviate these sub-phenomena as C_Passive, C_Reflexive and C_Middle, respectively.

The presuppositions needed for all three sub-phenomena are:
- C_Wordorder
- C_Agreement_subj(NP)_V
- C_Complementation

The TSNLP database so far only contains data for C_Passive, but the documentation given here should enable users to systematically construct data for C_Reflexive and C_Middle on the same model, because the three constructions present the same or similar sub-phenomena.

2.2.2 C_Passive

C_Passive_passivizability

Some verbs do not passivize, for instance intransitive, but also verbs taking prepositional objects (unlike English). This should not be expected to be tested by a test suite unless we perform extensive lexical replacement to test all the verbs of the lexicon.

Nevertheless, our TSNLP data contains a number of test sets with a well-formed active sentence and a wrong passive sentence which contrast with test sets containing both a well-formed active and its corresponding well-formed passive. These positive test items are based on the positive test items used in the phenomenon C_Complementation (which assumes by default that an active diathesis for the verbs) and illustrate the different behaviors of verbs which are transitive, ditransitive, predicative or verbs which take other kinds of complements.

Il a vu Jean.
Jean a été vu.

Il a parlé de Jean.
** Jean a été parlé de.*

C_Passive_auxiliary

The form of the passivized verb in the "Présent" is the same as that of an intransitive verb in the "Passé Composé". It must use the auxiliary *être* and not *avoir*.

However, the form with the auxiliary *avoir* is the correct form for the "Passé Composé" of a transitive verb. What must be checked is that a sentence is not both accepted with *avoir* as an the auxiliary and interpreted as a passive.

Il voit Jean.
Jean est vu.

Jean est venu.
** Jean a vu.*
** Jean a venu.*

The TSNLP data systematically covers the choice of auxiliary for the passive.

C_Passive_ppast

The verb form must be in the past participle. One type of error is the use of an infinitive instead of a past participle.

Jean est vu.
** Jean est voir.*

The TSNLP data systematically covers the use of the past participle as opposed to the infinitive in the passive. In these test sets, the correct passive sentence is the well-formed test item.

Past Participle Agreement

As in all all verbal forms conjugated with the auxiliary *être*, the past participle must agree with the syntactic subject in number and gender.

Jean est vu.
** Jean est vue.*

* *Jean est vus.*
* *Jean est vues.*

Jeanne est vue.
* *Jeanne est vu.*
* *Jeanne est vus.*
* *Jeanne est vues.*

This second type of error for the past participle in the passive construction is covered exhaustively and systematically in the TSNLP test data (with the correct passive sentence as the positive test item), under C_Agreement, where these data are cross-referenced to C_Passive.

C_Passive_syntactic-subject

The main characteristic of the passive is that its syntactic subject is the second semantic argument, which in French is the direct object of the active voice.

We may only test this characteristic if the system under test has some semantic information in the lexical entries of the predicates.

Jean a bu toute l'eau.
L'eau a été bue.

* *Jean a bu la voiture.*
* *La voiture a été bue.*

The TSNLP test data contains a number of test sets, in which the same verb is used both in a (semantically) well-formed and a (semantically) ill-formed active sentence, with the corresponding (semantically) well-formed and (semantically) and ill-formed passive sentences.

C_Passive_agent-phrase

Expressing the agent is never obligatory, but if it is expressed, the agent phrase must be interpreted as the "logical subject" of the predicate.

If the system can handle semantic restrictions on the arguments of the predicate, then we can test that these are correctly preserved in the passive.

Michel caresse le chien.
Le chien est caressé par Michel

* *La table regarde le chien.*
* *Le chien est regardé par la table.*

As with the phenomenon of C_Passive_syntactic-subject, the TSNLP test data for C_Passive_agent-phrase contains a number of test sets, in which the same verb is used both in a (semantically) well-formed and a (semantically) ill-formed active sentence, with the corresponding (semantically) well-formed and (semantically) and ill-formed passive sentences.

2.2.3 C_Reflexive

There are 2 sorts of reflexives in French:

• one merely expresses coreference between the subject and the object (which may be an indirect or prepositional object) of a predicate;

• the other has no meaning and must be listed in the lexicon: the "inherently reflexive" verbs.

Both types present the same morphological and syntactic characteristics and the distinction does not need to be reflected in the TSNLP testdata.

C_Reflexive_auxiliary

Both types of reflexive verbs must use the auxiliary *être*.

Jean voit Michel.
Jean a vu Michel.
Jean se voit dans le miroir.
Jean s'est vu dans le miroir.
** Jean s'a vu dans le miroir.*

Jean parle à Michel.
Jean a parlé à Michel.
Jean se parle tout seul.
Jean s'est parlé tout seul.
** Jean s'a parlé tout seul.*

Jean s'est évanoui.
** Jean s'a évanoui.*
L'échaffaudage s'est effondré.
** L'échaffaudage s'a effondré.*

C_Reflexive_ppast

The verb form must be in the past participle. One type of error is the use of an infinitive instead of a past participle.

L'échaffaudage s'est effondré.
** L'échaffaudage s'est effondrer.*

Reflexive: Past Participle Agreement

As in all all verbal forms conjugated with the auxiliary *être*, the past participle must agree with the syntactic subject in number and gender.

L'échaffaudage s'est effondré.
** L'échaffaudage s'est effondrée.*
** L'échaffaudage s'est effondrés.*
** L'échaffaudage s'est effondrées.*

Jeanne s'est évanouie.
** Jeanne s'est évanoui.*
** Jeanne s'est évanouies.*
** Jeanne s'est évanouis.*

This aspect of the reflexive construction is covered exhaustively and systematically in the TSNLP test under C_Agreement.

C_Reflexive_PRON-refl_Agreement

The reflexive pronoun must agree (in person and number) with the subject.

L'échaffaudage s'est effondré.
** L'échaffaudage m'est effondré.*
** L'échaffaudage t'est effondré.*
** L'échaffaudage l'est effondré.*
** L'échaffaudage nous est effondré.*
** L'échaffaudage vous est effondré.*

C_Reflexive_PRON-refl_Interpretation

If the system gives some semantic information, we can test that the reflexive pronoun is interpreted as being coreferential with the subject. This can only be done with parsers, once agreement has been checked.

2.2.4 C_Middle

The middle construction has the same morphological and syntactic structure as the reflexive: use of the auxiliary *être*, and of the reflexive pronoun. These characteristics can be tested in the same way.

The middle has the same argument structure as the passive, i.e. the active ayntactic object is the middle syntactic subject. This can be tested in the same way as for the passive.

C_Middle_auxiliary

Ce logiciel se vend bien.
Ce logiciel s'est bien vendu.
** Ce logiciel s'a bien vendu.*

C_Middle_ppast

The verb form must be in the past participle. One type of error is the use of an infinitive instead of a past participle.

Ce logiciel s'est bien vendu.
** Ce logiciel s'est bien vendre.*

Middle: Past Participle Agreement

As in all all verbal forms conjugated with the auxiliary *être*, the past participle must agree with the syntactic subject in number and gender.

Ce logiciel s'est bien vendu.
** Ce logiciel s'est bien vendue.*
** Ce logiciel s'est bien vendus.*
** Ce logiciel s'est bien vendues.*

Ces voiture se sont bien vendues.
** Ces voiture se sont bien vendu.*
** Ces voiture se sont bien vendue.*
** Ces voiture se sont bien vendus.*

This aspect of the middle construction is covered exhaustively and systematically in the TSNLP test under C_Agreement.

C_Middle_syntactic-subject

As with the passive, the syntactic subject is the second semantic argument of the verb, which in French is the direct object of the active voice. We may only test this characteristic if the system under test has some semantic information in the lexical entries of the predicates.

** Jean a bu les voitures.*
** Les voitures se sont bues.*

C_Middle_agent-phrase

Unlike with the passive, the agent cannot be expressed in a middle construction.

Ce logiciel se vend bien.
Les concessionaires vendent bien ce logiciel.
** Ce logiciel se vend bien par les concessionaires.*

2.2.5 Parameters

C_Passive_passivizability

Attribute	Value
passivizability	(yes,no)

C_Passive_auxiliary

Attribute	Value
aux	*être*

C_Passive_ppast

Attribute	Value
verbform	nfinite-ppast

2.2.6 Data

The total number of test items for C_Passive is:

C_Passive	
295	test items
176	well-formed items
114	ill-formed items
4	dubious items
1	relatively ill-formed item

2.3 German

German has three distinct auxiliaries to form the passive voice, the most common of which are the procedural (*werden*) and stative (*sein*) passive.

- *werden*

  Das Fenster wird geschlossen.

- *sein*

  Das Fenster ist geschlossen.

- *bleiben*

  Das Fenster bleibt geschlossen.

All of them govern a verb in the perfective form and represent different aspects of the passive. Passive is formed by removing or 'downgrading' a nominative complement and promoting an accusative complement (direct object) to the subject function. From the C_Complementation phenomenon all subcat frames were submitted to this variation.

Additionally, German also has impersonal passive, that can be formed with intransitive verbs.

Er schläft.
Es wird geschlafen.

However, there also is the possibility to form the passive voice with the indirect object raised to the subject position, the so-called 'bekommen'-passive:

Die Frau schenkt ihm *einen Schlüssel.*
Er *bekommt einen Schlüssel geschenkt.*

There is also a modal passive with *zu* and a *lassen* passive foreseen in the system, but not yet represented by data.

Das Buch ist zu lesen.
Er läßt das Buch lesen.

2.3.1 General Division of the Phenomenon

Passive test items were derived from corresponding grammatical sentences in the active voice in the following way:

- deletion of the subject
- promotion of the direct object into the subject position
- other complements remain unchanged

According to this schema, passive sentences were constructed for all wellformed test items from the C_Complementation phenomenon.

A second set of passive sentences were built in parallel to the full paradigm of C_Tense-Aspect-Modality items. For all sentences there are corresponding sentences in the passive voice. We did not construct passive sentences for the items built with modal verbs or copula verbs. Also the items where the verb can form the perfective forms with both, *haben* and *sein* were not passivized.

For one verb there are sentences in the third person singular in all relevant tenses where the passive voice was built with the wrong auxiliary.

Es ist geschlafen worden.
**Es hat geschlafen worden.*

2.3.2 Data

The total number of test items for C_Diathesis is:

C_Diathesis	
295	test items
147	well-formed items
147	ill-formed items
1	dubious items

Chapter 3

Tense, Aspect, Modality

The only domain for Tense, Aspect and Modality is that of the clause, so there is only the phenomenon of C_Tense-Aspect-Modality.

3.1 English

English has two tenses: present tense and past tense. As the names imply, the present tense usually applies to present time and past tense to past time (*He works today. He worked yesterday.*). English has no future tense, but a number of possibilities for expressing future time (eg. it may be rendered by means of modal auxiliaries (eg. *will*) or by simple present or progressive forms. Thus futurity, modality and aspect are closely related in English, and tense, aspect and modality are treated together in TSNLP.

Aspect refers to the way in which the verb action is viewed, and there are two sets of aspectual contrast in English: perfective/non-perfective (e.g. *He has worked.* (perfective)) and progressive/non-progressive (e.g. *He is working.* (progressive). Progressive aspect, as the name suggests, indicates an action in progress instead of the occurrence of an action, or the existence of a state.

Modality may be defined, following Quirk et al. [Quirk *et al.*, 1991] as the manner in which the meaning of a clause is qualified so as to reflect the speaker's judgment of the likelihood of the proposition it expresses being true and includes such concepts as volition, probability and obligation.

3.1.1 General Division of the Phenomenon

The data illustrates the paradigm of tense and modal auxiliaries in English.

The main verb in English can be preceded by up to four auxiliaries (*do* cannot combine with any of the others). The four auxiliary positions are:

	AUXILIARIES			MAIN VERB
modal	perfect	progressive	passive	
can	have	be	be	take, etc.
may				
...				

Each auxiliary determines the form (affix) of the following verb (affix-hopping).[1] The rules are:

Auxiliary	Inflectional form of following verb
modal	base form
perfect	-*en* form
progressive	-*ing* form
passive	-*en* form

Examples:

He may succeed.
He may be seen.
He may be succeeding.
He may be being seen.
He may have succeeded.
He may have been seen.
He may have been succeeding.
He may have been being seen.

The initial verb can carry a tense inflection (+/− PAST): all non-initial verbs are non-tensed.

Auxiliaries are frequently contracted in English: *he will* > *he'll*

Both full and contracted forms (with the subject pronoun *he*) are covered in the TSNLP data as follows:

- C_Tense-Aspect-Modality
- C_Tense-Aspect-Modality-Contractions

and both present and past tense forms are given.

The central set of modal verbs are exhaustively treated. They are: *may*, *will*, *can*, *must*, *shall*.[2] Marginal modals include *need*, *dare*, *ought* and *used to*. *Need* and *dare* are restricted in their use to nonassertive contexts (mainly negative and interrogative sentences), so do not appear in this data set.

C_agreement and standard Word Order is assumed in all the data.

3.1.2 C_Tense-Aspect-Modality

Ill-formed items are created by changing the form of the verb that auxiliaries subcategorise for:

* *he may succeeding*

[1] See Huddleston [Huddleston, 1984], p.129.
[2] This classification is due to Huddleston [Huddleston, 1984], who treats *will* as a modal verb.

C_Tense-Aspect-Modality-DO_aux

- **Present**

  He does succeed.
  ** He does succeeding.*

- **Past**

  He did succeed.
  ** He did succeeding.*

Only the present form is shown in the examples below:

C_Tense-Aspect-Modality-BE_HAVE

She succeeds.

She has succeeded.
** She has succeeding.*

She is seen.

C_Tense-Aspect-Modality-MAY

He may succeed.
** He may succeeding.*

C_Tense-Aspect-Modality-WILL

He will succeed.
** He will succeeding.*

C_Tense-Aspect-Modality-CAN

He can succeed.
** He can succeeding.*

C_Tense-Aspect-Modality-SHALL

She shall succeed.
** She shall succeeding.*

C_Tense-Aspect-Modality-MUST

She must succeed.
** She must succeeding.*

C_Tense-Aspect-Modality-OUGHT

She ought to succeed.
** She ought succeeding.*

C_Tense-Aspect-Modality-USED

He used to succeed.
** He used succeeding.*

3.1.3 C_Tense-Aspect-Modality-Contractions

The subdivisions of the phenomena are identical to the full-forms.

3.1.4 Parameters

C_Tense-Aspect-Modality

Attribute	Value
comp-type	(VP_bare-inf,VP_Perf, VP_to-inf)

3.1.5 Data

The total number of test items for C_Tense-Aspect-Modality is:

C_Tense-Aspect-Modality	
196	test items
157	well-formed items
38	ill-formed items
1	dubious items
0	relatively ill-formed items

3.2 French

When talking about tense, a distinction has to be drawn between "referential" (or semantic) tense and morphological tense. The former expresses a temporal relation between the event described by the verb and the moment of utterance. The latter concerns the morpho-syntactic features of the verb.

Unlike German and English, French has three **morphological** tenses: past, present and future:

Je marchais.	(imparfait)
Je marchai.	(passé simple)
Je marche.	(présent)
Je marcherai.	(futur simple)

As far as **semantic** tense is concerned, there are different possibilities in French to express past, present and future (with different aspectual properties). The following table summarizes the most important ones, distinguishing between **simple** and **compound** tenses:

	simple tenses	compound tenses
past	imparfait, passé simple	passé composé, plus-que-parfait
present	présent	
future	futur simple	futur composé, futur antérieur

According to Comrie [Comrie, 1976], "aspects are different ways of viewing the internal temporal consistency of a situation."(p.3). Comrie analyzes the aspectual system of French in the following way: "In written French, there is a three-way formal distinction among the past tenses: Past Definite (with perfective meaning. e.g. *j'écrivis* 'I wrote'), Imperfect (with imperfective meaning. e.g. *j'écrivais* 'I was writing, I used to write'), and Perfect (with perfective meaning. e.g. *j'ai écrit* 'I have written'). The opposition between Past Definite and Imperfect is restricted to the Past Tense, while there are Perfect forms for all tenses."

Modality refers to the attitude of the speaker towards the truth-value of the proposition expressed by the sentence.

3.2.1 General Division of the Phenomenon

As shown in the preceding section, Tense, Aspect and Modality are closely related to semantics. Since the TSNLP data is not meant to test for semantics (which would prove to be quite difficult), it only tests for some of their morpho-syntactic aspects.

Tense, Aspect and Modality are terated at the level of the clause, and there is only one domain, i.e. C_Tense-Aspect-Modality.

This phenomenon is then divided according to three aspects concerning Tense, where the first two have to do with morpho-syntactic features, whereas the last one concerns semantics. Thus, Aspect and Modality are in fact not yet tested in the existing data.

- C_Tense-Aspect-Modality_simple-tense
- C_Tense-Aspect-Modality_compound-tense
- C_Tense-Aspect-Modality_mod-temp

3.2.2 C_Tense-Aspect-Modality_simple-tense

The data in this section tests for the finiteness of the simple tenses, i.e. of “imparfait”, “passé simple”,“présent” and “futur simple”. In order to derive ill-formed items, the finite verb is systematically replaced with all possible non-finite verb forms: “infinitive”, “past participle” and “participe présent (gerundive)”.

The input is a sentence which consists of a subject pronoun and an intransitive verb. Since pronouns belong to a closed class, they are exhaustively tested.

Je marche.
Je marchais.
Je marchai.
Je marcherai.
** Je marcher.*
** Je marché.*
** Je marchant.*

3.2.3 C_Tense-Aspect-Modality_compound-tense

The data tests for compound tenses consisting of an auxiliary verb and a past participle. Thus, in order to derive ill-formed items, the auxiliary is changed and/or the past participle is replaced by either another non finite verb form (such as an “infinitive” or a “past participle”) or a finite verb form.

J'ai marché.
** Je suis marché.*
** J'ai marchant.*
** J'ai marcher.*
** J'ai marche.*
** Je suis marchant.*

The input is a sentence which consists of a subject pronoun and an intransitive verb. Since both pronouns and auxiliary belong to closed classes, they are exhaustively tested.

3.2.4 C_Tense-Aspect-Modality_mod-temp

Here, what is tested is a semantic feature: for each tense, a (semantically) corresponding temporal modifier expressing either past, present or future, i.e. *hier, maintenant* or *demain* is added to the input sentence.[3]

[3]We take *maintenant* since it is more restrictive than *aujourd'hui*. If the temporal modifier is *aujourd'hui*, the semantic tense of the verb is almost unrestricted, e.g. *Il a travaillé aujourd'hui.* is correct.

In order to derive ill-formed items, the tense of the verb is replaced with one which does not correspond to the tense of the modifier.

Il a travaillé hier.
** Il travaille hier.*

3.2.5 Parameters

C_Tense-Aspect-Modality_simple-tense

Attribute	Value	instance
verbform	finite	
EXAMPLES		
Je marche.		
verbform	finite	marche
** Je marcher.*		
verbform	finite	* marcher

C_Tense-Aspect-Modality_compound-tense

Attribute	Value	instance
verbform aux	nfinite-ppast avoir	
EXAMPLES		
** J'ai marché .*		
verbform aux	nfinite-ppast avoir	marché ai
** J'ai marcher.*		
verbform	nfinite-ppast	* marcher
** Je suis marché*		
aux	avoir	* suis

C_Tense-Aspect-Modality_mod-temp

Attribute	Value	instance
tense_sem	(present,past,future)	
EXAMPLES		
Il a travaillé hier.		
tense_sem tense_sem	past past	a travaillé hier
Il travaille hier.		
tense_sem	past	* travaille

3.2.6 Data

The total number of test items for C_Tense-Aspect-Modality is:

C_Tense-Aspect-Modality	
352	test items
77	well-formed items
273	ill-formed items
2	dubious items

3.3 German

German has two tenses: present and past tense. Both are realized by inflection of the verb stem. Additionally there are two auxiliaries (*haben, sein*) encoding perfective aspect in the widest sense that govern verbs in their perfective forms. Accordingly it is possible to form four tenses of each verb. The future is formed with the auxiliary *werden* that governs verbs in their bare infinitival forms. *Werden* is restricted to finite forms. It can combine, again, with *haben* and *sein* such that there result two future tenses.

Furthermore, there are two moods in German: indicative and conjunctive. The conjunctive mood comes in two versions: one is formed by inflection of the verb stem while the other is formed by the tense auxiliary in its conjunctive form and the verb in the appropriate form.

Modal verbs can occur at almost any position in the verb cluster. The typical order is the following:

	VV	(PV)	(TV)	(MV)*	(werden/haben)[4]
weil er	geschlagen	worden	sein	müssen	wird

3.3.1 General Division of the Phenomenon

There are sentences for a transitive verb and for two intransitive verbs — one building the perfective forms with *haben*, one with *sein* — in all possibile variations of the tense, i.e. present, past, present perfective, past perfective, future and future in the past in indicative as well as in conjunctive mood.

Er liebt den Manager.
Er wird geschlafen haben.
Er sei gegangen.

These variations over the tense were also made in the passive voice for the same verbs. The result is a grammatical personal passive for the transitive verb, a grammatical impersonal passive for the intransitive verb building the perfective forms with *haben* and a partly ungrammatical passive for the intransitive verb building the perfective forms with *sein*.

Der Manager wird geliebt.
Es wird geschlafen worden sein.
**Es sei gegangen worden.*

For these verbs the imperatives were also constructed.

Geh!
Geht!

The same variations over the tense were also made with one verb for each of the following classes. Corresponding items in passive voice were only built of the first five verbs and are always ungrammatical.

- verbs with a dummy subject

  Es regnet.
  **Es wird geregnet.*

[4] *werden/haben* can only occur if there is also a modal verb.

- verbs with no subject

  Den Manager schwindelt.
  **Der Manager wird geschwindelt.*

- psych verbs with a subject and a direct object

  Der Plan hat den Manager geschmerzt.
  **Der Manager wird geschmerzt.*

- psych verbs with a subject and an indirect object

  Der Manager fehlt ihm.
  **Ihm wird gefehlt.*

- *haben* as a main verb

  Er hat den Plan.
  **Der Plan wird gehabt.*

- *sein* as a main verb

  Er ist der Manager.

- a modal verb with a verbal complement

  Er darf spielen.

- a modal verb as a main verb where both options are represented: the modal verb as a participle or as a so-called 'Ersatzinfinitiv' in the perfective forms

  Er darf es[5]*.*
  Er hat es gedurft.
  Er hat es dürfen.

For verbs building the perfective forms with either *haben* or *sein* (according to their transitivity), there are items for all the relevant forms (i.e. present perfect, past perfect, future II) for one representative verb.

Er hat den Manager gefahren.
Er ist nach Saarbrücken gefahren.

There are also sentences where a free selection of the auxiliary for the perfective forms is possible for the verb.

Er hat gesegelt.
Er ist gesegelt.

Only in the present tense there are examples for modal verbs as main verbs with a non-verbal complement.

Er mag den Manager.

All these sentences are in the third person singular. Only for one intransitive verb there are variations over the person in the present tense.

[5] *es* here is a anaphoric for a verbal complement

Ich schlafe.
Du schläfst.
Er schläft.
Wir schlafen.
Ihr schlaft.
Sie schlafen.

The ill-formed items are constructed by

- building the passive voice where it is impossible, as mentioned above
- building the passive voice for one verb with an incorrect auxiliary

  **Es hat geschlafen worden.*
- the systematical substitution of the correct auxiliary with the incorrect one in the perfective forms for a transitive verb, an intransitive verb building the perfective forms with *haben*, an intransitive verb building the perfective forms with *sein*, a modal verb with an verbal complement, a modalverb with *es* and one verb with a transitive or intransitive meaning, according to the auxiliary used for the perfective forms

  **Er ist den Manager geliebt.*
  **Er ist geschlafen.*
  **Er hat gegangen.*
  **Er ist spielen dürfen.*
  **Er ist es gedurft.*
  **Er ist den Manager gefahren.*
  **Er hat nach Saarbrücken gefahren.*
- the substitution of the infinitive of the modal verb in the perfective forms with a participle

  **Er hat spielen gedurft.*

3.3.2 Data

The total number of test items for C_Tense-Aspect-Modality is:

C_Tense-Aspect-Modality	
322	test items
188	well-formed items
134	ill-formed items
0	dubious items

Chapter 4

Complementation

4.1 English

Complementation is the phenomenon containing the tests for completion of (the meaning of) a word of a certain category in a phrase. There are several domains possible: verbs, nouns, adjectives and prepositions. There are several points from which one can start: a list of all possible and different "surface" frames, a list of all frames with information about how they are interlinked in the lexicon (e.g. for verbs: a *that*-clause is nearly always a possibility next to a subordinate clause with an empty complementizer), or a list that focuses on the meanings of the word that needs complementation rather than on the syntax. We chose the first option. There test data for two domains, the clause (C) and the noun phrase (NP).

In contrast to Quirk [Quirk *et al.*, 1991], complementation includes the subject (and also prepositions and obligatory adverbs).

4.1.1 C_Complementation

We drew up a list of all occurring frames in English and we wrote a set for each of these schemata. Meanings are not tested. The system to be tested succeeds if for each well formed item, it "consumes" precisely all specified complements, so all complements are obligatory. Optional complements must be handled by assigning two or more frames to a verb.

The standard word order assumed is in the examples is

subj - verb - ind obj - obj - prep obj - subclause

General Division of the Phenomenon

The subdivision of the English data is based on the following criteria (see [Hornby *et al.*, 1985; Svartvik, 1985; Quirk *et al.*, 1991; Alvey (4), ; Hammond *et al.*, 1988; Balkan *et al.*, 1991]):

1. the number of complements

2. function of complement

3. category/form of complement

(They are not always all applied.)

The top level classification looks like this:

- C_Complementation-Monovalent
- C_Complementation-Divalent
- C_Complementation-Trivalent
- C_Complementation-Tetravalent

There are two additional phenomena, cross classified with the classification just given. These data (and the classification) were taken from the HP test suite.

- C_Complementation-Equi
- C_Complementation-Raising

About the annotations, the following remarks need to be made:

The well-formed items have got a full analysis, the ill-formed none, except for some phenomena. In that case, the function of the ill-formed part is indicated by "?". WF=2 and 3 cases are considered as being ill-formed. In the phenomenon files there is a description for the ill-formed items, but only of the part that contains the error. There is no indication of attribute and value.

The expansion of the closed classes was generated semi-automatically, using the Auto_TSG and the lexical replacement tool (cf. [Arnold *et al.*, 1994]). It is mentioned in the classification where closed class expansion is applied.

One general presupposition, that has not been filled in in the phenomenon file, states that the constituents that are not analyzed further, are grammatically correct. The other presuppositions are filled in.

Ill-formed input only contains one error (relative to the positive item in the set). The ill formed items were generated in the following way:[1]

- variation of gender, number, case, category, "instance" (wrong preposition e.g.) if relevant. E.g.
  ** Him gives him the proposal.* (*Him* subject instead of *He*)
  ** She depends from her.* (*from* instead of *on*)

[1]There is no ill-formed item containing a wrong category for `verb-particle` and there is no added complement for the phenomena: `Trivalent-Direct_object_VP_bare_infinitive` and `Trivalent-Direct_object_VP_ing`.

The sets consisting of items lifted from the HP test suite were not expanded to fully systematic sets, nor were the sets of the phenomena `Trivalent-Direct_object_predicate(PP)`, `Trivalent-PP_complementizer(0)_indicative`, `Trivalent-PP_complementizer(WH)_VP_to_infinitive` and `Trivalent-Indirect_object_complementizer(WH)_VP_to_infinitive`.

– omission of a complement. E.g.
** He gives him.* (omitted object, for instance *the proposal*)

– addition of a complement (usually "it"). E.g.
** She carries it it.*

This causes some redundancy in the construction of ill formed data, especially because we kept generation of the ill formed items very systematic.

A systematic example:
** Carries him.*
** He carries.*
** He carries he.*
He carries him.
** He carries it him.*
** He carries to him.*
** Him carries him.*
** If he works carries him.*

The test sets are made to be self-contained, so a lot of tests (e.g. case of subject) and also input strings will be repeated (especially in the items ill-formed due to deletion). E.g.
She goes off it. → ** She goes it.*
She goes beyond it. → ** She goes it.*

Some remarks about the phenomenon naming:
Where **NP** occurs in the phenomenon names, pronouns are also allowed. The test items will in most of the cases contain pronouns (because these are the only complements that show case in English).
Where a phenomenon name ends in "...", the name must be completed by one of bold completions given in the description.

C_Complementation-Monovalent- ...

This phenomenon treats the intransitive verb frames.

There are two kinds of verbs in this class: with a dummy (or prop) subject, and with a "normal" subject. Two subclasses: **Subj(it)** and **Subj(NP)**.

It rains.
He comes.

C_Complementation-Divalent-Complementizer(that)_indicative-Subj(it)

This contains the verbs with a dummy subject *it* that replaces a *that*-clause in the subject position.

It seems that it is important.

C_Complementation-Divalent-Complementizer(0)_indicative-Subj(it)

Equivalent to the previous example, with the difference in realization of the complementizer: it is empty.

It seems it is important.

C_Complementation-Divalent-Predicate

This contains the copular verbs like *be* and *seem*. They need a subject and a predicate that tells something about the subject. This predicate can be a noun, an adjectival phrase or a adverbial or prepositional phrase.

She is a manager.
He is new.
She is in a good mood.

C_Complementation-Divalent-Direct_object

This is the "normal" case of a transitive verb with an object and a subject.

He carries it.

C_Complementation-Divalent-Complementizer(that)_ ...

In this phenomenon, a *that*-clause is the object. This phenomenon is divided into three subphenomena, each designated by the form of the verb of the subordinate clause:

- **indicative**
  She knows that it is important.

- **subjunctive/should_bare_infinitive**
  He requests that she leave.
  He requests that she should leave.

C_Complementation-Divalent-Complementizer(0)_indicative

This is entirely equivalent to the previous phenomenon with indicative, there is only no complementizer.

She hopes it is important.

C_Complementation-Divalent-Complementizer(WH)

Here there is a *WH*-word instead of the *that*- complementizer.

He knows who comes.

C_Complementation-Divalent-VP_ing

The object of the verbs in this phenomenon is an *-ing* form of a verb.

She enjoys working.

C_Complementation-Divalent-VP_to_infinitive

Here the object is a *to* infinitive.

He starts to work.

C_Complementation-Divalent-VP_bare_infinitive

The object is a bare infinitive.

She can come.

C_Complementation-Divalent-PP

These verbs are the prepositional verbs. This subphenomenon is expanded using prepositions as a closed class (based on [Hornby *et al.*, 1985]).

He depends on her.

C_Complementation-Divalent-Particle

These verbs are the phrasal verbs. Expanded (based on the Alvey_4 lexicon: a test set for each particle).

She reaches out.

C_Complementation-Trivalent-Particle_PP

This is a combination of the two previous phenomena: a prepositional verb with a verb particle (Phrasal-prepositional verbs).

She looks up to her.

C_Complementation-Trivalent-Indirect_object_direct_object

This case is the "classical" case of an indirect followed by a direct object.

He gives her the proposal.

C_Complementation-Trivalent-Indirect_object_PP

The direct object is replaced by a prepositional phrase.

She tells him about the news.

C_Complementation-Trivalent-Direct_object_PP

The PP can have any meaning and function, but quite frequently it will be a *to*-PP as a substitution for an indirect object.

He compares her with the manager.

C_Complementation-Trivalent-Indirect_object_complementizer(that)_indicative

The second object is a subordinate *that*-clause.

She convinces her that it works.

C_Complementation-Trivalent-Indirect_object_complementizer(0)_indicative

The second object is a subordinate clause with an empty complementizer.

He tells him it works.

C_Complementation-Trivalent-PP_PP

A verb with this frame has two PP's as complements.

She applies to her for advice.

C_Complementation-Trivalent-Direct_object_particle

The verb has a verb particle and needs a direct object. This phenomenon contains only items with a pronoun in the object function. The word order is more free when the direct object is an normal NP.

He sends her away.

C_Complementation-Trivalent-Direct_object_ ...

Here the verb has an object and a phrase that is the predicate of the object. There are three variants within this subphenomenon (depending on the category of the predicate):

- **predicate(PP)**
  She leaves the proposal in the office.
- **predicate(AP)**
  He finds the building good.
- **predicate(NP)**
  She considers him a competitor.

C_Complementation-Trivalent-Direct_object_VP_ed

Closely linked to the previous subphenomenon (and especially to the **AP** subclass), is the construction where a past participle occurs in the predicate position.

He finds the office closed.

C_Complementation-Trivalent-Direct_object_VP_ing

This frame has a direct object and a verb in the *-ing*-form.

She sees her leaving.

C_Complementation-Trivalent-Direct_object_as_ ...

This phenomenon is in the intersection between the prepositional verbs (as it always uses the preposition *as*) and the verb frames with predicates. Adjectival phrases (**AP**) and noun phrases (**NP**) can follow the preposition. No adverbial phrases or PP's, but the *-ing*-form of the verb is allowed (**VP_ing**).

He considers him as good.
She considers her as a competitor.
He considers him as being good.

C_Complementation-Trivalent-Direct_object_VP_bare_infinitive

The object of the main verb is the subject of the infinitive (expanded based on the list in Quirk).

She feels her leave.

C_Complementation-Trivalent-Indirect_object_VP_to_infinitive

The object of the main verb is the subject of the *to-* infinitive.

He tells him to leave.

C_Complementation-Trivalent-PP_VP_to_infinitive

The argument of the preposition is the subject of the *to-* infinitive.

She arranges with her to work.

C_Complementation-Trivalent-Indirect_object_complementizer(WH)_indicative

This verb frame consists of an indirect object and a *WH-* clause.

She asks him how he works.

C_Complementation-Trivalent-Indirect_object_complementizer(WH)_VP_to_infinitive

The verb has an indirect object and an verb phrase consisting of a *WH*-word and a *to*-infinitive.

She tells her how to work.

C_Complementation-Trivalent-Indirect_object_prep_VP_ing

The verb is followed by an indirect object and a PP that contains (as argument of the preposition) a verb in the *ing*-form.

He helps him in working.

C_Complementation-Trivalent-PP_complementizer(that)_indicative

This frame is analogous to the frame with indirect object and *that*-complementizer.

He explains to him that it works.

C_Complementation-Trivalent-PP_complementizer(0)_indicative

Equivalent to the previous one, but no complementizer.

She explains to her it works.

C_Complementation-Trivalent-PP_complementizer(WH)_VP_to_infinitive

The PP is followed by a *WH*-word and a *to*-infinitive.

She explains to her how to proceed.

C_Complementation-Trivalent-PP_complementizer(WH)_indicative

A PP followed by a *WH*-clause as non-subject complements.

He explains to him how it works.

C_Complementation-Tetravalent

This frame has four complements. As it is not very frequent, only one class is made.

She bets her one pound that she leaves.

C_Complementation-Equi-Obj_control

The object of the verb is (referentially) the same as subject of the infinitive.[2]

[2]It is assumed that the user is more or less familiar with the names "raising" and "equi", as these are not used in traditional grammars. See e.g. [Soames and Perlmutter, 1979].

Abrams urges Browne to hire Chiang.

C_Complementation-Equi-Pobj_control

The prepositional object of the verb is (referentially) the same as the subject of the infinitive.
Abrams appeals to Browne to hire Chiang.

C_Complementation-Equi-Subj_control

The subject of the verb is (referentially) the same as the subject of the infinitive.
Abrams manages to hire Browne.

C_Complementation-Raising-Subj_to_obj-VP_bare_infinitive

The subject of the infinitive is "the same" as the object of the verb (by "raising" the subject of the infinitive).
Abrams makes Browne hire Chiang.

C_Complementation-Raising-Subj_to_obj-VP_to_infinitive

The subject of the *to*-infinitive is "the same" as the object of the verb (by "raising" the subject of the infinitive).
Abrams causes Browne to hire Chiang.

C_Complementation-Raising-Subj_to_subj

The subject of the infinitive is "the same" as the subject of the verb (by "raising" the subject of the infinitive).
Abrams fails to hire Browne.

4.1.2 NP_Complementation

NP_Complementation is treated as one phenomenon without subdivisions. It is not possible to make the fine grained distinctions like in C_Complementation, because there are only two kinds of "objects": subordinate clauses and prepositions. Furthermore, the means to express functions are more limited than with verbs.

The test suite does not contain any examples of NP_Complementation with subclauses. There is however an exhaustive set of variations on a noun: *the donation*, chosen for its three possible complements. Since the complements of an NP are very often optional (and certainly when they are PP's), it was useful to perform all possible deletions and permutations of PP's. There are no additions, nor replacements. (See [Estival *et al.*, 1994b].)

The following list contains the simple cases of NP_Complementation of the word chosen, which are all combined in the test set for NP_Complementation:[3]

> the donation of the manager
>
> the donation by the manager
>
> the donation from the manager
>
> the donation of the building
>
> the donation to the company

The functions of the complements are indicated as prepositional objects (p-obj), not as subject etc.

Data

Complementation	
1048	test items
158	well-formed items
775	ill-formed items
16	dubious items
101	relatively ill-formed items

And more in detail:

C_Complementation	
1011	test items
148	well-formed items
757	ill-formed items
7	dubious items
99	relatively ill-formed items

NP_Complementation	
37	test items
10	well-formed items
18	ill-formed items
9	dubious items
2	relatively ill-formed items

[3]Without duplication of any functions, *i.e.* constructions like *the donation from the manager by the manager* are not considered.

4.2 French

By Complementation (which could also be called "subcategorization", or "valency" for verbs), we mean the set of arguments required by the head of a phrase: for a verb, these are its subject and complements, while most nouns and adjectives only subcategorize for complements.

The distinction between arguments and modifiers (see section 5) is that an argument is obligatory, while a modifier is optional.

Two important aspects of the phenomenon of Complementation are:

- the number of arguments which a head takes;
- the (morpho-)syntactic categories of the arguments.

The TSNLP test data is organized along these two lines.

4.2.1 General Division of the Phenomenon

Complementation affects elements which can act as heads of phrases: verbs, nouns, adjectives or prepositions. There is as yet no TSNLP data for preposition complementation in French. Thus, we only distinguish between the following:

- C_Complementation
- NP_Complementation
- AP_Complementation.

Further subdivisions of these three phenomena differ in a number of respects and are discussed individually below.

4.2.2 C_Complementation

Usually, verbs are first classified according to the number of complements they take (also called "valency"). For instance, the distinction between intransitive and transitive verbs is that a transitive verb takes both a subject and a direct object (valency = 2) while an intransitive verb only takes a subject (valency = 1).

As the French TSNLP data does not include verbs which subcategorize for four complements, we obtain the following subdivision:

- C_Complementation-Zerovalent
- C_Complementation-Monovalent
- C_Complementation-Divalent
- C_Complementation-Trivalent

The second subdivision deals with the morpho-syntactic category and the function of the arguments. To remain as theory-neutral as possible, both category and function are specified in the anotations of the test data. In the phenomenon name, the function of the argument is given first, with the category between parentheses:[4]

- C_Complementation-Zerovalent_subj(PRON-impers)_V
  Il pleut.
- C_Complementation-Monovalent_subj(PRON)_V
  Je viens.
- C_Complementation-Divalent_subj(NP)_V_obj(NP)
  L'ingénieur accepte l'invitation.)
- C_Complementation-Trivalent_subj(NP)_V_obj(NP)_p-obj(PP)
  (*L'ingénieur offre le livre à la femme.*)

As the argument of a verb may be of different categories, and can sometimes even have different functions, the TSNLP test data contains test items for the different possibilities in the four subphenomena (C_Complementation-Zerovalent, C_Complementation-Monovalent, C_Complementation-Divalent and C_Complementation-Trivalent). The actual realization of the arguments is given in the phenomenon-name, as shown above.

There are three ways of deriving ill-formed test items for Complementation, corresponding to the three operations of **replacement**, **deletion** and **addition** (see [Estival *et al.*, 1994b]).

1. most often, the category of the argument and/or its function is changed (see the category and function lists given in A).

   L'ingénieur accepte l'invitation.

function	category	instance
subj	NP	L'ingénieur
obj	NP	l'invitation

   ** L'ingénieur accepte de l'invitation.*

function	category	instance
subj	NP	L'ingénieur
p-obj(de)-NP	PP	de l'invitation

2. an argument is deleted:

   L'ingénieur accepte l'invitation.
   ** L'ingénieur accepte.*

3. more rarely, an argument is added:

   L'ingénieur vient.
   ** L'ingénieur vient l'entreprise.*

   This is often more difficult, because inserted PP arguments might also be taken as modifiers, and the test item is then well-formed.

[4]Note that the function of the verb ("functor") is not given.

An argument such as NP, PP, etc. can often be replaced with a pronoun, which doubles the number of possible realizations (e.g. four for a divalent verb). This is not tested exhaustively in the TSNLP test data for French.

- C_Complementation-Divalent_subj(NP)_V_obj(NP)
  L'ingénieur accepte l'invitation.
- C_Complementation-Divalent_subj(PRON)_V_obj(NP)
  Il accepte l'invitation.
- C_Complementation-Divalent_subj(NP)_obj(PRON)_V
  L'ingénieur l'accepte.
- C_Complementation-Divalent_subj(PRON)_obj(PRON)_V
  Il l'accepte.

However, since pronouns belong to a closed class, they are exhaustively tested in C_Complementation.

The following sections present an exhaustive list of the different cases covered by the TSNLP data (each phenomenon name is followed by some examples).

C_Complementation-Zerovalent

- C_Complementation-Zerovalent_subj(PRON-impers)_V
  Il pleut.
  ** L'entreprise pleut.*

C_Complementation-Monovalent

- C_Complementation-Monovalent_subj(NP)_V
  L'ingénieur vient.
  ** A l'ingénieur vient.*
- C_Complementation-Monovalent_subj(PRON)_V
  Je viens.
  ** Me viens.*

C_Complementation-Divalent

- C_Complementation-Divalent_subj(NP)_V_obj(NP)
  L'ingénieur accepte l'invitation.
  ** L'ingénieur accepte de l'invitation.*
- C_Complementation-Divalent_subj(PRON)_obj(PRON)_V
  Je l'accepte.
  ** Je lui accepte.*
- C_Complementation-Divalent_subj(PRON-impers)_V_obj(NP)
  Il faut un ingénieur.
  ** Il faut pour un ingénieur.*

- C_Complementation-Divalent_subj(PRON-impers)_obj(PRON)_V
  Il le faut.
  ** Il lui faut.*

- C_Complementation-Divalent_subj(NP)_V_p-obj(PP)[5]
  L'ordinateur provient de Suisse.
  ** L'ordinateur provient la Suisse.*

- C_Complementation-Divalent_subj(PRON)_p-obj(PRON)_V
  Il en provient.
  ** Il le provient.*

- C_Complementation-Divalent_subj(NP)_V_s-comp(C-cpl)
  L'entreprise suppose que l'ingénieur vient.
  ** L'entreprise suppose de l'ingénieur.*

- C_Complementation-Divalent_subj(PRON)_obj(PRON-neuter)_V
  6z*Je le suppose.*
  ** Je lui suppose.*

- C_Complementation-Divalent_subj(NP)_V_v-comp(VP-inf)
  L'ingénieur pense diriger l'entreprise.
  ** L'ingénieur pense de diriger l'entreprise.*

- C_Complementation-Divalent_subj(NP)_V_p-obj(inf(VP)-inf)
  L'ingénieur commence à diriger l'entreprise.
  ** L'ingénieur commence diriger l'entreprise.*

- C_Complementation-Divalent_subj(NP)_v-comp(PRON-neuter)_V
  L'ingénieur le sait.
  ** L'ingénieur te sait.*

- C_Complementation-Divalent_subj(NP)_V-copula_subj-pred(attribute)[6]
  L'entreprise est un échec.
  ** L'entreprise est à un échec.*

  Le professeur apparaît comme un spécialiste.
  ** Le professeur apparaît pour un spécialiste.*

- C_Complementation-Divalent_subj(NP)_subj-pred(PRON-neuter)_V-copula
  Le professeur l'est.
  ** Le professeur m' est.*

- C_Complementation-Divalent_subj(PRON-impers)_V_s-comp(C-cpl)
  Il faut que l'ingénieur vienne.
  ** Il faut si l'ingénieur vient.*

[5] This is tested for the two prepositions *à* and *de*.

[6] Although French copula verbs do not form a closed class in TSNLP, the most important ones are tested here.

- C_Complementation-Divalent_subj(PRON-impers)_s-comp(PRON-neuter)_V
  Il le faut.
  ** Il lui faut.*

- C_Complementation-Divalent_subj(PRON-impers)_V_v-comp(VP-inf)
  Il faut diriger l'entreprise.
  ** Il faut de diriger l'entreprise.*

- C_Complementation-Divalent_subj(PRON-impers)_v-comp(PRON-neuter)_V
  Il le faut.
  ** Il lui faut.*

- C_Complementation-Divalent_subj(PRON-impers)_V_subj-pred(ADJ)
  Il fait beau.
  ** L'entreprise fait beau.*

- C_Complementation-Divalent_subj(PRON-impers)_V_subj-pred(attribute)
  Il est arrivé une catastrophe.
  ** Il est arrivé à une catastrophe.*

- C_Complementation-Divalent_subj(PRON-impers)_subj-pred(PRON)_V[7]
  ** Il l'est arrivé.*

C_Complementation-Trivalent

- C_Complementation-Trivalent_subj(NP)_V_obj(NP)_p-obj(PP)
  L'ingénieur offre le livre à la femme.
  ** L'ingénieur offre le livre la femme*

- C_Complementation-Trivalent_subj(NP)_obj(PRON)_V_p-obj(PP)
  L'ingénieur le présente à l'entreprise.
  ** L'ingénieur lui présente à l'entreprise.*

- C_Complementation-Trivalent_subj(NP)_p-obj(PRON)_V_obj(NP)
  L'ingénieur lui présente la femme.
  ** L'ingénieur le présente la femme.*

- C_Complementation-Trivalent_subj(NP)_obj(PRON)_p-obj(PRON)_V
  L'ingénieur le leur présente.
  ** L'ingénieur te me présente.*

- C_Complementation-Trivalent_subj(NP)_V_obj(NP)_s-comp(C-cpl)
  L'entreprise convainc le conseil que l'homme doit venir.
  ** L'entreprise convainc au conseil que l'homme doit venir.*

- C_Complementation-Trivalent_subj(NP)_obj(PRON)_V_s-comp(C-cpl)
  L'entreprise l'avertit si le conseil accepte ou non.
  ** L'entreprise lui avertit si le conseil accepte ou non.*

[7] There is no corresponding well-formed test item.

- C_Complementation-Trivalent_subj(NP)_s-comp(PRON)_V_obj(NP)
  L'entreprise en convainc l'homme.
  ** L'entreprise me convainc l'homme.*

- C_Complementation-Trivalent_subj(NP)_obj(PRON)_s-comp(PRON)_V
  L'entreprise l'en convainc.
  ** L'entreprise lui en convainc.*

- C_Complementation-Trivalent_subj(NP)_V_p-obj(PP)_s-comp(C-cpl)
  Le conseil dit à l'entreprise que l'homme doit venir.
  ** Le conseil dit l'entreprise que l'homme doit venir.*

- C_Complementation-Trivalent_subj(NP)_p-obj(PRON)_V_s-comp(C-cpl)
  Le conseil lui demande si l'homme vient.
  ** Le conseil le demande si l'homme vient.*

- C_Complementation-Trivalent_subj(NP)_s-comp(PRON)_V_p-obj(PP)
  Le conseil le demande à l'entreprise.
  ** Le conseil lui demande à l'entreprise.*

- C_Complementation-Trivalent_subj(NP)_s-comp(PRON)_p-obj(PRON)_V
  Le conseil le lui demande.
  ** Le conseil le le demande.*

4.2.3 NP_Complementation

Unlike C_Complementation, NP_Complementation is not divided according to the valency of the noun. The function of the arguments is not given either, but only the category of the argument(s) is mentioned.

NP_Complementation is subdivided in the following way:

- NP_Complementation_NP_PP
  La cause de la rupture
  ** La cause à la rupture*

- NP_Complementation_NP_PP-inf
  La crainte de combattre
  ** La crainte à combattre*

- NP_Complementation_NP_C-cpl(que)
  L'espoir que la femme vienne
  ** L'espoir si la femme vienne*

- NP_Complementation_NP_PP1_PP2
  L'invitation de l'ingénieur à la négociation
  ** L'invitation à l'ingénieur à la négociation*

- NP_Complementation_NP_PP_PP-inf
  La décision de l'ingénieur de venir.
  ** La décision à l'ingénieur de venir*

As in C_Complementation, the derivation of ill-formed data is based on the **replacement** and **deletion** operations (see [Estival *et al.*, 1994b]).

Since it is not always possible for an application to test for NPs in isolation, the data for NP_Complementation exists in two versions:

- as isolated NPs:
  the value of the attribute CATEGORY is "NP".
- as the subject of a sentence with an intransitive verb:
  the value of the attribute CATEGORY is "NP(augmented to S)".

4.2.4 AP_Complementation

As with NP_Complementation, the division of AP_Complementation is based on the category of the arguments.

So far, test items only exist as APs given in isolation, but these will be turned into test items which are complete sentences by using the AP as a predicate with a copular verb.

- AP_Complementation_ADJ_PP
  jaloux de l'ingénieur
  ** jaloux à l'ingénieur*
- AP_Complementation_ADJ_C-cpl(que)
  heureux que l'ingénieur vienne
  ** heureux si l'ingénieur vienne*
- AP_Complementation_ADJ_PP-inf
  difficile à accepter
  ** difficile pour comprendre*
- AP_Complementation_ADJ_PP_PP-inf
  difficile pour l'entreprise de comprendre
  ** difficile de l'entreprise de comprendre*

4.2.5 Parameters

Attribute	Value	instance
subcat	X(Y1,Y2,..,Yn) X: number of complements Y: function of the complement	
comp	function of the complement[8]	
EXAMPLE		
L'ingénieur accepte l'invitation.		
subcat	2(subj, obj)	
comp	subj	l'ingénieur
comp	obj	l'invitation
** L'ingénieur accepte de l'invitation.*		
comp	obj	* de l'invitation

4.2.6 Data

Total number of test items for Complementation:

Complementation	
827	test items
209	well-formed items
606	ill-formed items
8	dubious items
4	relative ill-formed items

More specifically:

- C_Complementation:

C_Complementation	
755	test items
188	well-formed items
565	ill-formed items
2	dubious items

- NP_Complementation (including augmented test items):

NP_Complementation	
40	test items
12	well-formed items
19	ill-formed items
6	dubious items
A 3	relatively ill-formed items

- AP_Complementation:

AP_Complementation	
32	test items
9	well-formed items
22	ill-formed items
1	relatively ill-formed items

4.3 German

4.3.1 C_Complementation

The phenomenon C_Complementation - in standard grammar [Helbig and Buscha, 1988; Drosdowski, 1984] known as *verbal government* or *verbal valency* - deals with verbal subcategorization frames. Verbs subcategorise for *obligatory* and *optional* complements. While the deletion of obligatory elements results in ungrammatical sentences the omission of optional elements doesn't effect the grammaticality. In contrast to adjuncts which are syntactically independent of the verb the form of the complements is determined by the verb.

Der Manager beschuldigt den Präsidenten des Diebstahls.
Der Manager beschuldigt den Präsidenten.
** Der Manager beschuldigt des Diebstahls.*
** Der Manager beschuldigt den Präsidenten dem Diebstahl.*

General Division of the Phenomenon

Because TSNLP data focuses on syntactic properties the aim was to compile a comprehensive list of combinations of *obligatory* complements of verbs, forming the basis of different sentence patterns. Obligatory complements can be tested by an operationalizable criterion, a specific sort of right extraposition:

Der Manager beschuldigt den Präsidenten, und zwar des Diebstahls.
** Der Manager beschuldigt des Diebstahls, und zwar den Präsidenten.*

It was tried to find instances of all possible combinations of nominal, prepositional, adjectival but also sentential and infinitival complements. The result of the collection is a list of about 130 verb frames [Colliander, 1983; Engel, 1970; Engel and Schumacher, 1976; Goetze, 1979; Helbig and Schenkel, 1980; Weisgerber, 1983]. The sentences illustrate:

- combinations of nominal, prepositional and adjectival complements
  - verbframes with only nominal complements
    Der Manager bedarf des Präsidenten.
    Der Manager gibt dem Präsidenten den Test.
  - verbframes with prepositional phrases
    Among the prepositional phrases two classes can be distinguished: *prepositional objects* and *prepositional complements.* They differ in the status of their prepositions. In the first case the verb determines the prepositions which are semantically empty but serve as functional element to assign the case to the governed nominal phrase.
    Der Manager bittet um den Test.
    ** Der Manager bittet auf den Test.*
    In prepositional complements the preposition together with the case of the depending nominal phrase express *location, direction, time and modality.*
    Der Manager wohnt in Saarbrücken.
    Der Manager eilt in den Saal.

Der Vorfall spielt während des Tests.
Der Manager wohnt in guten Verhältnissen.

- verbframes with predicative complements
  Der Manager heißt Peter.
  Der Manager wird krank.
  Der Manager schimpft den Präsidenten einen Idioten.
- verbframes with adjectival complements
  Der Manager wohnt gut.
  Der Test dauert lange.

• Sentential Complements

- verbframes with *daß*-clauses
  Daß der Manager schläft, stimmt.
  Daß der Manager schläft, bedarf des Beweises.
  Der Manager glaubt, daß der Präsident schläft.
- verbframes with *ob*-clauses
  Ob der Manager schläft , entgeht dem Präsidenten.
  Der Manager weiß, ob der Präsident schläft.
- verbframes with *"wh"*-clauses
  Wie der Manager schläft, ärgert den Präsidenten.
  Der Manager weiß, wie der Präsident schläft.

• Infinitival Complements

- verbframes with bare infinitives
  Der Manager läßt schlafen.
  Der Manager kann schlafen.
- verbframes with *zu*-infinitives
  Der Manager glaubt, den Präsidenten zu ärgern.
  Den Präsidenten zu ärgern, stimmt den Manager nachdenklich.

In describing the complement structure of the sentences various labels are used in the literature.[9] According to the analysis annotated to the sentences the complements are identified both by their function and by their morpho-syntactic category.[10]

An important characteristic of verbal government is the number of subcategorised complements in a verb frame. The subdivision of the phenomenon follows the three types:

• Monovalent Verbs

  Only five different verbframes were found for verbs which take one complement.[11] This group contains no ungrammatical sentences.

[9] See Engel/Schumacher 1976 for an overview of the notion of different authors.
[10] The abbreviations correspond to the labels used in the comparative function and category lists of TSNLP.
[11] Sentences of a verb and a subject of a NP, a *that* clause or a *zu*-infinitive; sentences built of a verb plus direct or indirect object.

Der Manager schläft.
Den Präsidenten friert.

- Divalent Verbs

  Divalent verbs subcategorise for two complements.

  Der Manager begrüßt den Präsidenten.
  ** Der Manager begrüßt.*

  Der Manager stimmt gegen den Präsidenten.
  ** Der Manager stimmt.*

  Der Manager gilt als fähig.
  ** Der Manager gilt als fähig.*

  Der Manager wagt zu programmieren.
  ** Der Manager wagt.*

- Trivalent Verbs

  Verbs of this class require three complements to form grammatical sentences.

  Der Manager lehrt den Präsidenten den Test.
  ** Der Manager lehrt den Präsidenten.*
  ** Der Manager lehrt den Test.*

  Der Manager hält den Test für gelungen.
  ** Der Manager hält den Test.*
  ** Der Manager hält für gelungen.*

  Der Manager zwingt den Präsidenten dazu, daß der Präsident schläft.
  ** Der Manager zwingt den Präsidenten.*
  ** Der Manager zwingt dazu, daß der Präsident schläft.*

Parameters

- Construction of ill-formed items

  From the relevant operations *Replacement, Deletion, Addition* only all possible omissions — except the subject — are carried out to create ungrammatical sentences.

- Annotations

  All test strings are annotated with two attributes describing the number (*n_complements*) and form (*verb frame*) of the subcategorised complements. For the ungrammatical example the remaining verbframe is given.

  The following sentences belong to the annotations in the following table:

  Der Manager schläft.
  Der Manager hört auf den Präsidenten.
  Der Manager bezichtigt den Präsidenten des Diebstahls.

  * Der Manager bezichtigt.
  Der Manager bezichtigt den Präsidenten.
  Der Manager bezichtigt des Diebstahls.

Attribute	Value
n_complements	1/2/3
frame	function(category)
GRAMMATICAL EXAMPLES	
n_complements	frame
1	subj(NP)
2	subj(NP) p-obj(auf)(PP)
3	subj(NP) obj(NP) iobj(NP_gen)
UNGRAMMATICAL EXAMPLES	
n_complements	frame
1	subj(NP)
2	subj(NP) obj(NP)
2	subj(NP) iobj(NP_gen)

Data

The total number of test items for C_Complementation is:[12]

C_Complementation	
464	test items
218	well-formed items
219	ill-formed items
27	dubious items

[12] 25 test item are dubious in the context of the phenomenon.

Chapter 5

Modification

5.1 English

5.1.1 NP_Modification

Following Quirk et al. [Quirk *et al.*, 1991], we recognise three component parts in a complex NP:

- the head
- premodification
- postmodification

The head is the central component around which the other components cluster, and dictates agreement with the rest of the sentence outside the NP. Thus in:

the old man who is standing in the corner

man is the head. Premodification comprises all the items placed before the head, including determiners, adjectives and nouns. In this example, *The old* are the premodifying elements. Postmodification comprises all the items placed after the head, including prepositional phrases, non-finite clauses, and relative clauses. The relative clause *who is standing in the corner* is thus an example of postmodification.

Modification can be restrictive or non-restrictive. In restrictive modification, the modifier forms part of the description that serves to pick out the referent of the head of the clause. In non-restrictive modification, on the other hand, the information provided by the modification is additional but not essential to the identification of the referent.

Contrast, for example the following two sentences:

(6) a. *The old man who is standing in the corner is waiting for his friend.* (restrictive)

b. *The old man, who is standing in the corner, is waiting for his friend.* (non-restrictive)

While (6 a.) picks out a certain old man, namely the one who is standing in the corner, in (6 b.) the identity of the old man does not depend on whether or not he is standing in the corner.

General Division of the Phenomenon

The data concentrates on a particular kind of postmodification, namely Relative Clauses.

There are two broad categories of relative clauses in English:

- Restrictive
- Non-restrictive

The difference between these two types of modification is described in the preceding section.

The relative pronoun can be:

- a wh-pronoun (e.g. *who*)
- *that*
- zero pronoun

The choice of relative pronoun depends on three factors:

- whether the clause is restrictive or non-restrictive (see above);
- the gender type of the antecedent:

  Uniquely for relative clauses, a two-term "gender" system, **personal** and **nonpersonal** is distinguished. Personality is ascribed basically to human beings but extends in some cases to pet animals and creatures in the supernatural world that can be thought of as having human characteristics:

  the manager who
  ** the manager which*

  Collective nouns are normally treated as personal when they have plural agreement and as non-personal when they have singular concord:

  the committee who/which

- the function of the relative pronoun as subject, object, etc or as a determiner (*whose*):

  the manager who comes
  ** the manager whom comes*

The data covers core cases of restrictive and non restrictive relative clauses. Relative clauses that contain categories other than relative pronouns (e.g. *the place WHERE he succeeds*). These are not covered by the data.

Sentential subjects (e.g. *He succeeds, which pleases the manager.*) are also not dealt with.

The set of relative pronouns (*who, which, whom, whose, that*) is covered exhaustively. Ill-formed items are generated systematically by placing wrong overt relative pronouns in the place of correct ones.

Restrictive relative clauses are analysed at NP level, non-restrictive at sentence level (because of the importance of punctuation in the latter case).

Discontinuous constituents such as stranded prepositions are analysed using a slash category (PP_slash-NP) as follows:

that	*he*	*approves*	*of*
|	|	|	|
PRON_rel	PRON	V	PP_slash-NP
p-obj	subj	func	p-obj
2:3	2:3	0:4	2:3

NP_Modification-Relative_clauses-Restrictive

- **Subject_extraction**
  - **Personal**
    The manager who succeeds
    ** The manager which succeeds*
  - **Nonpersonal**
    The firm which succeeds
    ** The firm who succeeds*
  - **Collectives**
    The committee which succeeds
    The committee who succeeds

  Only the personal forms are shown in the examples below:

- **Object_extraction**
  The manager whom he sees
  ** The manager who he sees*

- **PP_extraction**
  Core prepositions are treated exhaustively in this part of the data, and indicated in the Phenomenon name. For example:

  PP_extraction-Personal-(of)
  The manager of whom he approves
  ** The manager of that he approves*

- **PP_extraction(stranded)**
  These are cases with a stranded preposition:

  The manager whom he approves of
  ** The manager which he approves of*

- **Genitive**
  The firm whose manager succeeds

NP_Modification-Relative_clauses-Non_restrictive

- **Subject_extraction**
  The manager, who succeeds, succeeds.
  ** The manager, whom succeeds, succeeds.*

- **Object_extraction**
  The manager, whom he sees, succeeds.
  ** The manager, which he sees, succeeds.*

- **PP_extraction**
  Core prepositions are treated exhaustively in this part of the data.

  The manager, of whom he approves, succeeds.
  ** The manager, of which he approves, succeeds.*

- **PP_extraction(stranded)**
  The manager, whom he approves of, succeeds.
  ** The manager, which he approves of, succeeds.*

- **Genitive**
  The manager, whose advice he needs, succeeds.

Data

The total number of test items for NP_Modification is:

NP_Modification	
782	test items
300	well-formed items
482	ill-formed items
0	dubious items
0	relatively ill-formed items

5.2 French

Modification is a complex phenomenon which is not only syntactic in nature, but also has a large semantic component. We broadly characterize it as the addition of linguistic material (words or phrases) which is not directly required by any other element in the sentence, but which serves to qualify or modify what is expressed by other elements of the sentence.

It would be difficult to exhaustively describe or test all the types of modification, because almost any phrase can be modified, and it can be modified in a number of different ways. However, following the general guidelines for classification of phenomena, we can first determine the domains in which modification can occur, and for each domain the ways in which modifying elements can be expressed and their ordering with respect to the other elements.

Since TSNLP is mostly concerned with testable characteristics of the phenomena, we have chosen to concentrate on the morpho-syntactic characteristics of modification and largely ignore the semantic aspect of modification.

From a morpho-syntactic point of view, there are three main aspects which can be tested in all the domains concerned:

- the number of modifiers allowed;
- the (morpho-)syntactic categories of the possible modifiers;
- the various orderings of the modifiers.[1]

5.2.1 General Division of the Phenomenon

Although almost any linguistic element can be modified in some way, because the place of attachment of a modifier can only be tested when a syntactic representation in the form of a parse tree or labelled bracketing is produced by the system, the TSNLP test data for Modification only concerns three major domains: clause (C), noun phrase (NP) and adjective phrase (AP).

Indeed, the distinction between sentential modification (S) and clausal modification (C) can only be tested by looking at the attachment of the modifier in the syntactic tree (which could only be done when testing parsers or NLP systems which produce a parse representation), or by verifying scope relations (which could only be done if there was a semantic representation giving scope relations). Thus the distinction between sentential modification and clausal modification is not reflected in the TSNLP test data and all instances of sentential modification are given at the level of the clause (C).

Similarly, the distinction between clausal modification and verbal modification is also very hard to draw. For instance, in the sentence "*Il vient aujourd'hui.*", the adverb *aujourd'hui* could be attached either at the level of the clause (C: *il vient*) or to the verb (V: *vient*). Since this aspect of modification is difficult to represent in test items, and as it is not clear that representing it would actually help much in testing NLP systems, the distinction between

[1] Although the ordering of the modifiers is closely related to the question of their scope, since the notion of scope itself requires a semantic analysis and TSNLP does not at this stage provide a semantic level, the TSNLP test data for modification does not include any test item or annotation explicitly aimed at the interpretation of scope.

clause modification and VP modification is not present in the current TSNLP data.[2] However, one of the most important aspects of verbal modification, the realization of negation and the placement of the negative particle, is given under the phenomenon of Negation.

While it is possible to have modifiers in an adverbial phrase (ADVP) or propositional phrase (PP), these are taken as being less central and they are not covered in the TSNLP test data.[3]

Thus, we only distinguish between the following (∗ indicates the phenomena for which test data already exists).

- ∗ C_Modification
- NP_Modification
- AP_Modification.

For each of these, two further aspects can be tested, namely the **category** of the modifiers and the **word order** between the modifiers and the other elements in the domain. These two aspects are discussed in more detail in section 5.2.2, because the only test data for modification in French is currently that for C_Modification.

5.2.2 C_Modification

In order to reach a reasonable level of complexity, the test items used for C_Modification are based on the positive test items for the different types of verbs in C_Complementation. Starting with intransitive verbs, modifiers are successively added, resulting in sentences with one modifier, two modifiers, etc. The operation is repeated with 2-valency verbs (which take a direct object or a prepositional object), and 3-valency verbs (which take both a direct object and a prepositional indirect object).

An important characteristic of clausal modification is the **number** of modifiers which are allowed in a domain and this is the basis for the main subdivision of this phenomenon in the TSNLP test data.[4]

[2] The information that a given modifier could be taken as modifying either the whole clause or only the VP is nevertheless provided. For instance, the adverb in *Il vient aujourd'hui.* can be treated the following analysis in which the first value (*0:2*) of the attribute "domain" states that the adverb modifies the entire clause (*Il vient*), and the second value (*1:2*) states that it can also modify the verb (*vient*).

position	2:3
instance	*aujourd'hui*
category	ADV
function	mod-temp
domain	0:2, 1:2

[3] Examples would be:

- ADVP_Modification:
  Il a répondu [très correctement].
- PP_Modification:
  Il était [complètement derrière la machine].

[4] Again, ∗ indicates the phenomena for which test data already exists.

- * C_Modification_1-mod
- * C_Modification_2-mod
- C_Modification_3-mod
- etc...

As with all types of modification, there are two aspects of C_Modification which can be tested. One is the possible **category** of clausal modifiers: adverbs, prepositional phrases and (some) nominal phrases. The other concerns the **word order** among the elements of the clause.

The actual TSNLP data includes test items for the following subphenomena (note that no category distinction is specified in the phenomenon name when the test item contains more than one modifier).

- C_Modification_1-mod_category
  - C_Modification_1-mod_category(ADV)
  - C_Modification_1-mod_category(PP)
  - C_Modification_1-mod_category(NP)
- C_Modification_1-mod_Wordorder
- C_Modification_2-mod_category
- C_Modification_2-mod_Wordorder

C_Modification_1-mod

C_Modification_1-mod_category

In French, several categories can act as clausal modifiers: adverbs (ADV), prepositional phrases (PP) and (some) nominal phrases (NP).

C_Modification_1-mod_category(ADV)

Adverbs can be formed from adjectives by adding the suffix *-ment*. This morphological relation may be a source of error, and we thus derive ill-formed test items by substituting an adjective (ADJ) for an adverb (ADV).

L'ingénieur travaille rapidement.
** L'ingénieur travaille rapide.*

The data also contains non-derived adverbs and adverbial phrases.

C_Modification_1-mod_category(PP)

Ill-formed test items could theoretically be formed by changing the category PP to that of its complement (NP). But since, in some cases, this might correspond to another correct adjunct (see below), there are no ill-formed test items in the test data for this subphenomenon.

L'ingénieur vient à midi.

C_Modification_1-mod_category(NP)

A restricted set of NPs may be used as clausal modifiers. Unlike in English (where NP clausal modifiers can also be locative, e.g. *home*) NP clausal modifiers in French must be temporal.

L'ingénieur vient le soir.
** L'ingénieur vient la maison.*

Whether an NP is temporal is part of semantics and can not be tested unless the systems under test contain semantic information. Where that is the case, these temporal NPs can be replaced with non-temporal NPs to create ill-formed test data.

As the current TSNLP test data is only syntactic, the NP could in principle be replaced by a PP containing the same NP. However, as with C_Modification_1-mod_category(PP), this derivation would often generate items which can be well-formed, e.g.:

L'ingénieur vient l'hiver.
L'ingénieur vient en hiver.

Thus, there are no ill-formed test items in the test data for this subphenomenon.

C_Modification_1-mod_Wordorder

The actual position of the modifier is often a source of error. For clausal modifiers, this position is usually either after the verb or at the beginning of the clause. In the latter case, a comma is necessary. Thus, punctuation is also tested by ill-formed test items not containing the comma.

L'ingénieur travaille rapidement.
** L'ingénieur rapidement travaille.*

Rapidement, l'ingénieur travaille.
** Rapidement l'ingénieur travaille.*

As the possible positions for modifiers change in the presence of auxiliaries, the TSNLP data contains examples in both present tense and passé composé.

The supertypes of C_Modification_Wordorder are both C_Modification and C_Wordorder.

C_Modification_2-mod

As can be observed when working with corpora of texts, the number of possible modifiers in a clause is an extremely important aspect of modification and often causes problems to NLP systems.

The test data for multiple modification does not include coordinated modifiers, as this should be part of the phenomenon of Coordination.

When there is more than one modifier, the data also test for both their **category** and their **word order**.

C_Modification_2-mod_category

Although when there is more than one modifier, the actual names of the sub-phenomena do not include the different categories (thus, we do *not* have sub-phenomenon names such as C_Modification_2-mod_category(ADVP+PP) or C_Modification_2-mod_category(NP+PP)), the category of the different modifiers is systematically varied in the test data.

As far as possible, the data covers all possible combinations. For instance, since there are three different categories which can be used as clausal modifiers, there are six possible combinations for a clause with two modifiers.

L'ingénieur travaille rapidement à la maison. (ADV - PP)
L'ingénieur travaille rapidement le soir. (ADV - NP)
L'ingénieur travaille le soir à la maison. (NP - PP)
Aujourd'hui, l'ingénieur travaille rapidement. (ADV - ADV)
L'ingénieur travaille à la maison avec sa femme. (PP - PP)
L'hiver, l'ingénieur travaille le soir. (NP - PP)

C_Modification_2-mod_Wordorder

For the position of the modifiers, the starting data are the same inputs as above and the position of the modifiers is varied systematically.

The position of the two modifiers with respect to each other is tested, even though this might arguably depend on the semantics of the adverb, because it is a very important problem and also because purely surface oriented test items can give information about the kind of semantics which is mastered by a system.

** L'ingénieur travaille à la maison rapidement.* (order of the modifiers with respect to each other)
A la maison, l'ingénieur travaille rapidement.
** Rapidement, l'ingénieur travaille à la maison.*

5.2.3 NP_Modification

Although the data for this phenomeon, which would be called NP_Modification, has not been created at this point, we can indicate how it should be treated.

Testing for NP Modification presupposes that the following phenomena have already been tested:

- NP_Wordorder
- NP_Agreement
- NP_Complementation
- C_Rel[5]

As with C_Modification, the two aspects to be checked are the **category** of the modifiers and their **order** inside the NP. Unlike with C_Modification, the number of modifiers is not taken as the primary division for this phenomenon.

[5] The phenomenon C_Rel is not treated in the TSNLP data.

NP_Modification_category

NP_Modification_category(AP)

la voiture rapide
** la voiture rapidement*

NP_Modification_category(PP)

la voiture aux roues neuves
** la roues neuves voiture*

NP_Modification_category(C_rel)

l'usine que l'on a fermée
l'usine qui ferme
** l'usine que la direction a fermé l'atelier* (not a C-rel)

NP_Modification_Wordorder

NP_Modification_Wordorder(AP)

In French, adjectives can either precede or follow the N they modify.

la grande voiture
** la voiture grande*
la grande voiture rouge
** la grande rouge voiture*
** la voiture grande rouge*

NP_Modification_Wordorder(PP)

In French, a PP modifying an NP always follows the N.

la voiture aux roues neuves
** la aux roues neuves voiture*

NP_Modification_Wordorder(C_rel)

In French, a relative clause always follows the N it modifies.

l'usine que l'on a fermée
** la que l'on a fermée usine*

5.2.4 AP_Modification

Although the data for this phenomeon, which would be called AP_Modification, has not been created at this point, we can indicate how it should be treated.

Testing for NP Modification presupposes that the following phenomena have already been tested:

- AP_Wordorder
- AP_Complementation

As with C_Modification and NP_Modification, the two aspects to be checked are the **category** of the modifiers and their **order** inside the AP.

AP_Modification_category

Only adverbs (ADV) can modify adjectives.

L'usine est très productive.
** L'usine est grande productive.* (A)
** L'usine est productive de ses ouvriers.* (PP)

NOTE: For the latter example, AP_Complementation must have been checked, since a PP can be the complement of some adjectives:

Il est fier de ses enfants.

AP_Modification_Wordorder

The adverb must always precede the adjective.

La voiture est très grande.
** La voiture est grande très.*

5.2.5 Parameters

The TSNLP test data for C_Modification is underspecified with respect to the parameters, thus the following parameters are *not* in the database. However, the following documentation is intended to indicate how they should be set for this phenomenon.

C_Modification_1-mod_category

WF	Attribute	Value	input
1	cat	ADV	L'ingénieur travaille rapidement.
0	cat	ADV	L'ingénieur travaille rapide.
1	cat	PP	L'ingénieur vient à midi.
1	cat	NP	L'ingénieur vient ce soir.

C_Modification_1-mod_Wordorder

WF	Attribute	Value	input
1	position_mod	post_verb	L'ingénieur travaille rapidement.
1	position_mod	start_sentence	Rapidement, l'ingénieur travaille.
0	position_mod	(ante_verb;start_sentence)	L'ingénieur rapidement travaille.
1	punct	PUNCT(,)	Rapidement, l'ingénieur travaille.
1	punct	PUNCT(,)	Rapidement l'ingénieur travaille.

5.2.6 Data

The total number of test items for C_Modification is:

C_Modification	
392	test items
329	well-formed items
50	ill-formed items
2	dubious items
11	relatively ill-formed items

5.3 German

5.3.1 NP_Modification

The phenomenon NP_Modification deals with regularities concerning optional constituents clustered around the head of noun phrases. According to the position relative to the noun *premodification* and *postmodification* can be distinguished.

General Division of the Phenomenon

Three aspects are relevant in NP_Modification. Firstly, the modifying elements are classified according to their morpho-syntactic category. Secondly, interaction with word order principles is an important criterion affecting the grammaticality of the test phrases. The last point deals with the interaction of modifiers occurring in the same noun phrase.

The data is grouped into the following subphenomena:

- NP_Modification-Premodification
- NP_Modification-Postmodification
- NP_Modification-Complex

The first two classes reflect order principles of single modifiers while the last one illustrates what combinations of pre- and postmodifiers are allowed inside nominal phrases. The morpho-syntactic categories are used as labels for the modifying elements.

NP_Modification-Premodification

This data set comprises noun phrases with modifiers placed before the head noun. Elements of the following categories can occupy this position.

- adjectives

  Noun phrases with two adjectives in positive form as well as examples with one adjective in comparative and superlative form are included.

  der neue wichtige Test
  **der neue Test wichtige*
  **der Test neue wichtige*

  wichtigere Tests
  **Tests wichtigere*

- adjective phrases

  Besides adjectives modified by adverbs this data set includes an example with an adjective extended by a prepositional phrase.

  der besonders wichtige Test
  **der Test besonders wichtige*

der für alle Beteiligten wichtigste Test
**der Test für alle Beteiligten wichtigste*

- particles

  Grading particles such as *nur, auch, genug* can only modify noun phrases.

  genug Geld
  **genug Test*
  genug Tests
  ?Tests genug

- genitive NPs

  Genitive NPs can occur in both the position before and after the head noun.

  Peters Test
  der Test Peters

  In contrast to all other modifiers the omission of the genitive noun phrase in front of a singular head noun results in an ungrammatical phrase since this element has the function of a possessive determiner.

  des Managers Test
  **Test*
  **der des Managers Test*
  **der Managers Test*
  **Managers Test*

  As the following examples show there is a difference in the acceptance of noun phrases according to the genitive marking.

  des Managers Test
  ?des Präsidenten Test
  des Präsidenten wichtigster Test

- verbal participles

  This data set covers four different subcategories represented as inflected verbal forms and as inflected verbal phrases.

  V_pcp-bare-infl:

  der lesende Manager
  **der Manager lesende*

  V_pcp-zu-infl(passive):

  der mitzuteilende Bericht
  **der Bericht mitzuteilende*

  VP_pcp-perf-infl(passive):

  der dem Manager gegebene Bericht
  **der Bericht dem Manager gegebene*

  VP_pcp-perf-infl(sein):

  der soeben angekommene Bericht
  **der Bericht soeben angekommene*

- relative clauses

  In so called *pseudo cleft constructions* the sentence starts with a relative clause followed by the head noun. Since these constructions contain an emphasized head noun only definite or demonstrative determiners are possible specifiers.

  Den der Berater kennt, der Manager leitet das Projekt.
  Den der Berater kennt, dieser Manager leitet das Projekt.
  **Den der Berater kennt, ein Manager* leitet das Projekt.
  **Die der Berater kennt, Manager* leiten das Projekt.

  Besides, this form of left extraposition can also occur in parentheticals.

  Gestern angekommen sind *die der Berater kennt — die Kinder*.

NP_Modification-Postmodification

In this subphenomenon elements placed after the head noun are covered. Modifiers in post-position include:

- genitive NPs

  As mentioned before nominal phrases with genitive NPs placed before or after the head noun are grammatical.

  das Projekt Berlins
  Berlins Projekt

  der Test Mutters
  Mutters Test
  der Test der Mutter
  der Mutter Test

  Additionally, this data set comprises ungrammatical noun phrases resulting from non-genitive modifiers. In the last example an explicit genitive marker is missing.

  der Test des Präsidenten
  **der Test der Präsident*
  **der Test dem Präsidenten*
  **der Test den Präsidenten*

  **der Test Präsidenten*

- prepositional phrases

  In general, prepositional phrases must follow the head noun.

  der Test von gestern
  **von gestern der Test*

  der Bericht in dem Aktenkoffer
  **in dem Aktenkoffer der Bericht*

  To emphasize on the prepositional phrase and on the superlative a special ordering is possible. As the ungrammatical examples show, the noun phrase must contain the comparative or superlative form of an adjective.

für alle Beteiligten der wichtigere Test
für alle Beteiligten der wichtigste Test
**für alle Beteiligten der wichtige Test*
**für alle Beteiligten der Test*

- relative clauses

  This modification group contains four segments. Additionally to position principles the first data set also reflects agreement requirements between the head noun and the relative pronoun of the relative clause.

  derjenige der Manager, der einen Computer hat
  **der einen Computer hat, derjenige der Manager*

  **derjenige der Manager, die einen Computer haben*

  In the second set the complete paradigm of the singular and plural relative pronouns of all three genders is covered.

  der Manager, der kommt
  die Präsidentin, derer der Berater sich erinnert
  die Kinder, die der Berater kennt

  Relative clauses beginning with a relative pronoun plus preposition constitute the third data group.

  das Projekt, über das der Manager sich freut
  das Institut, von dem der Manager kommt

  Finally relative clauses containing contractions of relative pronouns and prepositions are covered.

  das Projekt, woran der Manager sich erinnert
  der Tag, da der Manager kommt
  die Art, wie der Manager geht

- adverbs

  Nominal structures where an adverb follows the head noun are restricted to a few cases.

  das Buch dort
  ?dort das Buch

  In the following example it is supposed that the adverb modifies the whole noun phrase.

  einige Tests mehr
  **einige mehr Tests*
  **mehr einige Tests*

- adjective phrases

  As post nominal parentheses uninflected adjective phrases can occupy the position after the head noun.

  der Test, wichtig für den Präsidenten
  **der wichtig für den Präsidenten Test*
  **wichtig für den Präsidenten der Test*

- adjectives

  The construction with an uninflected adjective placed behind the head noun is very rare and mostly restricted to lyrical texts.

  Röslein rot
  **rot Röslein*

NP_Modification-Complex

To cover complex modification exhaustively, all possible combinations of modifiers have to be constructed systematically. A first step towards this aim is a data segment covering the full range of combinations with two different modification types. To exemplify the procedure, a data set with a genitive NP as postmodifier combined with representative examples of all other pre- and postmodification types has been worked out.[6]

- adjective(s) and adjective phrases

  der neue, wichtige Test des Managers
  der für alle Beteiligten wichtigste Test des Managers

  der Test des Managers, wichtig für den Präsidenten
  ?der Test, wichtig für den Präsidenten, des Managers[7]

- adverb

  The postmodifying genitive NP must be placed before the adverb.

  der Test des Managers morgen
  **der Test morgen des Managers*

  einige Tests des Managers mehr
  **einige Tests mehr des Managers*

- particle

  nur Tests des Managers

- genitive NP

  The semantical interpretation influences the grammaticality of these phrases.

  Peters Besuch Berlins
  Peters Test des Projekts
  **Peters Test des Managers*

  des Managers Test des Projekts
  ?des Managers Test des Präsidenten

  ?der Test Peters des Managers[8]

  In the following examples it is not possible to attach the two genitive NPs to the head noun.

[6] Since the construction *Röslein rot* is very restricted it is not included into modification combinations.
[7] It is not possible to attach the genitive NP to the head noun; but the noun phrase is grammatical if the genitive NP refers to *Präsidenten*.
[8] If *des Managers* is an apposition referring to *Peters* the phrase is correct.

?der Test der Mutter des Managers
?der Test des Managers der Mutter

- prepositional phrase

  The postmodifying genitive NP must be placed before the prepositional phrase.

  der Test des Managers von gestern
  **der Test von gestern des Managers*

  In the second test string of the following example pair, the genitive NP modifies *Tisch* and its scope cannot be extended to the head noun.

  der Bericht des Managers auf dem Tisch
  ?der Bericht auf dem Tisch des Managers

  If the prepositional phrase is placed before the head noun it has no influence on the attached genitive NP.

  für alle Beteiligten der wichtigste Test des Managers

- verbal participle

  Since all verbal participles are premodifiers the addition of a postmodifying genitive NP results in grammatical strings.

  der dem Präsidenten mitzuteilende Bericht des Managers
  der soeben angekommene Bericht des Managers

- relative clause

  The genitive NP follows the head noun immediately and cannot be placed after the relative clause.

  der Manager des Projekts, der kommt
  **der Manager, der kommt, des Projekts*

  den der Berater kennt, den Manager des Projekts

Parameters

- Construction of ill-formed items

  In general, the ungrammatical examples for the first two subphenomena result from ordering variations.[9] The ungrammatical phrases of complex modification are caused by combination patterns which are not possible inside nominal groups.

- Annotations

  All test strings are annotated with two attributes describing the position of the modifiers (*modifier type*) and reflecting the categorial surface of the noun phrase (*nominal word order*).

[9] Additionally there are a few ill-formed string concerning specialities of modifying genitive NPs and premodifying relative clauses.

Attribute	Value
modifier position	pre, post, pre & post
nominal word order	category
GRAMMATICAL EXAMPLES	
modifier position	nominal word order
pre	DET ADJ N
post	DET N NP_gen PP
complex	DET VP_pcp-zu-infl(passive) N NP_gen

For the ill-formed phrases an error attribute describing the source of ungrammaticality is added.

Data

The total number of test items for NP_Modification is:[10]

NP_Modification	
198	test items
131	well-formed items
55	ill-formed items
12	dubious

[10] 7 test strings are dubious and 5 are ill-formed in context.

Chapter 6

Agreement

6.1 English

6.1.1 C_Agreement

Agreement can be broadly defined as the relationship between two grammatical elements such that if one of them contains a particular feature (e.g. plurality) then the other also has to have that feature.

In English, there is agreement in person and number between the subject and verb of a sentence. This shows up most clearly with pronouns and the verb *be*:

I am
** I are*

In general, a singular subject requires a singular verb:

The manager succeeds.
** The manager succeed.*

and a plural subject requires a plural verb:

The managers succeed.
** The managers succeeds.*

Following Quirk et al. [Quirk *et al.*, 1991], we call this generalisation the rule of Grammatical Concord. Occasionally there are conflicts between this rule and two other principles; the principle of Notional Concord and the principle of Proximity. Notional Concord, again following Quirk et al., is the agreement of verb and subject according to the *idea* of number, rather than the actual presence of the grammatical marker for that idea. Collective nouns often show Notional Concord:

The committee succeed/s.

The principle of Proximity (following Quirk) denotes agreement of the verb with a closely preceding noun phrase in preference to agreement with the head of the noun phrase that functions as subject. Partitives may illustrate this phenomenon:

A lot of people succeed.
** A lot of people succeeds.*

General Division of the Phenomenon

The following are covered by the data:

- Pron_V(be)
- Pron_V
- NP_V
- Collectives
- Relative_clauses
- Partitives
- Coordinated_subj(correlatives)
- Parentheticals

All positive items get a categorial analysis, but not all get a functional analysis. No parameters are given for ill-formed items, only comments and an indication of the position of error, where appropriate.

C_Agreement-Pron_V(be)

The full set of personal pronouns is given in the data: *I*, *you*, *he*, *she*, *it*, *we*, *they*.

I am new.
* *I is new.*

C_Agreement-Pron_V

I succeed.
* *I succeeds.*

NP_V

The manager succeeds.
* *The manager succeed.*

C_Agreement-Collectives

As mentioned above, collectives frequently show Notional Concord.

The committee succeed/s.

C_Agreement-Relative_clauses

In relative clauses, the number of the verb depends on the interpretation of the number of the relative pronoun (*who*, *which*, etc), which in turn depends on the number of the noun phrase it is modifying:

The manager who succeeds comes.
** The manager who succeed comes.*

C_Agreement-Partitives

As mentioned above, partitives illustrate the Principle of Proximity.

A lot of them are new.
** A lot of them is new.*

C_Agreement-Coordinated_subj(correlatives)

Coordinated subjects pose a particular problem, especially where correlatives are concerned. With *either ... or*, where each member of the coordination has a different number, the principle of Proximity is generally applied:

Either he or they succeed.
** Either he or they succeeds.*

Neither ... nor may either behave like *either ...or*, or in less formal usage, like *and*:

Neither he nor he succeed/s.

C_Agreement-Parentheticals

In parenthetical expressions the material in parenthesis is usually disregarded for purposes of agreement:

An employee, and sometimes the manager, succeeds.
** An employee, and sometimes the manager, succeed.*

Data

The total number of test items for C_Agreement is:

C_Agreement	
123	test items
68	well-formed items
50	ill-formed items
5	dubious items
0	relatively ill-formed items

6.1.2 NP_Agreement

In English, cooccurrence restrictions hold between a determiner and a noun. Three categories of noun are important:

- singular count (*firm*)
- plural count (*firms*)
- mass (*money, advice*)

Quirk et al. [Quirk *et al.*, 1991] classifies determiners into three groups, on the basis of their position in the NP in relation to each other:

- Predeterminers:
  These precede determiners and are mutually exclusive.
- Central_determiners
- Postdeterminers:
  These may follow determiners.

General Division of the Phenomenon

The data covers the following:

- Predeterminer
- Central_determiner
- Postdeterminer
- Central_determiner_Central_determiner combinations
- Predeterminer_Predeterminer combinations
- Predeterminer_Central_determiner combinations

Possessives, articles, demonstratives and the central set of quantifiers (*(a) few, (a) little, many, much, several*, etc) are exhaustively treated.

Where data is of the form DET1 DET2 N there is a presupposition that there is agreement between DET2 and N.

All items get a full **analysis**, but only ill-formed items get **parameters**. For the parameters at most comment, position and instance of error are filled in.

The data was produced partially using the test suite generation tool.

NP_Agreement-Predeterminers

All_both_half

all firms
** all firm*

Fractions

** half firm*

Multipliers

** double firm*

Such_what

such firms
** such firm*

NP_Agreement-Central_determiners

Class_1

Class_1 determiners are followed by singular count, plural count, or mass nouns:

- **Definite article**
  the firm
- **Interrogative_determiner**
  which firm
- **Negative_determiner** *no firm*
- **Possessives**
  his firm
- **Relative_determiner**
  whose firm
- **Some_and_any**
  some firm
- **Wh-determiner_in_ever**
  whatever firm

Class_2

Class_2 determiners are followed by plural count nouns and mass nouns:

- **Enough**
  enough money
  ** enough firm*

- **Zero_article**
  money
  ** firms*

Class_3

Class_3 determiners are followed by singular count nouns and mass nouns and include (Demonstratives):

this firm
** this firms*

Class_4

Class_4 determiners are followed by plural count nouns (Demonstratives):

these firms
** these firm*

Class 5

Class_5 determiners are followed by singular count nouns:

- **Neither**
  neither firm
  ** neither firms*

- **Either**
  either firm
  ** either firms*

- **Indefinite_article**
  a firm
  ** a firms*

- **Each_and_ every**
  each firm
  ** each firms*

NP_Agreement-Postdeterminers

Ordinals

- **Ordinal_numbers**:

  first firm?

- **General_ordinals**:

  another firm
  ** another firms*

Cardinals

one firm
** one firms*

Quantifiers

- **Closed_class-Class_1**
  Class_1 quantifiers cooccur with plural count nouns:

  several firms
  ** several firm*

- **Closed_class-Class_2**
  Class_2 quantifiers cooccur with mass nouns:

  a little money
  ** a little firm*

- **Closed_class-Class_3**
  Class_3 quantifiers cooccur with mass nouns or plural count nouns:

  more firms
  ** more firm*

- **Open_class-Class_1**
  Class_1 quantifiers cooccur with mass nouns or plural count nouns:

  plenty of firms
  ** plenty of firm*

- **Open_class-Class_2**
  Class_2 quantifiers cooccur with mass nouns:

  a great deal of money
  ** a great deal of firm*

- **Open_class-Class_3**:
  Class_3 quantifiers cooccur with plural count nouns:

a great number of firms
** a great number of firm*

NP_Agreement-Central_determiner_Central_determiner

Central determiners are mutually exclusive. An exception is that possessives may cooccur with *every*:

his many firms
** the this firm*

NP_Agreement-Predeterminer_Predeterminer

As noted above, predeterminers are mutually exclusive. An exception is *such* which may be preceded by *all*:

all such firms
** both all firms*

NP_Agreement-Predeterminer_Central_determiner

Only certain combinations of predeterminer and central determiners are permitted.

both the firms
** both another firms*

Data

The total number of test items for NP_Agreement is:

NP_Agreement	
1196	test items
201	well-formed items
928	ill-formed items
65	dubious items
2	relatively ill-formed items

6.2 French

Two elements x and y agree if they have at least one feature in common. In the following example, the subject and the finite verb are in an agreement relation: they must have the same features for person and number.

L'ingénieur vient.

	L' ingénieur	vient
pers-num	3-sg	3-sg

In French, three features are mainly involved for an agreement relation: person, number and gender.[1]

The TSNLP data for Agreement test for these features. Their possible values of are shown in the following table:[2]

attribute	value
pers	(1,2,3)
num	(sg, pl)
gen	(masc,fem)
case	(nom,acc)

To derive ill-formed test items for Agreement, the value of one of these attributes is replaced, as illustrated in the following examples:

L'ingénieur vient.
** L'ingénieur viens.* (Person is wrong)

L'ingénieur vient.
** L'ingénieur viennent.* (Number is wrong)

Le travail
** La travail* (Gender is wrong)

6.2.1 General Division of the Phenomenon

As the phenomenon of Agreement manifests itself in two domains, the clause and the NP, there are two subphenomena:

- C_Agreement
- NP_Agreement

The further subdivision of these two phenomena differs in certain aspects and they are discussed individually in the following sections.

[1] They are abbreviated as PERS, NUM, GEN.
[2] The attribute CASE is needed for the relative pronoun in NP_Agreement.

6.2.2 C_Agreement

C_Agreement mainly concerns what is traditionally called Subject-Verb Agreement. This is covered by five subphenomena:

- C_Agreement_subj(PRON)_V
- C_Agreement_subj(NP)_V
- C_Agreement_subj(NP-quant)_V
- C_Agreement_subj(PRON-coord)_V
- C_Agreement_subj(PRON-impers)_V_subj-pred(PRON)

An important agreement problem in French is that of the past participle, and this is covered by the last subphenomenon:

- C_Agreement_ppast

C_Agreement_subj(PRON)_V

The input consists of a pronoun and an intransitive verb. Since pronouns belong to a closed class, they are exhaustively tested. As for the verb, it is tested exhaustively in the "présent", "passé composé", "imparfait", "futur simple" and "conditional" forms.

Je viens.
** Je vient.*

C_Agreement_subj(NP)_V

L'ingénieur vient.
** L'ingénieur viens.*

The input consists of an NP subject and an intransitive verb. By default, the NP is masculine, but some test items also contain feminine NPs. Both singular and plural NPs are covered. The verb is tested for "présent", "passé composé", "imparfait" and "futur simple" forms.

C_Agreement_subj(NP-quant)_V

Some (plural) quantifiers permit only syntactic agreement, e.g. *deux*:

Deux litres de lait suffisent.
** Deux litres de lait suffit.*

Other quantifiers, which are actually singular NPs such as *la plupart*, allow both syntactic (i.e. singular) and semantic (i.e. plural) agreement:

La plupart des ordinateurs fonctionnent.
La plupart des ordinateurs fonctionne.

The input for these test items is based on the corresponding test items for NP_Agreement.

C_Agreement_subj(PRON-coord)_V

In French, when two (tonic) subject pronouns which are not in the same person are coordinated, the agreement of the verb depends on the particular person features involved: 1st person has priority, followed by 2nd person.[3]

Toi et moi venons.
** Toi et moi venez.*
Toi et lui venez.

C_Agreement_subj(PRON-impers)_V_subj-pred(PRON)

In French, the verb usually agrees with the impersonal subject pronoun *ce* (i.e. 3rd person singular) but if a predicate NP (which can be a pronoun) is in the 3rd person plural, the verb should agree with it.

C'est moi.
** Ce suis moi.*

Ce sont eux.
** C'est eux.*[4]

The input for this test data consists of an impersonal pronoun followed by the verb *être* and a predicative (tonic) pronoun.

C_Agreement_ppast

For the agreement of the past participle (which is extremely important in French grammar), the TSNLP data follows Grevisse (1991) [Grevisse, 1991] (§904–916).

Le discours que l'ingénieur a prononcé est excellent.
** Le discours que l'ingénieur a prononcée est excellent.*

6.2.3 NP_Agreement

NP Agreement in French involves a correspondence in number and gender between the head noun and the determiners, adjectives, quantifiers and other modifiers inside the NP, as well as with certain pronouns.

It also includes agreement in number and gender between the head noun and the relative pronoun in relative clauses (which are considered NP modifiers).

Since some applications are not able to test for isolated NPs, the data for NP_Agreement exists in two versions:

- as isolated NPs:
  the value of the attribute CATEGORY is "NP".

[3] C_Agreement_subj(PRON-coord)_V interacts with Coordination, and its supertypes are C_Agreement and NP_Coordination.

[4] This is only acceptable in the spoken language. In this case, the item is considered to be well-formed and the value of the attribute REGISTER is "informal".

- as the NP object in a sentence with an direct transitive verb:
  the value of the attribute CATEGORY is "NP(augmented to S)".[5]

NP_Agreement_DET_N

The data for NP Agreement tests exhaustively for both masculine and feminine nouns. Since determiners belong to a closed class, the TSNLP data also tests exhaustively for them.[6]

Le travail.
** La travail.*

NP_Agreement_NP-quant_P-det_N

Different kinds of quantified NPs are tested.

La plupart des questions.
** La plupart de la question.*

NP_Agreement_DET_N_ADJ

The input consists of a determiner and a noun followed by an adjective. Both masculine and feminine nouns are tested, in the singular and in the plural.

Le travail habituel.
** Le travail habituelle.*

The determiners are not tested exhaustively, since that should have been done in NP_Agreement_DET_N. This is also the case for all the phenomena given below.

NP_Agreement_DET_N_AP-sup

The input is the same as for NP_Agreement_DET_N_ADJ, except that the adjective is in the superlative.

Le travail le plus habituel.
** Le travail la plus habituelle.*

NP_Agreement_DET_N_AP-coord

The input is the same as for NP_Agreement_DET_N_ADJ, except that the noun is followed by coordinated adjectives.[7]

Le travail habituel et intelligent.
** Le travail habituelle et intelligente.*

---

[5] If the NP were the subject, subject-verb agreement would be problematic for ill-formed test items.
[6] See the list of French determiners in A.2.2.
[7] As NP_Agreement_DET_N_AP-coord interacts with Coordination, its supertypes are C_Agreement and AP_Coordination.

NP_Agreement_DET_ADJ_N

The input consists of a determiner followed by an adjective preceding the head noun. Both masculine and feminine nouns are tested, in the singular and in the plural.

Le bon travail.
** Le bonne travail.*

NP_Agreement_DET_AP-sup_N

The input consists of a determiner followed by a superlative adjective preceding the head noun.

Le plus beau livre.
** Le plus belle livre.*

NP_Agreement_DET_AP-coord_N

The input is the same as for NP_Agreement_DET_ADJ_N, except that the noun is preceded by coordinated adjectives.[8]

La grande et belle maison.
** La grand et beau maison.*

NP_Agreement_NP_C-rel(PRON-rel+V)

The data tests for the agreement in number and gender between the head noun and the relative pronoun in a relative clause (which is considered as an NP modifier). The relative clause consists of a relative pronoun and an intransitive verb.

La question qui provoque.
** La question que provoque.*

NP_Agreement_NP_C-rel(PRON-rel+V-copula+ADJ)

The input is the same as for NP_Agreement_C-rel(PRON-rel+V), except that the relative clause consists of a relative pronoun, a copular verb and an adjective.

La question qui est intelligente.
** La question qui sont intelligente.*

NP_Agreement_DET-quant_P-det_N

Since we consider quantified determiners such as *beaucoup de, trop de* etc. to belong to a closed class, they are exhaustively tested.

Un peu de vin.
** Un peu de vins.*

[8] As NP_Agreement_DET_AP-coord_N interacts with Coordination, its supertypes are C_Agreement and AP_Coordination.

NP_Agreement_DET1_DET2

Nouns can also be preceded by two determiners:

Tout le vin.
** Toute le vin.*

6.2.4 Parameters

Attribute	Value	instance
case-pers-num-gen	(nom,acc)-(1,2,3)-(sg, pl)-(masc,fem)	
EXAMPLES		
L'ingénieur vient.		
pers-num pers-num	3-sg 3-sg	l' ingénieur vient
** L'ingénieur viens.*		
pers-num	3-sg	vient

6.2.5 Data

The total number of test items for Agreement is (augmented test items are included):

Agreement	
1641	test items
376	well-formed items
1265	ill-formed items

More specifically:

- C_Agreement is covered in the database by:

C_Agreement	
287	test items
104	well-formed items
183	ill-formed items

- NP_Agreement is covered in the database by (augmented test items are indicated between parentheses):

NP_Agreement	
677 (677)	test items
136 (136)	well-formed items
541 (541)	ill-formed items

6.3 German

6.3.1 C_Agreement

The phenomenon C_Agreement — in standard grammar[Helbig and Buscha, 1988; Drosdowski, 1984] known as *subject verb agreement* — deals with the agreement between the subject and the finite verb in the morpho-syntactic features *person* and *number*. In general, the values of *person* and *number* for the subject and the verb must be the same (**grammatical agreement**).

Der Manager schläft.
** Der Manager schläfst.*
** Der Manager schlafen.*

Two other principles are responsible for exceptions to the rule of grammatical agreement: **proximity** and **semantic agreement**. Proximity means that the finite verb refers only to the closest subject element of complex subjects and shows agreement with that part.

Sowohl der Manager als auch der Berater schlafen.
Sowohl der Manager als auch der Berater schläft.

Semantic agreement denotes agreement between the finite verb and the subject on the basis of semantic interpretation.

Ein Dutzend Präsidenten schläft.
Ein Dutzend Präsidenten schlafen.

In German grammatical agreement predominates.

General Division of the Phenomenon

The sentences are built according to a classification of various subject forms mainly based on the Duden grammar [Drosdowski, 1984]. The verbform is always present tense, indicative. The following subject classes result form variations of the subject position.[9]

- Pronouns

  The complete paradigm of the personal pronouns with all existing verb forms is covered in the data because personal pronouns in subject position show *person* agreement with the verb most clearly.

  Wir schlafen.
  ** Wir schlafe.*
  ** Wir schläfst.*
  ** Wir schläft.*
  ** Wir schlaft.*

  Furthermore various singular demonstrative and possessive pronouns as well as indefinite pronouns in singular and plural fill the subject position.

  Diese schläft.
  Meines schläft.

[9] The data contain at least one example of each subject type.

Jeder schläft.
Alle schlafen.

Ein jemand schläft.

- Simple Noun Phrases

  Besides simple determiner plus countnoun subjects mass nouns and bare plurals are included to demonstrate the different number requirements.

  Der Manager schläft.
  Die Manager schalfen.
  ** Die Manager schläft.*

  Wein ist teuer.
  ** Wein sind teuer.*

  Manager schlafen.
  ** Manager schläft.*

- Interrogative Pronouns

  A third person singular verb is required for interrogative pronouns in subject position.

  Wer schläft?
  ** Wer schlafen?*

- Copula Constructions

  In copula constructions agreement shows which element the subject is. If there is no form difference between subject and predicative two analyses are possible.

  Ich bin der Gewinner.
  ** Ich ist der Gewinner.*

  Der Manager ist der Präsident.

- Impersonal Constructions

  Sentences without a personal subject are always third person singular.

  Dem Präsidenten ist schlecht.
  Den Präsidenten ist schlecht.
  ** Den Präsidenten sind schlecht.*

- *Es* in subject position

  In passive sentences with *es* in subject position the auxiliary must be third person singular.

  Es wird geschlafen.
  ** Es werden geschlafen.*

  If *es* has only topic function or serves as correlative it has no influence on agreement features.

  Es werden die Tests gemacht.
  ** Es wird die Tests gemacht.*

  Es gefällt dem Manager zu schlafen.

- Singular Partitives

  The attributive plural genitive has no influence on number agreement.

  Keiner der Manager schläft.
  ** Keiner der Manager schlafen.*

- NPs with postmodifying PPs

  Prepositional phrases in plural have no influence on number agreement.

  Die Frage in den Tests ist leicht.
  ** Die Frage in den Tests sind leicht.*

- NPs with relative clauses

  In general, the relative clause has no influence on the agreement between noun and finite verb.

  Einer dieser Manager, die einen Computer haben, schläft.
  ** Einer dieser Manager, die einen Computer haben, schlafen.*

  Personal pronouns modified by relative clauses agree with the finite verb only if the pronoun is repeated in the relative clause. Otherwise the finite verb is third person.

  Ich, die ich müde bin, schlafe.
  ** Ich, die müde ist, schlafe.*
  Ich, die müde ist, schläft.

  Free relative clauses are always singular.

  Wer den Computer hat, schläft.
  ** Wer den Computer hat, schlafen.*

- Measure Constructions, Appositions, Comparative Phrases

  Many measure constructions, appositions and comparative phrases show semantic agreement but some cases require grammatical agreement.

  Eine Handvoll Päsidenten schläft.
  Eine Handvoll Päsidenten schlafen.

  Ein Kilogramm Zucker genügt.
  ** Ein Kilogramm Zucker genügen.*

  Beides, Forschung und Lehre, ist wichtig.
  Beides, Forschung und Lehre, sind wichtig.

  Nichts als leere Versprechungen steht im Vertrag.
  Nichts als leere Versprechungen stehen im Vertrag.

- Coordinated Structures with *und*

  Subjects coordinated with *und* generally require plural.

  Der Manager und der Berater schlafen.
  ** Der Manager und der Berater schläft.*

  Sometimes alternative semantic interpretations allow semantic agreement.

Groß und klein schlafen.
Groß und klein schläft. (single concept)

Agreement based on proximity is possible to emphasize one subject part.

Der Bedarf und das Angebot steigen.
Der Bedarf und das Angebot steigt.

Hyphen compounds usually need a singular verb.

Die Vor- und Nachbereitung des Tests ist gut.
?Die Vor- und Nachbereitung des Tests sind gut.

Coordinations of the indefinite pronoun *jed-* require a singular verb.

Jeder und jede schläft.
?Jeder und jede schlafen.

Coordinated personal pronouns cause problems on person agreement. In general, first person dominates second and second dominates third person but contradictory forms also result in grammatical sentences. Often the collective plural pronoun is inserted to clarify the agreement relation.

Ich und du schlafen.
** Ich und du schläft.*

?Du und er schläft.
Du und er schlafen.
Du und er, ihr schlaft.

- Coordinated Structures with *oder*

  Since disjunctive conjunctions exclude one subject element the finite verb is mostly singluar.

  Der Manager oder der Präsident schläft.
  ?Der Manager oder der Präsident schlafen.

- Coordinated Structures with *nicht...sondern, sowie, sowohl...als auch, weder...noch*

  Exclusive expressions need singular verbs while inclusive expressions denote plural and thus prefer plural verbs.

  Nicht der Manager, sondern der Berater schläft.
  ** Nicht der Manager, sondern der Berater schlafen.*

  Sowohl der Manager als auch der Berater schlafen.
  Sowohl der Manager als auch der Berater schläft.

- Subject Infinitives and Subject Clauses

  Subject infinitives and subject clauses require a singular verb even if they are coordinated with *und*.

  Zu schlafen gefällt dem Manager.
  Daß der Manager kommt, stimmt.

  Zu schlafen und zu testen gefällt dem Manager.
  ** Zu schlafen und zu testen gefallen dem Manager.*

Parameters

- Construction of ill-formed items

  The ungrammatical sentences result from subject and verb variations on person (only for personal pronouns) and number.

- Annotations

  The grammatical test strings are annotated with the attribute *number* and only sentences with personal pronouns as subject have the additional attribute *person*. Since all other subjects are third person this value is taken as the default. Information on the subject form is found in a comment field.

Attribute	Value	Instance	Comment
person	1/2/3	subject/verb	subject form
number	sg/pl	subject/verb	subject form
		EXAMPLE	
person	number	instance	comment
1	pl	Wir	Personal Pronouns
1	pl	schlafen	
	sg	Eine Handvoll Präsidenten	Measure Construction
	pl	schlafen	

  The ungrammatical sentences are only annotated with an error attribute and an error code. In most cases it's impossible to judge the source of the error, e.g. in:

  Der Manager schläft.
  Die Manager schalfen.
  ** Die Manager schläft.*
  ** Der Manager schalfen.*

Data

The total number of test items for C_Agreement is:[10]

C_Agreement	
399	test items
224	well-formed items
154	ill-formed items
21	dubious items

6.3.2 NP_Agreement

The German nominal group — even when reduced to the core inventory of nouns, determiners and attributive adjectives — is a morphologically and syntactically complex structure. The interaction of agreement features and declension classes causes various different forms of noun phrases.

[10] One test item is dubious in the context of the phenomenon.

The phenomenon NP_Agreement deals with the cooccurrence of determiners, adjectives and nouns in noun phrases. All three elements must agree in *case*, *number* and *gender*.

General Division of the Phenomenon

The subclassification of NP_Agreement is motivated by the elements occurring in the noun phrase. The following four subclasses are distinguished:

- (adjective plus) noun
- determiner plus noun
- determiner, adjective plus noun
- (determiner), adjectives plus noun

NP_Agreement-(ADJ)-N

This data set comprises three different types of nouns both in isolation and in combination with an adjective.

- singular count nouns

  All the singular forms of *Mann, Projekt, Wand* plus these forms combined with all forms of the inflectional paradigm of *alt* are covered. Because singular count nouns obligatorily have to combine with a determiner, all test items of this group are ungrammatical.

  * *Mann* (der Mann)
  * *Mannes* (des Mannes)

  * *alter Mann*
  * *alte Mann*
  * *altes Mann*

  The bare noun forms are linked to one grammatical noun phrase containing a definite determiner. The adjective-noun patterns are grouped into sets with the same noun form.

- bare plurals

  Analogously to the single count nouns, the systematic variation of plural nouns resulted in all possible forms of *Mann, Projekt, Wand* and the combination of all of these forms with the complete inflectional paradigm of *alt*.

  Männer
  Männern

  alte Männer
  alter Männer

  The noun phrases modified by *alt* are grouped into sets according to the form of the head noun.

alte Wände
** altes Wände*
alter Wände
** alten Wände*
** altem Wände*

- mass nouns

  A mass noun of each gender (*Abfall, Getreide, Milch*) is chosen to represent this group.

  Getreide
  viel Getreide
  ** viele Getreide*

  To compare mass nouns and count nouns the following examples are grouped into test sets.

  viel Abfall
  ** viele Abfall*

  ** viel Männer*
  viele Männer

The ungrammatical items result from the systematic variations over the adjective and the noun forms.

The number of test items for NP_Agreement-(ADJ)-N is:

NP_Agreement-(ADJ)-N	
98	test items
28	well-formed items
66	ill-formed items
4	dubious items

NP_Agreement-DET-N

In this subphenomenon nouns are specified by different determiners. The first part of the data contains the complete inflectional paradigm of the nouns *Mann, Projekt* and *Wand* systematically combined with all inflectional forms of a definite (*der*), a demonstrative (*dieser*) and two indefinite (*ein, kein*) determiners.

eine Wand
einer Wand
** eines Wand ...*

die Männer
der Männer
** das Männer ...*

keines Projekts
** kein Projekts ...*

The sets comprise patterns where the noun form is constant while the determiner forms are varied.

For example, one set consists of:

der Mann
** die Mann*
** das Mann*
** des Mann*
den Mann
dem Mann

The ill-formed items are constructed by the systematic variation over determiner and noun forms.

To cover a broad range of determiners, noun phrases built with various determiner types (definite, indefinite, demonstrative, possessive) and the noun *Projekt* are added to the data. Since this subset focuses on determiner variation only grammatical example items in nominative case are constructed.

dasjenige Projekt
alle Projekte
beide Projekte
jenes Projekt
unsere Projekte

The number of test items for NP_Agreement-DET-N is:

NP_Agreement-DET-N	
335	test items
111	well-formed items
224	ill-formed items
0	dubious items

NP_Agreement-DET-ADJ-N

Determiners, adjectives and nouns form the noun phrases of this subclass. Parallel to the determiner-noun data the inflectional paradigm of a representative data subset is exhaustively tested. Additionally to the previous subphenomenon test strings combine the complete paradigms of the determiner and noun forms with all possible forms of the adjective *alt* (*alte,altes,alter,alten,altem*).

der alte Mann
** der alter Mann*
** der altes Mann*
** der alten Mann*
** der altem Mann*

Noun phrases with the same noun but different determiner and adjective forms are grouped together. The ill-formed items result from the systematic variation over determiner, adjective and noun forms.

The second part of the data contains examples where the adjective *alt* is inserted into the nominative noun phrases built of various determiner types.

dasselbe alte Projekt

irgendein altes Projekt
etliche alte Projekte

The ungrammatical examples result from wrong adjective inflection. If the determiner requires strong inflection the ill-formed item has weak inflection and vice versa.

das alte Projekt
** das altes Projekt*

ein altes Projekt
** ein alte Projekt*

Some indefinite determiners allow weak and strong inflection.

einige alte Projekte
einige alten Projekte

The number of test items for NP_Agreement-DET-ADJ-N is:

NP_Agreement-DET-ADJ-N	
1553	test items
122	well-formed items
1428	ill-formed items
3	dubious items

NP_Agreement-(DET)-ADJs-N

The data covers examples with two and three adjectives (*gut, alt, trocken*) modifying the noun *Wein*. Since the full inflectional paradigm of determiners, adjectives and nouns is already worked out only grammatical forms containing the indefinite determiner *ein* and the definite determiner *der* are listed.

ein guter alter Wein
der gute alte Wein

Furthermore examples without determiner for a mass noun and a bare plural are inserted. The most interesting forms here are dative singular and genitive plural since different adjective inflections are possible.[11]

gutem, altem Wein
gutem, alten Wein
** guten, altem Wein*

guter, alter, trockener Weine
guter, alten, trockenen Weine
guter, alter, trockenen Weine
** guter, alten, trockener Weine*

The number of test items for NP_Agreement-DET-ADJs-N is:

[11] As the ungrammatical examples show the order "strong—weak—strong" inflection is not possible.

NP_Agreement-DET-ADJs-N	
45	test items
38	well-formed items
7	ill-formed items
0	dubious items

Parameters

The components of well-formed noun phrases are annotated with the agreement features *gender*, *number* and *case*.

Attribute	Value	Instance	Comment
gender	masc/fem/neut	DET/ADJ/N	
number	sg/pl	DET/ADJ/N	
case	nom/gen/dat/acc	DET/ADJ/N	
		EXAMPLE	
gender	number	case	instance
fem	pl	nom	Wände
neut	pl	dat	den Projekten
masc	sg	acc	den alten Mann
masc	pl	gen	guter alter Weine

The ill-formed test items are only annotated with an error attribute and an error code since ungrammaticality can result from all parts of the noun phrases.

Data

The total number of test items for NP_Agreement is:

NP_Agreement	
2031	test items
299	well-formed items
1725	ill-formed items
7	dubious items

Chapter 7

Word Order

7.1 English

In the English test suite, here are no items testing for Word-Order as such. However, it is mentioned several times as one of the supertypes. This means that Word-Order is violated in the item, but that it is actually testing for another phenomenon.

In all other cases, it is assumed as a default presupposition.

7.2 French

Word Order is a rather special phenomenon since it is involved in every linguistic input which contains more than one element. Therefore, to be systematic, this phenomenon should be tested by changing the order of the elements in every well-formed test item. This would result in test items such as the following:

(7) well-formed item (WF =1): *la question*
ill-formed item (WF =0) : *question la*

(8) WF = 1 : *la femme pose une question*
WF = 0 : *la femme une question pose*
WF = 0 : *la femme une pose question*
WF = 0 : *la femme pose question une*
etc.

Although this method may seem to be the only systematic way to write ill-formed test items for Word Order, this is probably not the way the problem should be tackled, because in fact the same wrong order would be tested too many times. Therefore, the range of test items to test for word order should first be delimited.

In example (8) given above, we can already delimit this range by postulating that only **constituents** should be used. In this example, the constituents are *la femme*, *pose*, and *une question*.

But the number of derived ill-formed test items should be even more strictly limited, and we give below the general characteristics of the test items which are relevant for word order.

i) Input with only negative variants

An "input with only negative variants" is a well-formed test item from which only ill-formed test items can be derived by systematic permutation of the elements. This is often the case with fixed or rigid word order languages, and as French is not very flexible in this respect, such test items are very easy to find:

(9) a. WF = 1: *La femme veut venir.*
b. DERIVATIONS:
WF = 0 : *La femme venir veut.*
WF = 0 : *Venir la femme veut.*
WF = 0 : *Veut la femme venir.*
WF = 0 : *Venir veut la femme.*
WF = 0 : *Veut venir la femme.*

Some structures are more likely than others to give rise to actually occurring errors, in particular those which have variants with a different work order. Such an example in French is given by the combination of a compound tense, the presence of negation and the two constructions with a full NP or with a negative pro-form.

(10) Negation
a. WF = 1: *Je n'ai vu PERSONNE.*
WF = 1: *Je n'ai PAS vu Michel.*
b. DERIVATION:
WF = 0 : *Je n'ai PERSONNE vu.*

ii) Input with positive variant(s)

And "input with positive variant(s)" is a well-formed test item from which at least some well-formed test items may be derived by systematic permutation of the elements.

(11) a. WF = 1 : *Je viens aujourd'hui.*
b. DERIVATIONS:
WF = 0 : *Je aujourd'hui viens.*
WF = 0 : *Viens aujourd'hui je.*
WF = 0 : *Viens je aujourd'hui.*
WF = 0 : *Aujourd'hui viens je.*
WF = 1 : *Aujourd'hui je viens.*
WF = 1 : *Aujourd'hui, je viens.*

In this group of test items two items out the six possible permutations are actually well-formed. To be systematic, all six test items should be present in the test suite, but as some of them do not seem to be very important for the test data, we propose to only include the three items shown in (12):

(12) a. WF = 1 : *Je viens aujourd'hui.*
b. WF = 1 : *Aujourd'hui, je viens.*
c. WF = 0 : *Je aujourd'hui viens.*

This choice can be formalized as follows: only vary the element which is **interesting** for Word Order. Here the adverb *aujourd'hui* is the interesting element, because changing its position generates a new well-formed item.

The TSNLP test data includes both types of test items (i.e. having only negative variants, or having at least one positive variant) described above.

7.2.1 General Division of the Phenomenon

All the examples given above far can be related to phenomena other than Word Order. The examples in (8) (*Je viens aujourd'hui* vs. *Aujourd'hui, je viens*) concern Modification, others include clitics (involving Complementation), Negation etc.

This type of interaction is expressed in the database by having both Word Order and the other phenomenon involved listed as the SUPERTYPES for the test item.

This is actually often the case, and Word Order is most often linked with another phenomenon. Therefore, a large amount of TSNLP data relevant for Word Order (except for one phenomenon discussed below) is actually included with the data for other phenomena and discussed in the relevant chapters. The following list gives the names of the phenomena which include test data relevant for Word Order:

- S_Imperative_PRON_Wordorder
- S_Imperative_Negation_Wordorder
- S_Imperative_Negation_past_Wordorder
- S_Imperative_Negation_past_Wordorder
- S_Imperative_Negation_infinitive_Wordorder
- S_Imperative_Negation_PRON_Wordorder
- S_Imperative_Negation_infinitive_PRON_Wordorder
- S_Question_y-n_inverted-PRON_Wordorder
- S_Question_y-n_est-ce-que_subj_Wordorder
- S_Question_wh_inverted_p-obj(PRON-wh-det)_Wordorder
- S_Question_wh_inverted_p-obj(DET_wh+NP)_Wordorder
- S_Question_wh_inverted_p-obj(PRON_wh)_Wordorder
- S_Conditional_Wordorder

- C_Modification_1-mod_Wordorder
- C_Modification_2-mod_Wordorder
- C_Negation_Wordorder_pas
- C_Negation_Wordorder_rien
- C_Negation_Wordorder_personne
- C_Negation_Wordorder_jamais

All these phenomena have S_Wordorder or C_Wordorder (depending on the domain) in their list of SUPERTYPES, thus it is always possible to retrieve all the relevant test items together.

The TSNLP data includes one phenomenon which does not interact with another phenomenon, namely NP_Wordorder. The test data for NP_Wordorder is built on the test data for NP_Agreement (see 6.2), but there is no extended version of this test data yet (i.e. the NPs have not been integrated into sentences).

NP_Wordorder_DET_N

In French, the noun always follows the determiner. Thus, the TSNLP data tests for the position of the different kinds of determiners.

Le travail
* *Travail le*

7.2.2 Parameters

In general, the test data for Word Order (or for which Word Order is a supertype) contains the following parameters.

Wordorder

Attribute	Value	instance
position_X X = category or function	(ante_Y, post_Y) Y = category or function	
	EXAMPLES	
	De quoi parle - t - il ?	
position_subj	post_verb	parle - t - il
	* *De quoi il parle ?*	
position_subj	post_verb	* il parle

NP_Wordorder

Attribute	Value	instance
position_N	post_DET	
EXAMPLES		
le travail		
position_N	post_DET	le travail
** travail le*		
position_N	post_DET	* travail le

7.2.3 Data

The total number of test items for NP_Wordorder is (thus, the other phenomena containing wordorder are not included):

NP_Wordorder	
14	test items
7	well-formed items
7	ill-formed items

7.3 German

7.3.1 NP_Word-Order

The phenomenon NP_Word-Order deals with ordering regularities that hold among the elements inside noun phrases.

General Division of the Phenomenon

The data is classified into two sets. While the first set contains noun phrases with various elements in front of the head noun, the items of the second group have different components following the head.

Most of the constituents clustered around the head noun are modifiers so that there is an overlapping with the data illustrating NP_Modification. Except for the first two sets where determiners or determiner plus adjectives are placed in front of the head noun the test data is a subset of the NP_Modification data.

Elements of the following categories can occupy the pre-position:

- determiner

  To illustrate the word order between article and noun one representative determiner from each type — according to the following classification — is chosen.

Determiner Classification	
Determiner Types	Instances
DET_def	das, dasjenige, dasselbe
DET_indef	ein, irgendein, kein
	alle, jeder, einige, mehrere, manche
DET_indef-adj	beide, sämtliche, etliche
DET_dem	dieses, jenes
DET_poss	mein

  der Test
  **Test der*

  unsere Tests
  **Tests unsere*

- determiner(s) plus adjective(s)

  The data set illustrates three different classes. Firstly, one noun phrase represents word order variations between a determiner, two adjectives and a noun.

  der neue wichtige Test
  **der Test neue wichtige*
  **der neue Test wichtige*
  **neue wichtige der Test*

  Secondly, to cover a broad range of complex determination, this subgroup comprises all possible combinations with two different determiner types, according to the determiner schema above. Some representative constructions with more than two different determiners are included.[1]

**das dieses alte Projekt*
**jenes dieses alte Projekt*
**ein dieses alte Projekt*
**kein dieses alte Projekt*
alle diese alten Projekte
**einige diese alten Projekte*
beide diese alten Projekte
**etliche diese alten Projekte*
**mein dieses alte Projekt*

alle diese meine alten Projekte
alle diese meine etlichen alten Projekte
**diese meine einigen alten Projekte*

And finally, in all grammatical test strings with two and more determiners the adjective is placed at all possible positions.

alle diese alten Projekte
**alle alten diese Projekte*
**alten alle diese Projekte*

- adjective(s)

  The first example pair illustrates the word order between an adjective and a bare plural noun. Additionally an example with two adjectives is taken up.

  wichtige Tests
  **Tests wichtige*

  wichtige, neue Tests
  **wichtige Tests neue*
  **Tests wichtige neue*

- adjective phrases

  Besides adjectives modified by adverbs this data set comprises an example with an adjective extended by a prepositional phrase.

  der besonders wichtige Test
  **der Test besonders wichtige*

  der für alle Beteiligten wichtigste Test
  **der Test für alle Beteiligten wichtigste*

- particle

  This group contains grading particles such as *nur, auch, genug* which can be placed in front of nominal phrases.

  genug Tests
  ?Tests genug
  **genug Test*

[1] The test sets are built of strings with the same determiner type in front of the adjective.

- genitive NP

  The genitive NP — represented as a proper noun or as determiner-noun groups – can occupy the position in front of the noun phrase as well as the post-positon.

  Peters Test
  der Test Peters

  des Managers Test
  der Test des Managers

- determiner plus verbal participle

  Inflected participle types — represented as V or VP — can only be placed in front of the head noun.

  V_pcp-bare-infl:

  der lesende Manager
  **der Manager lesende*

  V_pcp-zu-infl(passive):

  der mitzuteilende Bericht
  **der Bericht mitzuteilende*

  VP_pcp-perf-infl(passive):

  der dem Manager gegebene Bericht
  **der Bericht dem Manager gegebene*

  VP_pcp-perf-infl(sein):

  der soeben angekommene Bericht
  **der Bericht soeben angekommene*

- relative clauses

  In general, relative clauses follow the head noun but in so called *pseudo cleft* constructions the sentence starts with a relative clause followed by the head noun.

  Den der Berater kennt, der Manager leitet das Projekt.
  Den der Berater kennt, dieser Manager leitet das Projekt.

  Additionally, this left extraposition can also occur in parentheticals.

  Gestern angekommen sind *die der Berater kennt — die Kinder.*

In the position following the head noun elements of five different categories are possible. Except for the last data set the singular noun phrases are specified by a definite determiner. The determiner-noun group is then treated as a single unit; thus, to construct ungrammatical items of this data set, the elements in post-position are not placed between determiner and head noun.

- genitive NP

  As mentioned above genitive NPs can be placed before and after the head noun.

  der Test Peters
  Peters Test

Berichte des Managers
des Managers Berichte

- prepositional phrase

  In general, prepositional phrases must follow the head noun.

  der Bericht über das Projekt
  **über das Projekt der Bericht*

- relative clause

  Except for pseudo cleft constructions and parentheticals relative clauses are placed after the head noun.

  der Manager, der kommt
  der Manager, dessen der Berater sich erinnert
  der Manager, dem der Berater vertraut
  der Manager, den der Berater kennt

  das Projekt, über das der Manager sich freut
  der Präsident, mit dem der Manager redet

- adverb

  Nominal structures where an adverb follows the head noun are restricted to a few cases.

  der Test morgen
  **morgen der Test*

  Jahre später
  **später Jahre*

- adjective phrases

  As post nominal parentheses uninflected adjective phrases can occupy the position after the head noun.

  der Test, wichtig für den Präsidenten
  **wichtig für den Präsidenten der Test*

- adjective

  This construction is very rare and mostly restricted to lyrical texts.

  Röslein rot
  **rot Röslein*

Parameters

- Construction of ill-formed items

  The ungrammatical noun phrases result either from order variations or in the case of the complex determination the errors are caused by determiner combinations which in general are not allowed.

- Annotations

  To make the order of the elements inside the noun phrase explicit — although it could be read of the *analysis* annotation — the input strings are annotated with the attribute *nominal word order*.

Attribute	Value
nominal word order	category
EXAMPLES	
input string	nominal word order
alle diese alten Projekte	DET_indef DET_dem ADJ N
** alle alten diese Projekte*	DET_indef ADJ DET_dem N
der Bericht von dem Manager	NP PP
**von dem Manager der Bericht*	PP NP

Data

The total number of test items for NP_Word-Order is:

NP_Word-Order	
329	test items
103	well-formed items
217	ill-formed items
9	dubious items

Chapter 8

Coordination

8.1 English

8.1.1 NP_Coordination

Coordination is where two or more elements (conjoins) are combined. Following Quirk et al. [Quirk *et al.*, 1991], we call the latter case "multiple" coordination.[1]

Coordination may be marked by explicit markers or not. Explicit markers of coordination are termed coordinators, the central class of which includes *and*, *or*, and *but*.

Other explicit markers of coordination include correlative pairs, the central class of which include *both ... and*, *either ... or*, and *neither ... nor*.

General Division of the Phenomenon

For NP coordination, the TSNLP data covers the following:

- Correlatives
- Coordinators
- Multiple_Coordination (with 3 conjoins)
- Case_assignment (within conjoins in subject and object position)

In TSNLP we distinguish between conjunctive coordinators (COORD_conj: *and, but*) and disjunctive coordinators (COORD_disj: *or*).

We term CORREL(atives) non-coordinators that occur in correlative pairs (*both, either, neither, nor*).

Only the central classes of coordinators and correlative pairs are covered by the data.

All items get a categorial analysis but no functional analysis. No parameters are filled in but error comments (and position/instance where appropriate) are given.

[1] Quirk et al. [Quirk *et al.*, 1991] use the term **conjoins** where other grammarians also employ **conjuncts**.

NP_Coordination-Coordinators

The data covers permissible coordinators in constructions with 2 (pronoun) conjoins. Ill-formed items are generated by inserting correlatives instead of coordinators.

he or she
* *he nor she*

NP_Coordination-Correlatives

The data covers permissible correlative pairs, ill-formed items being generated by systematically producing different combinations of combinators and correlatives.

neither he nor she
* *neither he or she*

NP_Coordination-Multiple_Coordination

The data shows constructions with 3 conjoins:

she, she and he
* *she, she but he*

NP_Coordination-Case_Assignment-Subject_position

In subject position, both conjoins must be either in nominative case, or in accusative case (in informal usage). Mixed nominative/accusative cases are not permitted. The data shows the full paradigm, with the closed class of personal pronouns expanded.

He and she succeed.
* *He and her succeed.*

NP_Coordination-Case_Assignment-Object_position

In object position, both conjoins must be in accusative case. The data shows the full paradigm of nominative/accusative cases, with the closed class of personal pronouns expanded.

He sees her and him.
* *He sees he and she.*

Data

The total number of test items for NP_Coordination is:

NP_Coordination	
333	test items
147	well-formed items
180	ill-formed items
6	dubious items
0	relatively ill-formed items

8.2 French

According to Grevisse (1991) [Grevisse, 1991] (§252–263), coordination is characterized by the following aspects:

1. the categories of the coordinated elements (or **conjuncts**):

   Conjuncts are most often of the same category (they can be clauses, subordinate clauses, NPs, Ns, etc.). Nevertheless, it is also often possible to coordinate "asymmetric" elements, i.e. the category of the conjuncts may be different. In that case, two cases must be distinguished with respect to the function of the conjuncts:

   a. the function is the same:
      - verbal arguments: NP and PRON can be conjoined;
        L'ingénieur et moi venons.
      - sentential modifiers: all categories which can act as adverbial complements (ADV, NP, infinitives introduced by a preposition, etc.) can be conjoined;
        Il répond calmement et en pesant ses mots.
      - modifiers of a noun (ADJ, participles, PPs)
        Augmentation successive et par degrés

   b. the function is different:
      - parentheticals
        Sa mère, qui ne voulait plus le voir– mais en était-elle sûre?– lui écrivait.

2. the number of conjuncts (two or more)

3. the relation between the conjuncts:

   Grevisse distinguishes two types of relations in coordination:

   a. In a **multilateral** relation, the meaning of the sentence remains the same when the order of the conjuncts is inverted or if an element is deleted. This type of coordination is also often called a **collection**.

   b. In a **unilateral** relation, each conjunct is explicitly related to another conjunct. Often specific terms are used, e.g. *respectivement*, *l'un... l'autre*, etc. This type of coordination is also called **distribution**.

   Only the first type of relation is currently covered in the TSNLP test data.

4. the presence or absence of a coordinating conjunction:

   a. **Explicit** coordination uses coordinating conjunctions between the conjuncts. Coordinating conjunctions form a closed class which includes *et, ou, ni, mais, car, or*.

      L'homme vient et la femme part.

   b. When there is no coordinating conjunction, the coordination is **implicit**, and the conjuncts are then usually separated by commas.

      La conscience, la religion ne sont pas de trop.

If a coordination contains more than two conjuncts, implicit and explicit coordination are usually mixed.

La conscience, la religion et l'opinion ne sont pas de trop.

5. restrictions on the use of the coordinating conjunctions:

   Only *et, ni, ou* and *mais* can conjoin phrasal elements.

8.2.1 General Division of the Phenomenon

As shown above, there are many aspects according to which Coordination can be described and a division of this phenomenon which would take all of them into account would be very complex. For the TSNLP test data, this division is not very fine-grained but has been simplified by only taking into account some of these aspects:

1. the domain of coordination:[2]
   - C_Coordination
   - NP_Coordination
   - PP_Coordination[3]
2. the number of coordinating conjunctions

   The TSNLP test data only contains items with one coordinating conjunction, but it can be extended as follows (∗ indicates the phenomena for which test data already exists).

   - ∗ C_Coordination
     - ∗ C_Coordination_1-coord
     - C_Coordination_2-coord
     - C_Coordination_3-coord, etc.
   - ∗ NP_Coordination
     - ∗ NP_Coordination_1-coord
     - NP_Coordination_2-coord

[2] This should not be confused with the elements which are coordinated. This may be a little confusing, since a term such as "NP Coordination" usually refers to the coordination of two (or more) NPs, whereas in TSNLP, NP_Coordination refers to a phenomenon which has an NP for domain. Although unusual, this classification presents for TSNLP the advantage of maintaining consistency across the test data, since the first subdivision of all the other phenomena is done according to their domains. However, as discussed below, a test item belonging to a domain "smaller" than a clause exists in two versions, as an isolated NP or PP and as an S containing that NP or PP:
Une foule ou un nombre de personnes
Une foule ou un nombre de personnes sont nécessaires.

Thus, the items corresponding to an intuitive definition of "NP Coordination" may be retrieved by searching for the PHENOMENON NP_Coordination and the CATEGORY "NP(augmented to S)".

Test items for agreement between subject and verb when the subject is a coordinated NP are included in the phenomenon NP_Coordination_Agreement (see below), but these test items only exist in the sentential version, and that the value of the attribute CATEGORY is "NP(augmented to S)".

[3] Because of time constraints, the TSNLP test data does not cover AP_Coordination, but some coordinated adjectives are included in NP_Coordination and NP_Agreement. Those test items have AP_Coordination as one of their supertypes.

  - NP_Coordination_3-coord, etc.
- ∗ PP_Coordination
  - ∗ PP_Coordination_1-coord
  - PP_Coordination_2-coord
  - PP_Coordination_3-coord, etc.

It is important to note that "1-coord" indicates that the test item includes **one** coordinating conjunction, i.e. that there are **two** conjuncts.

The next sections present in more detail the data for C_Coordination, NP_Coordination and PP_Coordination.

8.2.2 C_Coordination

Coordinating conjunctions do not all behave in the same way, and the phenomenon of C_Coordination is divided according to the conjunction which is used. Because of time constraints, the TSNLP data only covers the conjunction *et*. However, test data for *ou* would be mostly similar and *et* could easily be replaced by *ou* in existing test items to create the test data for *ou* (∗ indicates the phenomena for which test data already exists).[4]

- ∗ C_Coordination_1-coord(et)
- C_Coordination_1-coord(ou)
- C_Coordination_1-coord(mais)
- C_Coordination_1-coord(ni)
- C_Coordination_1-coord(car)

The test items for C_Coordination are based on those for C_Complementation.

For C_Coordination_1-coord(et), a further distinction is made according to whether the test input contains an ellipsis or not ("non-elliptic" is the default value).[5]

- ∗ C_Coordination_1-coord(et)

  (13) *L'ingénieur accepte la proposition et l'entreprise accepte le contrat.*

- ∗ C_Coordination_1-coord(et)_elliptic[6]

  (14) *L'ingénieur accepte la proposition et l'entreprise le contrat.*

[4] When there is more than one coordinating conjunction, it must be kept in mind that the test items probably use **implicit** coordination. For instance, the following example for C_Coordination_2-coord(et) contains a comma whose function is to coordinate:
Il pleut, il neige et il grêle.

[5] In the attribute ANALYSIS, the categories which include an ellipsis are referred to by adding the suffix "elliptic" and giving in parentheses the omitted element: C_elliptic(subj), C_elliptic(subj,V).

Note that an example such as *L'ingénieur accepte et signe le contrat.* is a test item for both C_elliptic(obj) and C_elliptic(subj).

[6] The TSNLP data does not include test items in which the conjuncts are of different categories, so the phenomenon name does not indicate which elements are coordinated, since this further subdivision would be too fine-grained for our purposes. It is nevertheless possible to retrieve test items according to the category of the coordinated elements by using the attribute ANALYSIS.

The type of coordination which Grevisse calls **collective** (see above) is included in elliptic coordination.

Note that a test item such as (15 a.) belongs to C_Coordination_1-coord(et)_elliptic, because it is considered to be the elliptic form of (15 b.):

(15) a. *L'ingénieur accepte le contrat et l'invitation.*
b. *L'ingénieur accepte le contrat et l'ingénieur accepte l'invitation.*

8.2.3 NP_Coordination

Because some NLP applications cannot deal with NPs in isolation, all the test data included under this heading exists in two versions: as an isolated NP, as in (16 a.), and as an S, as in (16 b.).[7]

(16) a. NP: *Moi et toi.*
b. S: *Moi et toi venons.*

If it is augmented to S, the NP is always in subject position and the verb is intransitive.

For NP_Coordination, the TSNLP data treats mainly the conjunctions *et* or *ou*, and the phenomenon name does not include the conjunction.[8] Like C_Coordination, NP_Coordination is divided according to whether the test item contains an ellipsis, and "non-elliptic" is the default value. Thus we have the following division:

- NP_Coordination_1-coord
  Moi et toi.
  L'homme et moi.
- NP_Coordination_1-coord_elliptic
  Beaucoup ou peu de vin.[9]

To make the test items semantically more plausible, most of the test data for NP_Coordination_1-coord_elliptic uses the conjunction *ou*. However, if the subject of a clause is a coordinated NP, an important question is the number of the Verb. This aspect of Coordination interacts with the phenomenon of Agreement and the test data for it falls under NP_Coordination_1-coord_Agreement.

NP_Coordination_1-coord_Agreement

More precisely NP_Coordination_1-coord_Agreement contains data such as:[10]

[7]This is indicated in the value of the attribute CATEGORY of the test item: it is either "NP" or "NP(augmented to S)".

[8]Note tha in this respect, the division of NP_Coordination is different from that of C_Coordination.

[9]Other examples for NP_Coordination_1-coord(et)_elliptic would be:
Plusieurs ou quelques entreprises
since we consider this item to be an elliptic version of:
Plusieurs entreprises ou quelques entreprises.

[10]Note that the phenomenon name NP_Coordination refers to coordination which occurs inside NPs and that the problem of subject-verb agreement with coordinated subjects should logically be included in the phe-

Moi et toi venons.
** Moi et toi venez.*

Here, the distinction between the conjunctions *et* and *ou* is particulary important, since the number of the verb depends on it. Thus, we have:

- NP_Coordination_1-coord_Agreement(et)
- NP_Coordination_1-coord_Agreement(ou)

For both, the TSNLP test data covers different kinds of subjects:

- coordinated pronouns (all persons are exhaustively tested)
- coordination of a pronoun and an NP
- coordination of two NPs (where the NP can be simple or complex)

The test data for NP_Coordination_1-coord_Agreement(ou) can be seen as the "complement" of that for NP_Coordination_1-coord_Agreement(et), because it only contains well-formed test items and test items which behave the same way with *ou* as with *et* are not included:

** Moi et toi viennent.*
** Moi ou toi viennent.* (not included)

The data for NP_Coordination is built on the test items for NP_Agreement. The data for NP_Coordination_1-coord_Agreement is in turn based on the data for NP_Coordination.

8.2.4 PP_Coordination

We assume that only a few NLP applications are able to deal with PPs in isolation and all the test data included under this heading exists in two versions: as an isolated PP and as an S.[11]

De l'entreprise et de l'ingénieur (PP)
L'homme parle de l'entreprise et de l'ingénieur. (S)

Division of PP_Coordination

- PP_Coordination_1-coord(et)
  - PP_Coordination_1-coord(et)_prep(de/à/en)
  - PP_Coordination_1-coord(et)_mixed-prepositions
  - PP_Coordination_1-coord(et)
- PP_Coordination_1-coord(ou)

nomenon C_Coordination. But since coordinated NPs, whether they are in subject position or not, are all tested under NP_Coordination, it seems more intuitive to include test items for agreement under the same heading. The supertypes of NP_Coordination_1-coord_Agreement are thus NP_Coordination and NP_Agreement.

[11]This is indicated in the value of the attribute CATEGORY of the test item: it is either "PP" or "PP(augmented to S)".

- PP_Coordination_1-coord(ou)_prep(de/à/en)
- PP_Coordination_1-coord(ou)_mixed-prepositions
- PP_Coordination_1-coord(ou)

Test data exists for the two coordinating conjunctions *et* and *ou*.[12]

Further division of the data is done according to the preposition of the PP. Indeed, in French, the three prepositions *à*, *de* and *en* behave differently than the other prepositions (e.g. *dans*, *avec*, etc.).[13] Thus, they are specifically named in the phenomenon name, while the test data for the other prepositions (the default) are found under PP_Coordination_1-coord(et) and PP_Coordination_1-coord(ou).

Dans l'entreprise et le restaurant

Both PP_Coordination_1-coord(et)_mixed-prepositions and PP_Coordination_1-coord(ou)_mixed-prepositions include test items which contain two different prepositions, as shown in the following example.

De ou à l'entreprise

8.2.5 Parameters

Except for NP_Coordination_Agreement, all the subphenomena of Coordination have the following parameters:

Attribute	Value	instance
coord_cat	X(list of categories), where X is a conjunction and list of categories corresponds to the list of coordinated elements	
EXAMPLE		
coord_cat	et(PRON_ton,PRON_ton)	Toi, moi
coord_cat	ou(NP,NP_elliptic(DET,N))	Le travail le plus habituel, le plus intelligent

NP_Coordination_Agreement has the same parameters as all the phenomena including agreement (see 6.2):

[12] The test data for *et* and *ou* is basically the same.
[13] See Grevisse (1991) [Grevisse, 1991] (§995)

Attribute	Value	instance
pers-num-gen not always all three attributes, it depends on the particular phenomenon	1/2/3-sg/pl-masc/fem	
	EXAMPLE	
pers-num num-gen	3-pl pl-masc	viennent intelligents

8.2.6 Data

The total number of test items for Coordination is:

Coordination	
698	test items
379	well-formed items
291	ill-formed items
28	dubious items

More specifically:

- C_Coordination:

C_Coordination	
165	test items
105	well-formed items
49	ill-formed items
11	dubious items

- NP_Coordination:

NP_Coordination	
371	test items
190	well-formed items
164	ill-formed items
17	dubious items

- PP_Coordination:

PP_Coordination	
162	test items
84	well-formed items
78	ill-formed items

8.3 German

8.3.1 C_Coordination

The phenomenon C_Coordination deals with the syntactic properties of sentential components connected either by conjunctive or disjunctive elements. The connecting words don't cause any structural difference on the sentences. The permutation and the deletion of sentence parts as well as the interaction of both operations are mainly responsible for the big variety of different sentences structures. ([Batori, 1975],[Höhle, 1991],[Kohrt, 1976], [van Oirsouw, 1985],[Renz, 1989])

General Division of the Phenomenon

The basis of the data collection are two complex sentences comprised of two simple clauses connected by *und*. These clauses each contain a main verb and three nominal complements (with unambiguous nominative, dative and accusative case marking). The finite main verb occurs in final or second position, according to whether the coordinated structure is the root sentence or extended by a complementizer.

Ich glaube, daß der Professor dem Präsidenten den Aktenkoffer schenkt, und der Student dem Kommilitonen den Roman verkauft.

Der Professor schenkt dem Präsidenten den Aktenkoffer und der Student verkauft dem Kommilitonen den Roman.

Thus, the phenomenon is subdivided into:

- C_Coordination-v2
- C_Coordination-vfinal

The two sentences were submitted to systematic variations. On the one hand all possible deletions of constituents in both sentences are exemplified. All patently impossible combinations, e.g., sentences where the same grammatical function (e.g., the nominative subject noun phrase) was deleted in both conjuncts are ommitted. For both sentence types several coordination phenomena like *gapping, right-node-raising, left-deletion* and *right-deletion* are covered in the data.

- Gapping

  Deletion of a main verb and of one or several nominal phrases inside a sentence.

  Ich glaube, daß der Professor dem Präsidenten, und der Student dem Kommilitonen den Roman verkauft.
  Der Professor schenkt dem Präsidenten den Aktenkoffer, und der Student den Roman.

- Right-Node-Raising

  Deletion of *one* constituent — placed right to the verb in the first conjunct — which has a structurally identical constituent in the right conjunct.

  Ich glaube, daß der Professor dem Präsidenten den Aktenkoffer, und der Student dem Kommilitonen den Roman verkauft.

Der Professor schenkt dem Präsidenten den Aktenkoffer, und der Student verkauft dem Kommilitonen den Roman.

- Right-Deletion

  Deletion of structurally identical material placed right to the verb.

  Ich glaube, daß der Professor dem Präsidenten, und der Student dem Kommilitionen den Roman verkauft.
  Der Professor schenkt, und der Student verkauft dem Kommilitonen de Roman.

- Left-Deletion

  Deletion of structurally identical material placed left to the verb.

  Ich glaube, daß der Professor dem Präsidenten den Aktenkoffer schenkt und den Roman verkauft.
  Der Professor schenkt dem Präsidenten den Aktenkoffer und verkauft dem Kommilitonen den Roman.

Second, the order of the nominal phrases in the second conjunct was systematically permuted. The permutations are constrained to one of the conjuncts in order to avoid a combinatory explosion.

Der Professor schenkt dem Präsidenten den Aktenkoffer, und der Student verkauft den Roman dem Kommilitonen.
Der Professor schenkt dem Präsidenten den Aktenkoffer, und dem Kommilitonen verkauft der Student den Roman.

Ich glaube, daß der Professor dem Präsidenten den Aktenkoffer schenkt, und der Student den Roman dem Kommilitonen verkauft.
Ich glaube, daß der Professor dem Präsidenten den Aktenkoffer schenkt, und dem Kommilitonen der Student den Roman verkauft.

Since all forms are based on the same two basic sentences, there are only two test sets comprised of:

- all variations of the coordinated *daß*-clauses
- all variations of the coordinated declarative sentences

Parameters

- Construction of ill-formed items

  For the sentential coordination data a finegrained error classification was worked out. The ungrammatical sentences are grouped into the following error classes:

  - four deletions
    **Der Professor schenkt, und dem Kommilitonen den Roman.*
  - word order
    **Ich glaube, daß der Professor den Aktenkoffer, und der Student dem Kommilitonen den Roman verkauft.*

  - gapping in first conjunct for verb-second
    **Der Professor den Aktenkoffer und der Student verkauft dem Kommilitonen den Roman.*
  - left deletion in first conjunct
    **Ich glaube, daß den Aktenkoffer schenkt, und der Student dem Kommilitonen den Roman verkauft.*
  - deletion of constituent in first conjunct
    **Ich glaube, daß der Professor den Aktenkoffer schenkt, und der Student dem Kommilitonen den Roman verkauft.*
  - deletion of constituent in second conjunct
    **Der Professor schenkt dem Präsidenten, und verkauft der Student den Roman.*
  - right deletion in second conjunct
    **Der Professor schenkt dem Präsidenten den Aktenkoffer, und der Student verkauft dem Kommilitonen.*
  - new topic after deletion of topic
    **Der Professor schenkt dem Präsidenten den Aktenkoffer und dem Kommilitonen verkauft den Roman.*
  - left and right deletion in same conjunct
    **Schenkt dem Präsidenten und der Student verkauft den Roman dem Kommilitonen.*
  - double right deletion in first conjunct
    **Der Professor schenkt, und den Roman der Student dem Kommilitonen.*

- Annotations

  All test strins are annotated with two attributes reflecting the number of deletions (*n_deletions*) and the surface form of the remaining elements (*coordination pattern*). The annotations in the table describe the following two example sentences.

  Der Professor schenkt den Aktenkoffer und verkauft den Roman dem Kommilitonen.
  Ich glaube, daß der Professor den Aktenkoffer und den Roman dem Kommilitonen verkauft.

Attribute	Value
n_deletions	1-4
coordination pattern	category plus suffix indicating conjunct
GRAMMATICAL EXAMPLES	
n_deletions	coordination pattern
2	N1 V1 A1 V2 A2 D2
3	N1 A1 A2 D2 V2

  According to the error classification the ungrammatical input strings are annotated with an additional attribute describing the source of the error. The sentences reflecting the first two error types are annotated in the following table.

UNGRAMMATICAL EXAMPLES		
n_deletions	coordination pattern	error
4	N1 V1 D2 A2	four deletions
2	N1 A1 N2 D2 A2 V2	word order

Data

The total number of test items for C_Coordination is:

C_Coordination	
534	test items
105	well-formed items
429	ill-formed items
0	dubious items

The number of test items for C_Coordination-v2 is:

C_Coordination	
297	test items
56	well-formed items
241	ill-formed items
0	dubious items

The number of test items for C_Coordination-vfinal is:

C_Coordination	
237	test items
49	well-formed items
188	ill-formed items
0	dubious items

Chapter 9

Negation

9.1 English

9.1.1 C_Negation

Negation of a simple sentence is accomplished by the insertion of the word *not*. Where a finite clause contains an auxiliary (or auxiliaries), *not* is placed before the first auxiliary.

He does succeed.
He does not succeed.

Clauses with full verbs, by contrast, cannot form negation in this way:

He succeeds.
** He succeeds not.*

Not is frequently contracted to *n't* and attached to the auxiliary:

He doesn't succeed.

The negative particle *not* or *n't* is frequently followed by one or more non-assertive items (eg. *any, anything, ever ...*):

He hasn't done any work.

The combination of *not* (with a non-assertive form) can be replaced in most cases by a corresponding negative word (e.g. *no*, *neither*, *never* ...).

Interesting questions arise concerning the scope of negation. For example, the negation in the following example (from Quirk et al. [Quirk *et al.*, 1991]):

John didn't search the room carefully.

could apply either to the predication as a whole or to the last element.

General Division of the Phenomenon

The C_Negation data is derived from the tense and aspect data, and covers simple sentences only (with no non-assertive words). Both full forms (with *not* and contractions (with *n't* and the pronoun *he*)) are covered:

- C_Negation_Tense-Aspect-Modality
- C_Negation_Tense-Aspect-Modality-Contractions

The auxiliaries *do*, *be* and *have* are covered, as well as the central class of modal auxiliaries:
may, *will*, *can*, *must*, *shall*
and the marginal modals:
need, *dare*, *ought used to*.
All forms are given in both the present and past tense.
All positive items get a full analysis, but not ill-formed items, and errors are reported as comments, not parameters.

C_Negation_Tense-Aspect-Modality

For the full form items, ill-formed items are generated by inserting *not* before a full verb:
** He works not.*
or by placing the *not* before instead of after the first auxiliary:
** He not will succeed.*

C_Negation_Tense-Aspect-Modality-DO_aux

- **Present**

  He does not succeed.
  ** He not does succeed.*

- **Past**

  He did not succeed.
  ** He not did succeed.*

Only the present form is shown in the examples below:

C_Negation_Tense-Aspect-Modality-BE_HAVE

** He succeeds not.*

He has not succeeded.
** He not has succeeded.*

He is not seen.
** He not is seen.*

C_Negation_Tense-Aspect-Modality-MAY

He may not succeed.
** He not may succeed.*

C_Negation_Tense-Aspect-Modality-WILL

He will not succeed.
** He not will succeed.*

C_Negation_Tense-Aspect-Modality-CAN

He can not succeed.
** He not can succeed.*

C_Negation_Tense-Aspect-Modality-SHALL

He shall not succeed.
** He not shall succeed.*

C_Negation_Tense-Aspect-Modality-MUST

He must not succeed.
** He not must succeed.*

C_Negation_Tense-Aspect-Modality-OUGHT

He ought not to succeed.
** He not ought to succeed.*

C_Negation_Tense-Aspect-Modality-USED

He used not to succeed.
** He not used to succeed.*

C_Negation_Tense-Aspect-Modality-DARE

He dare not succeed.
** He not dare succeed.*

C_Negation_Tense-Aspect-Modality-NEED

He need not succeed.
** He not need succeed.*

C_Negation_Tense-Aspect-Modality-Contractions

The data for the contracted forms follow the same subdivisions as the non-contracted forms.

Data

The total number of test items for C_Negation is:

C_Negation	
418	test items
289	well-formed items
129	ill-formed items
0	dubious items
0	relatively ill-formed items

9.2 French

The notion of negation is usually used to describe a semantic phenomenon whose properties come from logic: negation turns the truth value of a proposition into its opposite. At the grammatical level, negation refer to the markers or the constructions which are used to perform this semantic operation.

As discussed by Grevisse (1991) [Grevisse, 1991] (§973-983), sentential negation in French is a discontinuous construction, i.e. sentential negation is usually expressed by the negative particle *ne* and a negative adverb such as *pas*, *rien*, *personne*, etc.[1]

Il ne voit pas l'ingénieur.
Il ne voit rien.
Il ne voit personne.
Il ne voit jamais l'ingénieur.

Negation can be tested for two aspects:

- the position of the negative markers with respect to the verb;
- the presence of the particle and/or of the negative adverb.

9.2.1 General Division of the Phenomenon

The division of the phenomenon of Negation is based on the two aspects mentioned above.[2]

9.2.2 C_Negation

C_Negation_Wordorder

The TSNLP test data covers the most important negative adverbs, i.e. *pas, rien, personne* and *jamais*. Since they exhibit different behaviors with respect to position, we distinguish the following:[3]

- C_Negation_Wordorder_pas
- C_Negation_Wordorder_rien
- C_Negation_Wordorder_personne
- C_Negation_Wordorder_jamais
- C_Negation_Wordorder_ne

C_Negation_Wordorder_pas/rien/personne/jamais

For the first four negative markers, all the data is built on several kinds of clauses:

[1] The value of the attribute CATEGORY for these negative markers are "NEG(ne)", "NEG(pas)", "NEG(rien)", "NEG(personne)" etc.

[2] The TSNLP test data for C_Negation is currently underspecified with respect to the attribute ANALYSIS.

[3] As the position of the negative markers interacts with the phenomenon of Word Order, the supertypes of C_Negation_Wordorder are C_Negation and C_Wordorder.

• a transitive verb with a direct object (in the case of *rien* and *personne*, the direct object is the negative marker). To derive ill-formed test items, the negative marker is put in front of the verb and (if possible) after the direct object.

Il ne voit pas l'ingénieur.
** Il ne voit l'ingénieur pas.*
** Il ne pas voit l'ingénieur.*

• a transitive verb whose direct object is a clitic. The derivation of the ill-formed items is performed as above.

Il ne le voit pas.
** Il ne le pas voit.*
** Il ne pas le voit.*

• a ditransitive verb with a direct object and an indirect object. In addition to the above, the negative adverb can be put between the two clitics.

Il ne le lui donne pas.
** Il ne le lui pas donne.*
** Il ne le pas lui donne.*
** Il ne pas le lui donne.*

• a finite verb with an infinitive complement which is a transitive verb with its direct object. The negation is on the finite main verb. We only give here one example of an ill-formed item:

Il ne veut pas voir l'ingénieur.
** Il ne veut voir pas l'ingénieur.*

• a finite verb with an infinitive complement which is transitive verb with its direct object. The negation is on the infinitive complement. Again, we only give here one example of an ill-formed item:

Il veut ne pas voir l'ingénieur.
** Il veut ne voir pas l'ingénieur.*

• a finite verb with an infinitive complement whose direct object is a clitic (thus, in front of the infinitive verb). The negation is on the finite main verb.

Il ne veut pas le voir.
** Il ne veut le pas voir.*

• a finite verb with an infinitive complement whose direct object is a clitic (thus, in front of the infinitive verb). The negation is on the infinitive complement.

Il veut ne pas le voir.
** Il veut ne le pas voir.*

• a causative verb with an infinitive complement which is a transitive verb with its direct object. The negation is on the finite main verb.

Il ne fait pas démarrer le moteur.
** Il ne fait démarrer pas le moteur.*

• a causative verb with an infinitive complement whose direct object is a clitic (thus, in front of the infinitive verb). The negation is on the finite main verb.

Il ne le fait pas démarrer.
** Il ne pas le fait démarrer.*

• a causative verb with an infinitive complement which is a transitive verb with its direct object. The infinitive is in the past tense and the negation is on the infinitive complement.

Il avoue ne pas avoir fait démarrer le moteur.
Il avoue n'avoir pas fait démarrer le moteur.
** Il avoue n'avoir fait pas démarrer le moteur.*

• a causative verb with an infinitive complement whose direct object is a clitic (thus, in front of the infinitive verb). The infinitive is in the past tense and the negation is on the infinitive complement.

Il avoue ne pas l'avoir fait démarrer.
Il avoue ne l'avoir pas fait démarrer.
** Il avoue ne l'avoir fait pas démarrer.*

• a transitive verb which is in the passive voice and in the past tense.[4]

Il n'a pas été vu.
** Il n'a été pas vu.*

C_Negation_Wordorder_ne

The TSNLP data also contains items which test for the position of the negative particle *ne*.

• a transitive verb whose direct object is a clitic. To derive ill-formed items, the negative marker is put after the clitic.

Il ne le voit pas.
** Il le ne pas voit.*

• a ditransitive verb with a direct object and an indirect object, both of which are clitics. In addition to the above, the negative marker can be put between the two clitics.

Il ne le lui donne pas.
Il le ne lui donne pas.
** Il le lui ne donne pas.*

[4]The supertypes of these items are C_Negation, C_Wordorder and C_Passive.

• a causative verb with an infinitive complement whose direct object is a clitic.

Il ne le fait pas démarrer.
** Il le ne fait pas démarrer.*

C_Negation_ne

The presence of the particle *ne* with a negative adverb is tested as follows:

- C_Negation_ne(facultative)

  In French, with a restricted set of verbs, it is possible to use only the particle *ne*. This use is facultative and the presence of the negative adverb *pas* is not incorrect.[5]

  Il n'ose venir.
  Il n'ose pas venir.

- C_Negation_pas

  In written French (except for the verbs mentioned above), it is obligatory to use both elements of the discontinuous negation, the negative particle *ne* and the negative adverb *pas*. However, as the absence of *ne* is accepted in spoken language, this is mentioned in the attribute COMMENT for these ill-formed test items. In this case, the items are considered to be well-formed and the value of their attribute REGISTER is "informal".

  Il ne voit pas l'ingénieur.
  ** Il ne voit l'ingénieur.*
  ** Il voit pas l'ingénieur.*

- C_Negation_rien

  As in C_Negation_pas, we test for the presence of the negative particle *ne* and the negative adverb, i.e. for *rien*. Furthermore, the TSNLP data test for the absence of *pas*, an error which occurs quite often. We only give an example of this last aspect:

  Il ne voit rien.
  ** Il ne voit pas rien.*

- C_Negation_personne

  As with C_Negation_rien, but for the negative adverb *personne*.

- C_Negation_jamais

  As with C_Negation_rien, but for the negative adverb *jamais*.

- C_Negation_mixed

  Unlike English, French permits double negation, i.e. more than one negative adverb in a clause. The TSNLP data tests for the different combinations of negative adverbs. To derive ill-formed test items, we replace the negative adverb is replaced by its positive variant. Again, both the presence of the negative particle *ne* and the absence of *pas* are tested.

[5]According to Grevisse (1991) [Grevisse, 1991] (§975), these verbs are *cesser, oser* and *pouvoir*, and the TSNLP data tests exhaustively for them.

Il ne voit jamais rien.
** Il ne voit jamais quelque chose.*
** Il voit jamais rien.*
** Il ne voit pas jamais personne.*

Negation in other phenomena

Negation is also an important phenomenon for what has been called "Sentence Types" and the TSNLP test data for French covers it for Imperatives (see section 1.2.2). The relevant test items are included in the following subphenomena (as they have both S_Imperative and C_Negation among their supertypes, they can be retrieved easily).

- S_Imperative_Negation_Wordorder
- S_Imperative_Negation_past_Wordorder
- S_Imperative_Negation_infinitive_Wordorder
- S_Imperative_Negation_PRON_Wordorder
- S_Imperative_Negation_PRON_form
- S_Imperative_Negation_infinitive_PRON_Wordorder

9.2.3 Parameters

The TSNLP test data for C_Negation is partly underspecified with respect to the parameters: for the phenomena under C_Negation_Wordorder, only C_Negation_Wordorder_pas has any parameters, and we only give examples for this phenomenon. However, as this documentation is intended to indicate how parameters should be set for the other phenomena, the examples are preceded by a more general definition of the parameters.

C_Negation_Wordorder_pas

Attribute	Value	instance
position_NEG(X) X: "ne","pas","rien" personne" or "ne"	(ante_Y, post_Y) Y: post_verb, ante_obj(NP), ante_PRON, post_PRON, ante_infinitive, ante_ppast, etc.	
	EXAMPLES	
	Il ne voit pas l'ingénieur.	
position_NEG(pas)	post_verb,, ante_obj(NP)	voit pas l'ingénieur
	** Il ne voit l'ingénieur pas.*	
position_NEG(pas)	post_verb,, ante_obj(NP)	* voit l'ingénieur pas

Here are the parameters for all the of subphenomena of C_Negation except for C_Negation_Wordorder.

Attribute	Value	instance
negation	(Y1,Y2,etc) list of the negative markers	
EXAMPLES		
Il ne voit rien.		
negation	(ne,rien)	ne, rien
** Il voit rien.*		
negation	(ne,rien)	* rien
** Il ne voit pas rien.*		
negation	(ne,rien)	* ne, pas, rien

9.2.4 Data

The total number of test items for C_Negation is:

C_Negation	
168	test items
68 100	well-formed items ill-formed items

9.3 German

For the three negation phenomena S_Negation, C_Negation and NP_Negation the strategy for building ungrammatical sentences as well as the linguistic description are the same.

- Construction of ill-formed items

  The ungrammatical sentences result from variations over the position of the negation particles.

- Annotations

  As the ungrammatical examples show, the main syntactic aspect in negation is the position of the negation particles. This attribute is already covered by the sentence analyses so that there is no additional parameter to describe negation.

9.3.1 S_Negation

The phenomenon S_Negation deals with declarative sentences accomplished by the insertion of negation particles such as *nicht, niemals, nie, nie und nimmer, keinesfalls, keineswegs, nirgends, nirgendwo, nirgendwoher, nirgendwohin.* The negation particle always refers to the finite verb and thus its scope — reflected in the domain of the dependency analysis — covers the whole sentence.

Among the negation particles *nicht* has an extra status. While all negation particles can occupy the position in front of the finite verb (topic position), sentences with *nicht* at this position are ungrammatical.

Niemals *kommt der Manager.*
*Nicht *kommt der Manager.*

In sentences with unmarked word order both particle types cannot intersperse between the topic position and the finite verb.

**Der Manager* nicht *kommt.*
**Der Manager* niemals *kommt.*

On the other hand, sentences with *nicht* in front of nominal phrases are well-formed. But with particles like *niemals* word order and tense seem to influence grammaticality.

Nicht *der Manager gibt dem Präsidenten den Test.*
Der Manager gibt nicht *dem Präsidenten den Test.*

*Niemals *der Manager gibt dem Präsidenten den Test.*
?Dem Präsidenten gibt niemals *der Manager den Test.*
Dem Präsidenten hat niemals *der Manager den Test gegeben.*

Word order variation except for the finite verb in topic position are not covered in the data.

General Division of the Phenomenon

The first data set covers the complete active paradigm of the intransitive verb *kommen* with the negation particles *nicht* and *niemals* at all possible positions.

Der Manager ist nicht *gekommen.*
Nicht *der Manager ist gekommen.*
**Der Manager ist gekommen* nicht.
**Der Manager* nicht *ist gekommen.*
*Nicht *ist der Manager gekommen.*

The second group covers the complete passive paradigm of the transitive verb *begrüßen* with the negation particles *nicht* and *niemals* at all possible positions.

Der Manager wird nicht *begrüßt worden sein.*
Nicht *der Manager wird begrüßt worden sein.*
**Der Manager wird begrüßt worden sein* nicht.
**Der Manager wird begrüßt worden* nicht *sein.*
**Der Manager wird begrüßt* nicht *worden sein.*
**Der Manager* nicht *wird begrüßt worden sein.*
*Nicht *wird der Manager begrüßt worden sein.*

For the third data part negated sentences are derived from C_Complementation data. Monovalent, divalent and trivalent verbs with the negation particle *nicht* at all possible sentence positions are included.

Daß der Manager kommt, stimmt nicht.
Der Manager glaubt nicht, *daß der Student kommt.*
Es liegt dem Studenten nicht *daran, den Manager zu ärgern.*

One data set with a trivalent verb in present tense and present perfect illustrates the difference between the negation particles *nicht* and *niemals* in a more complex sentence structure.

Nicht *der Manager gibt dem Studenten den Test.*
*Niemals *der Manager gibt dem Studenten den Test.*

*Nicht *hat der Manager dem Studenten den Test gegeben.*
Niemals *hat der Manager dem Studenten den Test gegeben.*

Data

The total number of test items for S_Negation is:

S_Negation	
172	test items
72	well-formed items
99	ill-formed items
1	dubious items

9.3.2 C_Negation

The phenomenon C_Negation deals with embedded clauses accomplished by the insertion of negation particles such as *nicht, nie, keinesfalls.*

General Division of the Phenomenon

The data covers the complete active paradigm of the intransitive verb *kommen* in the embedded *daß*-clause with the negation particles *nicht* and *nie* at all possible positions.

?Der Manager glaubt , nicht *daß der Präsident gekommen sein wird.*[6]
Der Manager glaubt , daß nicht *der Präsident gekommen sein wird.*
Der Manager glaubt , daß der Präsident nicht *gekommen sein wird.*
**Der Manager glaubt , daß der Präsident gekommen* nicht *sein wird.*
**Der Manager glaubt , daß der Präsident gekommen sein* nicht *wird.*
?Der Manager glaubt , daß der Präsident gekommen sein wird nicht.[7]

Data

The total number of test items for C_Negation is:

	C_Negation
58	test items
24	well-formed items
17	ill-formed items
17	dubious items

9.3.3 NP_Negation

The phenomenon NP_Negation deals with negated nominal phrases.

General Division of the Phenomenon

The following nominal phrases with negation elements are covered.

- negation determiner *kein* plus noun

  The most common form of NP_Negation is the combination of the negation determiner *kein* with a noun.

  Kein *Manager kommt.*
  **Manager* kein *kommt.*

- *nicht* plus nominal phrase

  To result in a grammatical phrase the negation particle *nicht* must be placed immediately in front of the nominal phrase.

  Nicht *der Manager gibt dem Präsidenten den Test.*
  **Der Manager* nicht *gibt dem Präsidenten den Test.*
  **Der* nicht *Manager gibt dem Präsidenten den Test.*

[6]cf. *Nicht daß der Präsident kommt, glaubt der Manager, aber daß ...* requires a special intonation and is probably only possible if the sentence is extended by another *daß*-clause.
[7]The sentence is grammatical if the negation particle refers to the matrix verb *glaubt*.

- *nicht* plus coordinated nominal phrase

  The scope of the negation particle depends on the coordination type. If the nominal elements are connected by *und* or *oder* for example the complex nominal phrase is negated while *nicht* in structures with *nicht...sondern* can only refer to the adjacent noun phrase.

  Nicht *der Manager und der Präsident kommen.*
  **Der Manager und der Präsident* nicht *kommen.*

  Nicht *der Manager, sondern der Präsident kommt.*
  **Der Manager* nicht *sondern der Präsident kommt.*

- negation pronouns

  Some pronouns with inherent negation are covered in the data.

  Niemand schläft.
  Nichts schläft.
  Keiner schläft.

- negation determiner plus nominalized adjective

  A special case of NP_Negation are nominal phrases built of *nichts* with nominalized adjectives. Only adjectives in neutrum singular can be combined with this negation determiner.

  Der Manager sagt nichts *Falsches.*
  **Der Manager sagt Falsches* nichts.
  *Nichts *sagt der Manager Falsches.*

- emphasized negation pronoun *kein*

  Präsidenten kommen keine.
  Keine *Präsidenten kommen.*

  Ein Präsident kommt keiner.
  **Keiner ein Präsident kommt.*
  **Ein keiner Präsident kommt.*

Data

The total number of test items for NP_Negation is:

NP_Negation	
25	test items
14	well-formed items
11	ill-formed items
0	dubious items

Chapter 10

Parentheticals

10.1 English

Parenthetical expressions are remarks or extra items inserted into a sentence. Syntactically, they can range from words or even parts of words to whole sentences, and may not correspond to constituents in the normal sense.

Parenthetical expressions may be marked by a single dash (-) or by correlative parenthetical marker (term from Quirk et al. [Quirk *et al.*, 1991]).

Parenthetical markers include:

- commas (,)
- dashes (-)
- brackets ((), [],)

The comma is the least intrusive, the dash is more dramatic and less formal. Dashes or parentheses are used if the parenthetical item is an independent clause.

Certain punctuation rules obtain with parenthetical expressions (see below).

10.1.1 General Division of the Phenomenon

The data covers the following:

Parentheticals-Punctuation_markers

This illustrates allowable parenthetical punctuation markers.

- **NP_Parentheticals-Punctuation_markers**

  The manager (John) succeeds.
  ** The manager - John , succeeds.*

Parentheticals-Embedded_parentheses

Any of the correlative punctuation markers may enclose each other.

- **NP_Parentheticals-Embedded_parentheses**

  She knows the employee (John Smith [alias Wilson]).
  ** She knows the employee (John Smith (alias Wilson]].*

- **C_Parentheticals-Embedded_parentheses**

  He (she knows, in the firm,) succeeds.
  He , she knows, in the firm, succeeds.

Parentheticals-Omitted_item

- **NP_Parentheticals-Omitted_item**

  Square brackets ([]) are used in serious writing for the explanations of the reference of pronouns, and to replace pronouns or other omitted items in quotations.

  He says that she [the manager] succeeds.

Parentheticals-Punctuation_interaction

- NP_Parentheticals-Punctuation_interaction

  At the opening point of an inclusion none of the correlative punctuation markers can cooccur with any separation marker, so the following are not allowed:
  ** I spoke to John, (the manager) when I next saw him.*

  At the closing point of an inclusion only a closing bracket can cooccur with another separation marker:

  ** I spoke to John - the manager - , when I next saw him.*
  I spoke to John (the manager), when I next saw him.

- **S_Parentheticals-Punctuation_interaction**
  A parenthetical clause enclosed within a sentence may contain a question or exclamation mark but not a full stop:

  He sees the manager (it is John) today.
  ** He sees the manager (it is John.) today.*

10.1.2 Data

The total number of test items for Parentheticals is:

Parentheticals	
40	test items
12	well-formed items
27	ill-formed items
1	dubious items
0	relatively ill-formed items

Chapter 11

Abbreviations

11.1 English

Abbreviations are short versions of words which embrace a large number of semantic types (for example, titles, measures, dates, etc.).

There are three productive ways in which abbreviation is involved in word-formation (this classification is due to Quirk et al. [Quirk *et al.*, 1991]).

- Clippings:
  Part of the word is clipped off, usually the initial part:
  ad = advertisement
  but sometimes the end part:
  phone= telephone.
- Acronyms:
  Acronyms are formed from the initial letters of words that make up a name. There are two main types:
  - Those that are pronounced as sequences of letters.
    UFO = unidentified flying object
    eg = for example
  - Those that are pronounced as a word:
    NATO the North Atlantic Treaty Organisation
- Blends
  Blends are formed by blending one word with another.
  brunch = *br*eakfast and l*unch*

Some variation in the use of periods within abbreviations is permitted in some cases (*e.g.* or *eg*) but not in others (*TV* but not * *tv*). In much modern practice no periods are preferred.

Likewise, some leeway is sometimes possible with respect to capitalisation (*o.k.* and *O.K.* are both allowed in American English), but not in others (*Mrs* but not *MRS*).

Abbreviations can be marked in several ways:

- By an oblique:
  c/o = care of

- By an apostrophe
  '95 = 1995

- By a dash
  1985-95 = 1985-1995

Plurality may be marked with an apostrophe or (increasingly) without:
PhD's
PhDs

11.1.1 General Division of the Phenomenon

The data covers clippings and acronyms.

Clippings-Punctuation

- **ADVP_Clippings-Punctuation**
  approx.
  approx

- **NP_Clippings-Punctuation**
  '95
  ** -95*

  1985-95
  ** 1985'95*

Acronyms-Punctuation

- **PP_Abbreviations-Acronyms-Punctuation**
  e.g.
  eg

- **NP_Abbreviations-Acronyms-Punctuation**
  PhD's
  PhDs

Acronyms-Capitalisation

- **PP_Abbreviations-Acronyms-Capitalisation**
  e.g.
  E.G.

- **NP_Abbreviations-Acronyms-Capitalisation**
  Mrs
  * *mrs*

11.1.2 Data

The total number of test items for Abbreviations is:

Abbreviations	
18	test items
12	well-formed items
6	ill-formed items
0	dubious items
0	relatively ill-formed items

Part I

Annexes

Appendix A

Functions and Categories

Although we are aware that any description must in some sense be based on some theory, this is not a proposal for a linguistic theory, only a list of the labels needed for the description of test items. We of course do not claim that this list is complete or that it would meet the demands of descriptions for other purposes. As it stands, this list covers the distinctions needed for the analysis of the test items.

A.0.3 Categories

TSNLP is not meant to embody a specific syntactic theory and we tried to be as neutral as possible in establishing the list of syntactic categories to be used in the annotations for the test data. Thus the names of the categories are taken from standard linguistic practice and the few hierarchical relations we provide are hopefully uncontroversial. For the major syntactic categories, we distinguish between the syntactic phrases (NP, C, AP, ADVP, PP) and their corresponding lexical categories: N V, ADJ, ADV, P). For the minor categories (DET, CPL, COORD, etc.) there is probably no need to distinguish these two levels. The more complex categories S, C and VP are further described in A.0.3.

LIST OF SYNTACTIC CATEGORIES	
S	(Sentence)
C	(Clause)
DET	(Determiner)
NP	(Noun Phrase)
N	(Noun)
PRON	(Pronoun)
AP	(Adjective Phrase)
ADJ	(Adjective)
PP	(Prepositional Phrase)
P	(Preposition)
ADVP	(Adverb Phrase)
ADV	(Adverb)
VP	(Verb Phrase)
V	(Verb)
PART	(Particle)
CPL	(Complementizer)
NEG	(Negation)
COORD	(Coordination)
COMP	(Comparative)
PUNCT	(Punctuation)

We have chosen to use English labels for categories at the highest level, but language-particular additions may be used for languages other than English, and all syntactic categories may be further specified by features when necessary (e.g. "NP_nom", "C_rel", "PUNCT(!)", etc...).

The category labels used in the English, French and German test data are presented in parallel in section A.0.3. The three lists of category labels used for each language can also be found in Annex A, with examples for each language.

Agreement features in category names

For some categories, agreement features may be added when they are relevant (German and French use the order "case-person-number-gender", English uses the order "person-case-number-gender").

The values for **case**, **person**, **number** and **gender** are as follows:

case	: nom, acc, dat, gen
person	: 1, 2, 3
number	: sg, pl
gender	: fem, masc, neut

Some examples are:

– V_3-sg
– V_refl-3-sg
– NP_pl-fem
– PRON_nom-2-sg

– ADJ_pl-fem

Distinction between S, C and VP

The two categories S and C are not equivalent, although they may sometimes seem to label the same input string, because a sentence (S) may be composed of only one clause (C). Nevertheless, they are different:

• A clause is defined as the syntactic phrase containing a verb (finite or non-finite) and all of its arguments (including the subject) and modifiers. On the other hand, a sentence is the whole unit starting with a capital letter and ending with a strong punctuation mark in normal written text.

• A sentence may contain several clauses, but it may also be composed of only one clause.

In our test data, the input for most test items is a single sentence, ending with a period. Future data might include test items whose input is a paragraph composed of several sentences.

• The category VP is included to account for VP complements, e.g. infinitives or participials. Its inclusion in the list of TSNLP categories does not imply that the TSNLP analysis of a clause must involve a VP level or follow a particular linguistic theory: the test data writers are free to use it or not to use it, depending on the particular language and the analysis they wish to adopt.

From the point of view of phenomenon description, 'subcategorization' (or 'valency') includes the subject, and so is better treated at the clause level (i.e. C_Complementation). Subject-Verb agreement is also a phenomenon which is treated at the clause level (C_Agreement).

List of TSNLP categories (English, French and German)

Note that "-" means explicit absence of the corresponding category.

C A T E G O R I E S		
ENGLISH	FRENCH	GERMAN
	SENTENCE (S)	
S	S	S
S_imp	S_imp	S_imp-v1
S_int	S_int	S_int-wh-v2
		S_int-v1
		S_cond(wenn-dann)
		S_cpl-adv(als-da)
		S_cpl(daß)-vfinal
	CLAUSE (C)	
C_sub-cpl(that)-ind	C_cpl(que)	C_cpl(daß)-vfinal
C_sub-cpl(that)-should		
C_sub-cpl(that)-subj		
C_sub-cpl(0)		
C_sub-cpl(if,whether)	C_cpl(si)	C_cpl(ob)-vfinal
C_sub-cpl-adv(word)	C_adv-cpl(word)	C_cpl-adv(word)-vfinal
		C_cpl(wenn)-vfinal
C_main-decl	C_main	C_decl-v2
C_main-int-wh-inv	C_wh	C_wh-v2
C_main-int-wh		
C_main-int-y/n		
C_sub-cpl(WH)		C_wh-vfinal
C_sub-rel	C_rel	C_rel-vfinal
C_main-imp		
C_main-cond	C_cond	C_cond-v1
C_sub-cond		
	VERB PHRASE (VP)	
VP_ing	VP_ppres	VP_pcp-bare-infl
	VP_ppres-infl	VP_pcp-bare-uninfl
VP_perf	VP_ppast	VP_perf
	VP_ppast-infl	VP_perf(sein)
		VP_perf(passive)
		VP_perf(sein/passive)
		VP_pcp-perf-infl(sein)
		VP_pcp-perf-infl(passive)
		VP_pcp-perf-uninfl(sein)
VP_bare-inf	VP_inf	VP_sup-bare
		VP_sup-bare(sein)
		VP_sup-bare(passive)
		VP_sup-bare(sein/passiv)

ENGLISH	FRENCH	GERMAN
VERB PHRASE (VP) (continued)		
VP_to-inf	VP_à-inf VP_de-inf	VP_sup-zu VP_sup-zu(sein) VP_sup-zu(passive) VP_sup-zu(sein/passive) VP_pcp-zu-infl(passive) VP_pcp-zu-uninfl(passive)
	VP_pour-inf	VP_cpl(um,...)
VERB (V)		
V	V	
—	V_refl	
V_copula	V_copula	V_copula
V_aux, V_aux-neg	V_aux	V_aux
TO_aux		
—	-	V_mod
V_ing	V_ppres V_ppres-infl	V_pcp-bare-infl V_pcp-bare-uninfl
V_perf	V_ppast V_ppast-infl	V_perf V_perf(sein) V_perf(passiv) V_perf(sein/passive) V_pcp-perf-infl(sein) V_pcp-perf-infl(passive) V_pcp-perf-uninfl(sein)
V_bare-inf	V_inf	V_sup-bare V_sup-bare(sein) V_sup-bare(passive) V_sup-bare(sein/passive) V_sup-zu V_sup-zu(sein) V_sup-zu(passiv) V_sup-zu(sein/passive) V_pcp-zu-infl(passive) V_pcp-zu-uninfl(passiv)
NOUN PHRASE (NP)		
NP, N (_pl or _sg)	NP, N	NP, N
NP_mass	N_mass	N_mass
NP_pn (proper name)	NP_pn	NP_pn

ENGLISH	FRENCH	GERMAN
PRONOUN (PRON)		
	(default: non tonic)	
PRON_dummy		PRON_dummy
	PRON_impers	PRON_expl-topic
PRON_expl		PRON_expl-cor
PRON_rel	PRON_rel	PRON_rel
PRON_wh	PRON_wh	PRON_wh
PRON_pers	PRON_pers	PRON_pers
	PRON_pers-ton	
PRON_poss	PRON_poss	PRON_poss
PRON_refl	PRON_refl	PRON_refl
PRON_reci	PRON_reci	PRON_reci
PRON_dem	PRON_dem	PRON_dem
PRON_indef	PRON_indef	PRON_indef
CORRELATIVES (CORREL)		
CORREL		
DETERMINER (DET)		
DET_poss	DET_poss	DET_poss
DET_def	DET_def	DET_def
DET_indef	DET_indef	DET_indef
		DET_indef-adj
	DET_part	
DET_dem	DET_dem	DET_dem
DET_wh	DET_wh	DET_wh
DET_rel		
DET_quant	DET_quant	DET_indef
PREPOSITION PHRASE (PP)		
PP, P	PP, P	PP, P
PP_slash-NP		
—	P_det	P_det
ADJECTIVE PHRASE (AP)		
AP, ADJ	AP, ADJ	AP, ADJ
		ADJ_e
		ADJ_es
		ADJ_er
		ADJ_em
		ADJ_en
ADJ_comp	AP_comp	ADJ_comp_e
		etc.
ADJ_sup	AP_sup	ADJ_sup_e
		etc.
		ADJ_dem

ENGLISH	FRENCH	GERMAN
ADVERB PHRASE (ADVP)		
ADVP, ADV	ADVP, ADV	ADVP, ADV
		ADV_pron
		ADV_pron-correl
ADV_wh	ADV_wh	ADV_wh(prep)
		ADV_reci
ADV_grade		ADV_grade
COMPLEMENTIZER (CPL)		
CPL(that)	CPL(que)	CPL(dass)
CPL(if)	CPL(si)	CPL(ob)
CPL_adv(word)		CPL_adv(word)
COORDINATION (COORD)		
COORD_conj	COORD_conj	COORD_conj
COORD_disj	COORD_disj	COORD_disj
NEGATION (NEG)		
NEG_part	NEG(ne)	NEG_part(word)
	NEG(pas)	
	NEG(personne)	
	NEG(rien)	
COMPARATIVE (COMP)		
COMP_adj	COMP_adj	COMP_adj
COMP_np	COMP_np	COMP_np
COMP_vp		
PARTICLE (PART)		
	PART_wh(est-ce que)	-
	PART_euph	-
		PART_comp
		PART_grade
		PART_mod
PUNCTUATION (PUNCT)		
PUNCT(!)	PUNCT(!)	PUNCT(!)
PUNCT(,)	PUNCT(,)	PUNCT(,)
PUNCT(-)	PUNCT(-)	PUNCT(-)
PUNCT(.)	PUNCT(.)	PUNCT(.)
PUNCT([)	PUNCT([)	PUNCT([)
PUNCT(?)	PUNCT(?)	PUNCT(?)
ABBREVIATIONS (ABBR)		
ABBR	ABBR	ABBR

A.0.4 Functors and Arguments

In addition to the categorial labels, we provide a level of description of the input in which functor-argument relations are indicated. The motivation for this level of description is the hope that, in the case of non-configurational languages, it will be easier to state an accepted/acceptable dependency structure rather than a constituent structure in terms of syntactic categories.

The list of function and argument labels actually used for each language can be found in Annex A, and here we will only make some general remarks about the notions involved.

The notion of function adopted in TSNLP is a syntactic notion (in the style of LFG/HPSG) in which functors roughly coincide with heads and non-functors comprise the large functional classes of arguments (ARG), modifiers (MOD) and specifiers (SPEC).

A functional relation consists of two parts, the functor and the domain of the function. The basic distinction between functors and non-functors is mirrored in the specification of the **domain** given in the TSNLP annotations. For each element, **domain** specifies either:

• (for a functor) the domain within which this element functions as the functor (or "head"); or

• (for elements which are not functors), the functor in relation to which this element plays the role of ARG, MOD, or SPEC.

For non-functors, either the basic distinction ARG, MOD, SPEC or subclassifications can be used, according to the requirements specific to the particular language. The function labels used in the English, French and German test data are presented in parallel in section A.0.4 . The three lists of function labels used for each language can also be found in Annex A, with examples for each language.

A.1 shows an example of a functor-dependency structure for the test input *Der Plan wird den Manager geschmerzt haben.* It is followed by the values of the analysis such as it appears in a test item.

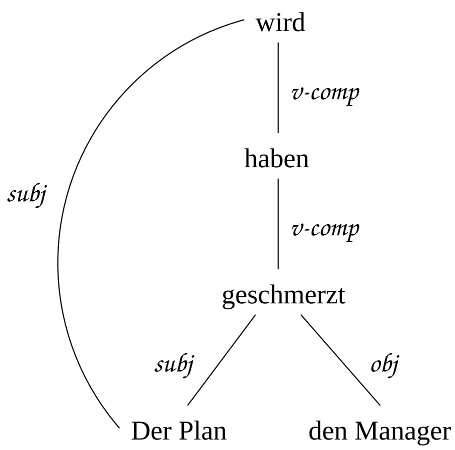

Figure A.1: Function-dependency structure of the sentence *Der Plan wird den Manager geschmerzt haben*

item	**position**	**category**	**function**	**domain**
wird	2:3	V_aux	func	0:7
Der Plan	0:2	NP	subj	2:3
Der Plan den Manager geschmerzt haben	0:2,3:7	—	v-comp	2:3
haben	6:7	V_sup-bare	func	0:2, 3:7
Der Plan den Manager geschmerzt	0:2, 3:6	—	v-comp	6:7
geschmerzt	5:6	V_perf	func	0:2, 3:6
Der Plan	0:2	NP	subj	5:6
den Manager	3:5	NP	obj	5:6

List of TSNLP functions (English, French and German)

F U N C T I O N S		
ENGLISH	FRENCH	GERMAN
FUNCTOR (FUNC)		
func	func	func
ARGUMENTS/COMPLEMENTS		
	arg arg-loc arg-dir arg-manner arg-temp	arg arg-loc arg-dir arg-manner arg-temp
subj	subj subj-impers	subj expl-topic
obj	obj	obj
i-obj		i-obj
p-obj: p-obj(about) p-obj(across) p-obj(after) p-obj(against) p-obj(among) etc. p-obj-loc p-obj-dir p-obj-modal p-obj-temp	p-obj: p-obj(à)-NP p-obj(de)-NP p-obj(pour)-NP p-obj(avec)-NP p-obj(contre)-NP etc. p-obj(à)-inf p-obj(de)-inf p-obj(pour)-inf	p-obj: p-obj(zwischen) p-obj(auf) p-obj(über) p-obj(wie) p-obj(als) etc. p-obj-loc p-obj-dir p-obj-modal p-obj-temp
s-comp v-comp n-comp	s-comp v-comp	s-comp v-comp
subj-pred: subj-pred-adj subj-pred-np subj-pred-pp	subj-pred: subj-pred-adj subj-pred-np subj-pred-pp	subj-pred subj-pred-adj subj-pred-np subj-pred-pp
obj-pred: obj-pred-adj obj-pred-np obj-pred-pp	obj-pred: obj-pred-adj obj-pred-np obj-pred-pp	obj-pred obj-pred-adj obj-pred-np obj-pred-pp
		iobj-pred-adj
verb-particle		

F U N C T I O N S		
ENGLISH	FRENCH	GERMAN
MODIFICATION (MOD)		
mod	mod	mod
mod-loc	mod-loc	mod-loc
mod-loc-dir	mod-loc-dir	mod-loc-dir
mod-loc-origin	mod-loc-origin	mod-loc-origin
mod-loc-place	mod-loc-place	mod-loc-place
mod-loc-path	mod-loc-path	mod-loc-path
mod-temp	mod-temp	mod-temp
mod-temp-duration	mod-temp-duration	mod-temp-duration
mod-temp-start	mod-temp-start	mod-temp-start
mod-temp-end	mod-temp-end	mod-temp-end
mod-temp-moment	mod-temp-moment	mod-temp-moment
mod-manner	mod-manner	mod-manner
mod-instr	mod-instr	mod-instr
mod-cause	mod-cause	mod-cause
mod-purpose	mod-purpose	mod-purpose
mod-possessive	mod-possessive	mod-possessive
mod-rel	mod-rel	mod-rel
mod-quant	mod-quant	mod-quant
mod-dem		mod-dem
		mod-grade
		mod-neg
SPECIFIER (SPEC)		
spec	spec	spec
spec-poss		spec-poss
		spec-neg

A.1 English

A.1.1 Functions with English Examples

FUNCTION	ENGLISH EXAMPLE
FUNCTORS	
func	[*The firms*, with which she deals,] succeed. [He *adds* it together.] He'll [*be* succeeding]. She cannot have [*been* seen]. The manager, [in whom he *believes*], succeeds. [How *does* she come ?] [He *is* a manager.] [the *money*]
ARGUMENTS/COMPLEMENTS	
subj	*A great number of them* are new. *She and she* succeed. *He* alternates between them. *She and I* succeed. It is requested *that she leave*. The committee, *that* succeeds, comes.
obj	She calms *her* down. He hears *him* leave. The firm, *that* she accepts, comes. Who does he give *the proposal* to ?
i-obj	She asks *him* how he works.
p-obj(about) p-obj-loc	She argues *about her*. He walks *towards it*.
obj-pred	He regards him *as good*.
subj-pred-ap subj-pred-np subj-pred-pp	*How ill* is she ? *Who* is he ? She is *a manager*. He is *in a good mood*.

FUNCTION	ENGLISH EXAMPLE
ARGUMENTS/COMPLEMENTS (continued)	
n-comp	a great deal *of advice*
s-comp	It seems *that it is important*
v-comp	She cannot be *seen*. What manager does he *give the proposal to* ? She could *be succeeding*.
verb-particle	He sends her *away*.
MODIFIERS	
mod-loc-place	*Where* does she come ?
mod-temp-pos	*When* does he come ?
mod mod-rel	*How late* does she come ? the manager *whom he approves of*
SPECIFIERS	
spec spec-poss	*a few* firms *a* firm *Peter's* friend

A.1.2 Categories with English Examples

CATEGORY	ENGLISH EXAMPLE
ADVERB (ADV) & ADVERB PHRASE (ADVP)	
ADV	An employee, and *sometimes* the manager, succeeds.
ADVP	*How late* does she come ?
ADJECTIVE (ADJ) & ADJECTIVE PHRASE (AP)	
AP	*How ill* is he ?
AP_comp	a *more important* company
AP_sup	the *most important* company
ADJ	She is *new*.
ADJ_comp	a *newer* proposal
ADJ_sup	the *newest* proposal
COORDINATION (COORD)	
COORD_conj	He *and* you succeed.
COORD_disj	Either she *or* John is absent.
CORRELATIVES (CORREL)	
CORREL	*Both* he and the manager are new.
	Either she or they are new.
COMPLEMENTIZER (CPL)	
CPL(that)	He says *that* the manager succeeds.
CPL_adv(how)	*How* does she come ?
CLAUSE (C)	
C_sub-cpl(if)	He asks him *if it works*.
C_sub-cpl(that)-ind	It seems *that it is important*.
C_sub-cpl(that)-should	She requests *that she should leave*.
C_sub-cpl(that)-subj	He requests *that she leave*.
C_sub-cpl(whether)	She asks him *whether it works*.
C_sub-cpl-adv(when)	He asks him *when she works*.
C_sub-rel	The committee, *about which he argues*, succeeds.
C_sub-wh	She asks him *what he sees*.
	She asks him *who he sees*.
DETERMINERS (DET)	
DET	*another* firm
	last week
DET_def	*the* firm
DET_dem	*that* firm
DET_indef	*a* firm
DET_poss	*her* firm
DET_quant	*many* firms
DET_rel	*whose* firm
DET_wh	*which* firm
NEGATION (NEG)	
NEG_part	Can she *not* carry it ?

CATEGORY	ENGLISH EXAMPLE
NOUN (N) & NOUN PHRASE (NP)	
NP	*every firm*
NP_mass	He is given *money* for it.
NP_pl	*The committees* succeed.
NP_sg	*The committee*, after whom she goes, succeed.
N_pl	a few *firms*
N_sg	a *firm*
PREPOSITION (P) & PREPOSITIONAL PHRASE (PP)	
P	He depends *on* the manager.
PP	She explains *to him* that it works.
PP_slash-NP	What does he look *at* ?
PRONOUN (PRON)	
PRON_dummy	*It* is raining.
PRON_expl	*It* is requested that she should leave.
PRON_pers-2	*You* succeed.
PRON_pers-3-acc-sg-f	She calls *her* up.
PRON_pers-3-acc-sg-m	He asks *him* how she works.
PRON_pers-3-nom-pl	*They* succeed.
PRON_pers-3-nom-sg-f	*She* is compared with the manager.
PRON_pers-3-nom-sg-m	*He* comes.
PRON_pers-3-sg-n	She puts *it* by.
PRON_rel	the committee he sees
	The committee, *that* she accepts, comes.
	the committee *which* he sees
	The committee, *who* she accepts, come.
	The committee, at *whom* he aims, succeed.
PRON_wh-acc	*Whom* does she see ?
PRON_wh	*What* does he look at ?
PUNCTUATION (PUNCT)	
PUNCT(()	An employee (and sometimes the manager) succeeds.
PUNCT(,)	The committee, after whom she goes, succeed.
PUNCT(-)	An employee - and sometimes the manager - succeeds.
PUNCT(.)	A lot of them are new.
PUNCT(?)	How ill is he *?*
PUNCT([)	She says that she [the manager] succeeds.
PUNCT(')	He says the manager 'succeeds'.
SENTENCE (S)	
S	*She asks him who he sees.*
S_int	*Can she carry it ?*
S_int-wh	*When does he come ?*

CATEGORY	ENGLISH EXAMPLE
VERB PHRASE (VP)	
VP_bare-inf	He can *see him*.
VP_ing	She sees him *leaving the office*.
VP_perf	Has he *carried it* ?
VP_to-inf	She starts *to work*.
VERB (V)	
TO_aux	He ought *to* succeed.
V	She *abstains* from it.
V_aux-neg	He *cannot* be seen.
	Doesn't she carry it ?
V_aux	*Can* he carry it ?
	Does she carry it ?
V_copula	How ill *is* she ?
V_bare-inf	He can *see* him.
V_ing	He sees him *leaving* the office.
V_perf	Has he *carried* it ?
COMPARATIVE (COMP)	
COMP_np	He considers her *as a competitor*.
COMP_adj	She regards her *as good*.
COMP_vp	He considers him *as being good*.

A.2 French

A.2.1 Functions with French Examples

FUNCTION	FRENCH EXAMPLES
FUNCTORS	
func	L'ingénieur *travaille.* *Un litre* des vins
ARGUMENTS/COMPLEMENTS	
arg : arg-loc arg-dir arg-manner arg-temp	 L'ingénieur habite *ici.* L'ingénieur va *là-bas.* L'ingénieur travaille *beaucoup.* L'ingénieur a travaillé *toute la journée.*
subj subj-impers	*L'ingénieur* travaille. *Il* pleut.
obj	L'ingénieur accepte *le travail.*
p-obj : p-obj(à)-NP p-obj(de)-NP p-obj(pour)-NP p-obj(avec)-NP p-obj(contre)-NP p-obj(à)-inf p-obj(de)-inf p-obj(pour)-inf	 Le directeur parle *à l'ingénieur.* Le directeur parle *de l'ingénieur.* Le directeur demande une voiture *pour l'ingénieur.* Le directeur parle *avec l'ingénieur.* Les ouvriers étaient *contre le projet.* Le directeur demande *à parler.* Le directeur demande *de partir.* Le directeur demande une voiture *pour partir.*
s-comp v-comp	Il faut *que l' ingnieur vienne.* L'ingénieur sait *diriger l' entreprise* .
subj-pred : subj-pred-np subj-pred-adj subj-pred-pp	 L'ingénieur est devenu *professeur* . L' entreprise est *grande* . Le professeur apparaît *comme ingénieux* .
obj-pred : obj-pred-np obj-pred-adj obj-pred-pp	 Les ouvriers ont élu l'ingénieur *leur représentant syndical.* Le directeur considère l'ingénieur *très brillant.* Le directeur traîte l'ingénieur *comme son égal.*

MODIFIERS	
mod :	
mod-quant	*Toutes* les machines marchent.
mod-loc-dir	L'ingénieur veut aller *à Genève.*
mod-loc-ori	L'ingénieur vient *de Paris.*
mod-loc-place	L'ingénieur rencontrera le directeur *à l'usine.*
mod-loc-path	L'ingénieur ira *de Paris à Genève.*
mod-temp-duration	L'ingénieur travaille *toute la semaine.*
mod-temp-start	L'ingénieur travaille *depuis sept heures.*
mod-temp-end	L'ingénieur travaille *jusqu'à vingt heures.*
mod-temp-moment	La présentation commence *à seize heures.*
mod-manner	La nouvelle machine marche *sans problème.*
mod-instr	L'ingénieur ouvre l'enveloppe *avec des ciseaux.*
mod-cause	L'ingénieur vient *à cause du nouveau projet.*
mod-purpose	L'ingénieur vient *pour le nouveau projet.*
mod-poss	Le projet *de ce nouvel ingénieur* sera intéressant.
mod-rel	L'ingénieur *qui vient d'arriver* est très brillant.
SPECIFIERS	
spec	*Le* nouveau projet sera intéressant.
	Tous *ses* projets ont réussi.

A.2.2 Categories with French Examples

CATEGORY	FRENCH EXAMPLES
	SENTENCE (S)
(default) S	*L' ingénieur vient .*
S_imp	*Sors !*
S_int	*Viens - tu ?*
	CLAUSE (C)
(default) C_main	*Il a dit* qu'il viendrait.
C_cpl(que)	Il a dit *qu'il viendrait.*
C_cpl(si)	Il demande *si vous viendrez.*
C_wh	Il demande *qui vient.*
C_adv-cpl(lorsque)	*Lorsque l'entreprise ferme*, les machines s'arrêtent.
C_adv-cpl(quand)	*Quand l'entreprise ferme*, les machines s'arrêtent.
C_rel(qui)	L'entreprise *qui produit les roues* vient de fermer.
C_rel(que)	L'entreprise *que l'on vient de fermer* produisait du vin.
C_cond	*Si l'entreprise ferme*, les ingénieurs seront au chômage.
	VERB PHRASE (VP)
VP_inf	L'ingénieur veut *partir.*
VP_à-inf	L'ingénieur demande *à venir.*
VP_de-inf	L'ingénieur demande *de partir.*
VP_pour-inf	L'ingénieur demande une voiture *pour partir.*
VP_ppres	*Les voitures venant de la droite* ont priorité.
VP_ppres-infl	*Les chaises roulantes* y ont accès.
VP_ppast	L'ingénieur a *acheté un livre.*
VP_ppast-infl	*Les ingénieurs partis*, le directeur est sorti.
	VERB (V)
(default) V	Ils *viennent.*
V_refl	Ils *se lavent.*
V_copula	Ils *sont/semblent/paraissent* heureux.
V_aux	Ils *sont* venus. Ils *ont* marché.
VP_ppres	Les voitures *venant* de la droite ont priorité.
VP_ppres-infl	Les chaises *roulantes* y ont accès.
VP_ppast	L'ingénieur a *marché.*
VP_ppast-infl	Les ingénieurs *partis*, le directeur est sorti.
V_inf	Ils demandent à *venir.*
	NOUN (N) & NOUN PHRASE (NP)
NP	*La vieille entreprise*
NP_pn	*Pierre*
N	La vieille *entreprise*
N_mass	Le bon *vin*

PRONOUN (PRON)	
PRON_impers	*Il* pleut.
	Ce sont eux.
PRON_rel	L' ingnieur pose la question *qui* est intelligente.
PRON_wh	*Qui* vient ? *Lequel* vient ?
PRON_pers	*Il* vient.
PRON_pers_ton	Ce sont *eux.*
PRON_poss	Notre entreprise a fermé, *la leur* est toujours ouverte.
PRON_refl	Les ingénieurs *se* sont aperus de leur erreur .
PRON_reci	Les ingénieurs *se* sont rencontrés .
PRON_dem	*Celle-ci* vient.
PRON_indef	*Aucun* ne vient.
DETERMINER (DET)	
DET_def	*la* vieille entreprise
DET_indef	*une* entreprise
DET_poss	*son* entreprise
DET_part	*du* bon vin
DET_dem	*cette* entreprise
DET_quant	*plusieurs* personnes
DET_wh	*Quelle* entreprise ?
PREPOSITION (P) & PREPOSITION PHRASE (PP)	
P	*à* l'entreprise
P_det	*du* vin
PP	*dans les entreprises*
ADJECTIVE (ADJ) & ADJECTIVE PHRASE (AP)	
ADJ	Le travail *habituel.*
AP	des entreprises *assez anciennes*
AP_comp	des entreprises *plus anciennes*
AP_sup	Le travail *le plus habituel.*
ADVERB (ADV) & ADVERB PHRASE (ADVP)	
ADV	L'ingénieur a travaillé *efficacement.*
ADVP	L'ingénieur a travaillé *très efficacement.*
ADV_wh	*Comment* va Jean ?
COMPLEMENTIZER (CPL)	
CPL(que)	Il a dit *que* l'entreprise fermerait.
CPL(si)	Il demande *si* vous viendrez.
COORDINATION (COORD)	
COORD_conj	Le directeur *et* l'ingénieur viendront à la réunion.
COORD_disj	Le directeur *ou* l'ingénieur viendra à la réunion.

NEGATION (NEG)	
NEG(ne)	L'ingénieur *ne* travaille pas.
NEG(pas)	L'ingénieur ne travaille *pas*.
NEG(personne)	*Personne* ne travaille.
NEG(rien)	L'ingénieur ne voit *rien*.
COMPARATIVE PHRASE (COMP)	
COMP_adj	Le directeur considère l'étudiant *comme brillant*.
COMP_np	Le directeur considère l'étudiant *comme un débutant*.
PARTICLE (PART)	
PART_wh(est-ce-que)	*Est-ce-que* tu viens ?
PART_euph	A-*t*-il terminé?

A.3 German

A.3.1 Functions with German Examples

FUNCTION	GERMAN EXAMPLE
FUNCTORS	
func	Der Manager *schläft.*
	Der Manager glaubt, [daß der Student *schläft*].
	Der Manager wird [geschlafen *haben*].
	[Das alte *Projekt*] war erfolgreich.
	Der Manager hofft [*darauf*, den Präsidenten zu treffen].
	Der Manager schimpft [*über* den Präsidenten].
ARGUMENTS/COMPLEMENTS	
arg	Der Manager schimpft [über *den Präsidenten*].
arg-loc	Der Manager wohnt *dort.*
arg-dir	Der Manager eilt *dorthin.*
arg-manner	Der Manager wohnt *gut.*
	Der Test stimmt den Manager *nachdenklich.*
arg-temp	Der Vortrag dauert *den ganzen Tag.*
	Der Vortrag dauert *lange.*
subj	*Der Manager* schläft.
	Es regnet.
obj	Der Manager hält *den Vortrag.*
iobj	Der Manager dankt *dem Präsidenten.*
expl-topic	*Es* wird geschalfen.
subj-pred	Der Manager gilt *als Präsident.*
	Der Manager gilt *als fähig.*
	Der Manager tut *wie der Präsident.*
	Der Manager dient *als Präsident.*
subj-pred-np	Der Manager ist *der Präsident.*
subj-pred-adj	Der Manager ist *krank.*
subj-pred-pp	Der Manager ist *in der Stadt.*
	Der Manager ist *in Hochstimmung.*
obj-pred	Der Manager bezeichnet den Präsidenten *als Idioten.*
obj-pred-np	Der Manager nennt den Präsidenten *einen Idioten.*
obj-pred-adj	Der Manager nennt den Präsidenten *faul.*
obj-pred-pp	Der Manager vermutet den Präsidenten *in der Stadt.*
iobj-pred-adj	Dem Manager ist *schlecht.*
p-obj(zwischen)	Der Manager wählt *zwischen den beiden Tests.*
p-obj(auf)	Der Manager besteht *auf den Test.*
p-obj(über)	Der Manager schimpft *über den Präsident.*
p-obj(um)	Der Manager bittet *um den Test.*
p-obj-loc	Der Manager wohnt *in der Stadt.*
p-obj-dir	Der Manager legt den Bericht *auf den Tisch.*
p-obj-modal	Der Manager wohnt *in guten Verhältnissen.*
p-obj-temp	Der Manager zieht den Test *in die Länge.*
	Der Test dauert *bis morgen.*

FUNCTION	GERMAN EXAMPLE
ARGUMENTS/COMPLEMENTS	
s-comp	Der Manager glaubt, *daß der Präsident schläft.*
v-comp	Der Manager kann *programmieren.*
	Der Manager glaubt, *den Präsidenten zu kennen.*
MODIFIERS	
mod	Alle [*diese* meine] alten Projekte waren erfolgreich.
mod-quant	[*Alle* diese] Manager schlafen.
mod-neg	Der Manager kommt *nicht*
mod-dem	[*Diese* meine] Projekte sind erfolgreich.
mod-loc-dir	Der Manager begleitet den Präsidenten *in die Stadt.*
mod-loc-origin	Der Manager kommt *aus Saarbrücken.*
mod-loc-place	Der Manager schläft *in der Stadt.*
mod-loc-path	Der Manager fährt *über Saarbrücken nach Paris.*
mod-temp-duration	Der Manager schläft *den ganzen Tag.*
mod-temp-start	Der Manager schläft *seit 7 Uhr.*
mod-temp-end	Der Manager schläft *bis 18 Uhr.*
mod-temp-moment	Der Vortrag beginnt *um 15 Uhr.*
mod-manner	Der Manager lobt den Vortrag *in höchsten Tönen.*
	Das *alte* Projekt war erfolgreich.
mod-instr	Der Manager hält den Vortrag *mit lauter Stimme.*
mod-cause	Der Manager hält den Vortrag *aus Langeweile.*
mod-purpose	Der Manager hält den Vortrag *zu seinem Vergnügen.*
mod-rel	[Der Manager, *der schläft,*] ist kompetent.
mod-possessive	das Buch *Peters*
mod-grade	*nur* Tests
	der *besonders* wichtige Test
SPECIFIERS	
spec	*Der* Manager schläft.
	Alle diese *meine* alten Projekte waren erfolgreich.
spec-poss	*Peters* Freund
spec-neg	Der Manager sagt *nichts* Falsches

A.3.2 Categories with German Examples

CATEGORY	GERMAN EXAMPLE
SENTENCE (S)	
S	*Der Manager denkt an den Test.*
S_imp-v1	*Denke an den Test!*
S_int-wh-v2	*Wer kommt?*
S_int-v1	*Denkt der Manager an den Test?*
S_cond(wenn-dann)-v2	*Wenn der Manager kommt, dann schläft der Student.*
S_cpl-adv(als-da)-v2	*Als der Manager kam, da schlief der Student bereits.*
S_cpl(daß)-vfinal	*Daß der Student nur ja an den Test denkt.*
CLAUSE (C)	
C_cpl(daß)-vfinal	Der Manager glaubt, *daß der Student schläft.*
C_cpl(ob)-vfinal	Es ist fraglich, *ob der Student schläft.*
C_cpl-adv(als)-vfinal	*Als der Manager kommt,* schläft der Student.
C_cpl(wenn)-vfinal	*Wenn der Manager kommt,* schläft der Student.
C_decl-v2	Der Manager sagt, *Peter halte den Vortrag.*
C_wh-v2	Der Manager fragt: *Wer hält den Vortrag?*
C_wh-vfinal	Der Manager fragt, *wer den Vortrag hält.*
C_rel-vfinal	Der Manager, *der kompetent ist*, schläft.
C_imp-v1	Der Manager sagt: *Denke an den Test*!
C_cond-v1	*Wäre der Test gelungen*, hätte sich der Manager gefreut.
COMPLEMENTISER (CPL)	
CPL(daß)	Der Manager glaubt, *daß* der Student schläft.
CPL(ob)	Es ist fraglich, *ob* der Student schläft.
CPL_adv(wenn)	Der Student schläft, *wenn* der Manager kommt.
COORDINATION (COORD)	
COORD_conj	Der Manager *und* der Student schalfen.
COORD_disj	Der Manager schläft *oder* der Student schläft.
NEGATION (NEG)	
NEG_part(nie)	Der Manager schläft *nie.*
NEG_part(nicht)	Der Manager schläft *nicht.*
COMPARATIVE PHRASE (COMP)	
COMP_adj	Der Manager bezeichnet den Studenten *als faul.*
	Der Manager hält den Test *für gelungen.*
COMP_np	Der Manager bezeichnet den Studenten *als Idioten.*
	Der Manager tut *wie der Präsident.*
PARTICLE (PART)	
PART_grade	Der Manager liest *nur* den Bericht.
PART_mod	Der Manager liest *wohl* den Bericht.
PART_comp	Der Manager ist *so* fleissig *wie* der Student.

CATEGORY	GERMAN EXAMPLE
VERBPHRASE (VP)	
VP_pcp-bare-infl	Der *den Bericht lesende* Manager
VP_pcp-bare-uninfl	*Den Bericht lesend* geht der Manager hinein.
[1]VP_perf	Der Manager hat *den Bericht gelesen.*
[1] VP_perf(sein)	Der Manager ist *den Weg entlanggekommen.*
[1]VP_perf(passive)	Der Bericht wird *in Windeseile gelesen.*
	Der Bericht ist *in Windeseile gelesen.*
[1]VP_perf(sein/passive)	Der Bericht wurde *von dem Manager durchgegangen.*
VP_pcp-perf-infl(sein)	Der *soeben angekommene* Bericht
VP_pcp-perf-infl(passive)	Der *dem Manager gegebene* Bericht
VP_pcp-perf-uninfl(sein)	*Zu Hause angekommen* schläft der Manager.
[1]VP_sup-bare	Der Manager kann *den Test bestehen.*
	Der Manager wird *den Test bestehen.*
	Der Manager läßt *den Präsidenten schlafen.*
	Der Manager läßt *den Präsidenten den Vortrag halten.*
[1]VP_sup-bare(sein)	Der Manager kann *in einer Stunde ankommen.*
	Der Manager wird *in einer Stunde ankommen.*
	Der Manager läßt *den Präsidenten kommen.*
[1]VP_sup-bare(passive)	Der Manager läßt *den Vortrag halten.*
[1]VP_sup-bare(sein/passive)	Der Manager läßt *in der Halle antreten.*
VP_sup-zu	Der Manager versucht, *den Bericht durchzulesen.*
	Der Manager hat *den Bericht zu lesen.*
	[1]Der Manager scheint *den Bericht zu lesen.*
VP_sup-zu(sein)	Der Manager versucht, *rechtzeitig zu kommen.*
[1]VP_sup-zu(passive)	Der Bericht ist *in einer Minute zu lesen.*
[1]VP_sup-zu(sein/passive)	Jetzt ist *in den Saal einzutreten.*
VP_pcp-zu-infl(passive)	Der *dem Manager mitzuteilende* Bericht
VP_pcp-zu-uninfl(passive)	*Von niemandem zurückzuhalten* geht der Manager.
VP_cpl(um)	Der Manager kommt, *um den Bericht zu lesen.*
VERB (V)	
V_copula	Peter *ist* der Manager.
V_aux	Der Manager *hat* geschlafen.
V_mod	Der Manager *kann* schalfen.
V_pcp-bare-infl	Der *lesende* Manager
V_pcp-bare-uninfl	*Lesend* geht der Manager hinein.
V_perf	Der Manager hat den Bericht *gelesen.*
V_perf(sein)	Der Manager ist *gekommen.*
V_perf(passive)	Der Bericht wird *gelesen.*
	Der Bericht ist *gelesen.*
V_perf(sein/passive)	Der Bericht wurde *durchgegangen.*

[1]These VP's are only annotated, where the corresponding phrase is actually continuous, otherwise the category is left empty in the data. As a *v-comp* of raising verbs, *VP_sup-zu* also can be discontinuous.

CATEGORY	GERMAN EXAMPLE
VERB (V)	
V_pcp-perf-infl(sein)	Der *angekommene* Bericht
V_pcp-perf-infl(passive)	Der *abgegebene* Bericht
V_pcp-perf-uninfl(sein)	Zu Hause *angekommen* schläft der Manager.
V_sup-bare	Der Manager kann den Test *bestehen.*
	Der Manager wird den Test *bestehen.*
	Der Manager läßt *schlafen.*
	Der Manager läßt den Präsidenten den Vortrag *halten.*
V_sup-bare(sein)	Der Manager kann in einer Stunde *ankommen.*
	Der Manager wird in einer Stunde *ankommen.*
	Der Manager läßt den Präsidenten *kommen.*
V_sup-bare(passive)	Der Manager läßt den Vortrag *halten.*
V_sup-bare(sein/passive)	Der Manager läßt *antreten.*
V_sup-zu	Der Manager versucht, den Bericht *durchzulesen.*
	Der Manager hat den Bericht *durchzulesen.*
	Der Manager scheint den Bericht *durchzulesen.*
V_sup-zu(sein)	Der Manager versucht *durchzukommen.*
V_sup-zu(passive)	Der Bericht ist *durchzulesen.*
V_sup-zu(sein/passive)	Jetzt ist *einzutreten.*
V_pcp-zu-infl(passive)	Der *mitzuteilende* Bericht
V_pcp-zu-uninfl(passive)	Von niemandem *zurückzuhalten* geht der Manager.
PRONOUN (PRON)	
PRON_dummy	*Es* regnet.
	Es heißt, daß Peter kommt.
	Es graut Peter davor zu schalfen.
PRON_expl-topic	*Es* wird geschlafen.
	Es werden die Tests gemacht.
	Es glauben viele Manager an das Projekt.
PRON_expl-cor(relat)	*Es* ärgert mich, daß Peter kommt.
PRON_rel-nom-sg-masc	Der Manager, *der* den Test durchführt, kommt.
PRON_rel-acc-sg-neutr	Der Manager leitet das Projekt, *das* morgen beginnt.
PRON_wh	*Wer* kommt?
PRON_pers-2-sg	*Du* schläfst.
PRON_poss-1-sg-masc	*Meiner* schläft.
PRON_refl-3-sg-masc-acc	Der Manager ärgert *sich.*
PRON_reci-3-pl-masc-acc	Der Manager und der Student kennen *einander.*
PRON_dem-sg	*Jener* schläft.
PRON_indef-sg	*Jemand* schläft.
NOUN (N) & NOUN PHRASE (NP)	
N	Der alte *Manager*
N_mass	Der alte *Wein*
NP_nom-sg	*Der alte Manager*
NP_pn	*Peter*

CATEGORY	GERMAN EXAMPLE
DETERMINER (DET)	
DET_poss	*Sein* alter Bericht
DET_def	*Der* alte Bericht
DET_indef	*Ein* alter Bericht
	Einige alte/alten Berichte
DET_indef-adj	*Beide* alten Berichte
	Die *beiden* alten Berichte
DET_dem	*Dieser* alte Bericht
DET_wh	*Welcher* Manager schläft.
	Wieviele Manager schlafen.
ADJECTIVE (ADJ) & ADJECTIVE PHRASE (AP)	
AP	Der *mit Erfolg gekrönte* Test
ADJ	Das *rosa* Kleid
	Der Manager ist *krank*.
ADJ_e	Der *alte* Manager
ADJ_es	Ein *altes* Projekt
ADJ_er	Ein *alter* Manager
ADJ_em	*Altem* Wein
ADJ_en	Dem *alten* Manager
ADJ_comp_e	Der *erfolgreichere* Test
ADJ_sup_en	Die *erfolgreichsten* Tests
ADJ_dem	*Solche* alten Projekte
	Ein *solches* altes Projekt
ADVERB (ADV) & ADVERB PHRASE (ADVP)	
ADVP	Der *ganz besonders* fleißige Manager
ADV	Der Manager informiert *gründlich*.
ADV_pron	Der Manager denkt *daran*.
ADV_pron-cor	Der Manager denkt *daran*, den Bericht zu schreiben.
ADV_wh(an)	Der Manager weiß, *woran* er den Bericht erkennt.
ADV_reci	Der Manager heftet die Berichte *aneinander*.
ADV_grade	Der Manager ist *sehr* fleißig.
PREPOSITION (P) & PREPOSITION PHRASE (PP)	
P	*In* dem Bericht
P_det	*Im* Bericht
PP	*In dem Bericht*

Appendix B

Vocabulary

B.1 English Vocabulary List

V			
abstain (32)	do (42)	operate (24)	throw (9)
accept (12)	draw (40)	overhear (9)	urge (7)
account (32)	enjoy (9)	ought (35)	use (16)
add (41)	explain (23)	peer (32)	used (51)
aim (32)	fail (8)	permeate (8)	watch (9)
alternate (32)	feel (9)	practise (32)	will (43)
appeal (4)	find (16)	press (10)	work (89)
apply (9)	follow (10)	proceed (2)	would (39)
approve (59)	get (84)	promise (5)	
argue (32)	give (18)	pull (10)	
arrange (9)	go (192)	push (20)	
ask (17)	have (328)	put (9)	
bargain (8)	hear (9)	rain (6)	
battle (32)	help (26)	reach (6)	
be (766)	hire (16)	read (9)	
believe (32)	hope (15)	regard (8)	
bet (8)	know (26)	renew (1)	
bring (30)	lead (30)	request (13)	
call (10)	leave (143)	revolve (31)	
calm (10)	let (19)	say (2)	
can (52)	look (42)	seem (15)	
carry (8)	lose (32)	see (476)	
cause (2)	make (22)	send (20)	
close (6)	manage (4)	serve (32)	
come (119)	may (39)	set (10)	
compare (11)	might (39)	shall (35)	
consider (25)	miss (10)	should (43)	
convince (9)	must (39)	show (16)	
could (39)	need (30)	start (9)	
dare (40)	noise (11)	succeed (933)	
deal (35)	notice (8)	(to 87)	
depend (32)	observe (9)	tell (21)	

N	**DET**		**ADV**
'95 (1)	a (116)	past (3)	about (8)
Abrams (28)	all (47)	plenty (6)	abroad (8)
advice (10)	an (16)	several (3)	across (8)
answer (2)	another (3)	some (67)	ahead (8)
bookcase (10)	any (67)	such (44)	along (9)
Browne (21)	both (44)	that (64)	aloud (9)
building (8)	double (44)	the (1023)	apart (8)
Chiang (13)	each (64)	their (64)	around (8)
committee (271	either (64)	these (64)	approx. (1)
company (10)	enough (64)	this (64)	aside (8)
competitor (16	every (64)	those (64)	astray (8)
day (2)	few (12)	twice (44)	away (16)
employee (9)	first (3)	two (1)	back (9)
everything (6)	half (47)	what (109)	by (12)
firm (1427)	her (35)	whatever (64)	behind (8)
friendship (6)	his (64)	which (192)	down (9)
interview (13)	its (64)	whichever (64)	flat (1)
John (36)	last (3)	whose (74)	forth (12)
majority (2)	least (3)	whosever (64)	forward (8)
manager (339)	less (3)	your (64)	home (9)
money (70)	little (6)		in (1)
mood (2)	lot (5)		low (1)
news (9)	lots (3)		off (8)
number (5)	many (30)		on (12)
office (18)	more (5)		open (7)
patience (6)	most (5)		out (11)
PhD (2)	much (3)		over (11)
pound (8)	my (64)		round (9)
proposal (9)	next (3)		sometimes (6)
quantity (3)	neither (63)		through (8)
sacrifice (6)	no (64)		to (8)
Smith (3)	none (3)		today (5)
time (7)	one (10)		together (7)
week (1)	one-third (44)		up (16)
Wilson (3)	other (3)		under (7)
	our (64)		waste (2)
CPL	**NEG**	**CORREL**	**COORD**
if (12)	not (274)	both ... and (14)	and (296)
that (311)	(exclusive "-n't")	either ... or (14)	but (8)
how (3)		neither...nor (22)	or (23)
that (88)			
when (4)			
where (1)			
whether (8)			
why (9)			

ADJ	**PRON**	**ADV**	
good (22)	he (1413)	AD (1)	for (50)
great (8)	her (445)	about (37)	from (34)
important (40)	him (423)	across (29)	in (81)
large (3)	I (44)	after (55)	into (29)
new (80)	it (398)	against (29)	of (164)
true (5)	me (37)	alias (3)	off (32)
	she (1058)	among (29)	on (191)
	that (192)	around (24)	over (38)
	them (84)	as (20)	over with (7)
	there (10)	at (34)	round (29)
	they (78)	before (6)	through (29)
	us (39)	behind (25)	to (157)
	we (48)	between (29)	towards (29)
	which (75)	beyond (29)	under (33)
	who (204)	by (149)	upon (29)
	whom (186)	c/o (1)	with (49)
	you (53)	e.g. (1)	

B.2 French Vocabulary List

V			N
aborder (1)	espérer (1)	requérir (10)	âme (2)
accepter (688)	être (864)	rester (28)	accès (10)
acheter (83)	faire (82)	séparer (11)	accord (66)
agir (1)	falloir (66)	savoir (33)	aide (8)
aimer (12)	fonctionner (2)	sembler (6)	ami (16)
aller (25)	hésiter (6)	signer (9)	art (8)
apercevoir (4)	indiquer (1)	sortir (16)	avenir (12)
apparaître (18)	intéresser (32)	suffire (4)	bâtiment
appeler (6)	interroger (25)	suggérer (1)	bouquet (11)
apprendre (33)	laisser (2)	supposer (91)	cérémonie (2)
arriver (10)	lire (16)	terminer (30)	cause (22)
asseoir (2)	manger (2)	tomber (17)	chaleur (4)
avancer (11)	marcher (187)	toucher (4)	changement (4)
avertir (6)	mentir (1)	travailler (43)	ciel (1)
avoir (721)	mettre (2)	trouver (2)	collègue (8)
avouer (26)	montrer (76)	venir (974)	collaboration (6)
boire (4)	neiger (38)	vivre (3)	connaissance (8)
cesser (2)	obtenir (66)	voir (178)	conscience (2)
chanter (2)	offrir (27)	vouloir (60)	conseil (132)
combattre (24)	oser (2)		consultation (16)
commencer (20)	ouvrer (1)		contrat (14)
comparer (20)	ouvrir (1)		courage (8)
comprendre (8)	paraître (8)		crainte (22)
congédier (1)	pardonner (1)		décision (62)
connaître (5)	parler (170)		détermination (20)
constituer (9)	partir (108)		diable (1)
convaincre (10)	passer (24)		dieu (1)
courir (1)	penser (47)		directeur (75)
croire (4)	permettre (11)		discours (12)
décider (24)	persuader (7)		dommage (2)
démarrer (43)	pleuvoir (68)		eau (33)
demander (104)	poser (134)		échec (33)
devenir (20)	pouvoir (9)		entreprise (587)
devoir (61)	préférer (13)		erreur (4)
dire (15)	présenter (180)		espoir (18)
diriger (36)	prendre (2)		faillite (2)
discuter (16)	prononcer (12)		farine (10)
dissocier (49)	proposer (1)		femme (175)
diviser (2)	provoquer (12)		fille (2)
donner (51)	rêver (77)		fleur (4)
dormir (3)	raconter (1)		foule (8)
durer (4)	recevoir (5)		homme (251)
écouter (15)	regretter (19)		huile (4)
écrire (20)	remercier (103)		idée (18)
emporter (1)	remuer (4)		information (3)
entendre (5)	rencontrer (10)		ingénieur (1862)
entreprendre (4)	reposer (1)		intérêt (2)
entrer (13)	représenter (9)		invitation (68)

N	ADJ	ADV	PRON
Jean (50)	adroit (2)	assez (14)	auquel (12)
jour (4)	affectueux (4)	aujourd'hui (12)	ce (8)
kilo (6)	attentif (4)	aussi (2)	cela (1)
lait (8)	beau (92)	beaucoup (41)	duquel (12)
lieu (39)	bon (42)	bon (6)	elle (146)
litre (24)	capable (4)	certainement (39)	elles (126)
littérature (2)	civil (2)	comment (5)	en (78)
livre (88)	difficile (12)	correctement (1)	eux (40)
métier (8)	ennuyeux (4)	couramment (2)	il (689)
maison (18)	espagnol (2)	demain (58)	ils (134)
Marie (10)	évident (54)	fréquemment (23)	je (149)
midi (81)	excellent (6)	hier (37)	la (350)
minorité (2)	fréquent (3)	jamais (59)	laquelle (15)
moteur (18)	gentil (6)	maintenant (6)	le (400)
négociation (20)	grand (70)	où (9)	lequel (25)
nombre (10)	habituel (122)	peu (14)	les (100)
obligation (10)	heureux (4)	pourquoi (24)	leur (109)
ordinateur (6)	insupportable (4)	quand (9)	lui (159)
ordre (2)	intelligent (85)	que (3)	me (107)
participe (6)	irresponsable (1)	rapidement (60)	moi (69)
permission (6)	italien (2)	sûrement (2)	nous (208)
personne (32)	jaloux (4)	souvent (40)	on (22)
Pierre (6)	juif (2)	tant (8)	plusieurs (6)
Pierrette (4)	malade (3)	toujours (5)	qui (82)
plan (69)	meilleur (36)	très (41)	quoi (22)
plupart (14)	nécessaire (47)	trop (16)	se (67)
problème (24)	neuf (4)	volontiers (56)	te (164)
professeur (59)	nouveau (7)		toi (63)
proposition (8)	pire (22)		tu (142)
protestant (2)	possible (10)		vous (236)
quantité (6)	présent (2)		y (23)
question (190)	prêt (6)		
réunion (63)	rapide (3)		
regret (8)	rare (4)		
restaurant (48)	religieux (2)		
roi (2)	sensé (4)		
rupture (22)	volontaire (1)		
soir (81)			
soutien (20)			
spécialiste (42)			
sucre (6)			
travail (288)			
usine (2)			
vie (21)			
vin (180)			
voiture (11)			

DET	P	PART	CPL
aucun (4)	à (473)	est-ce que (49)	que(214)
aucune (2)	avant (20)	t (42)	si
ce (137)	avec (47)		
certain (8)	contre (3)		
certaines (2)	dans (32)		
certains (2)	de (503)		
ces (17)	en (34)		
cet (17)	entre (29)		
cette (25)	par (62)		
chaque (10)	pour (145)		
de (9)	sans (20)		
des (54)	sur (22)		
différentes (2)			
diverses (2)			
du (43)			
l' (1854)			
la (250)			
le (443)			
les (189)			
leurs (10)			
ma (12)			
maintes (2)			
mes (15)			
mon (36)			
nos (10)			
notre (10)			
nul (2)			
nulle (2)			
quel (24)			
quelle (22)			
quelles (22)			
quelque (8)			
quelques (18)			
quels (22)			
sa (41)			
ses (11)			
son (20)			
ta (9)			
tes (11)			
ton (14)			
tous (32)			
tout (44)			
toute (38)			
toutes (27)			
un (110)			
une (31)			
vos (10)			
votre (16)			

B.3 German Vocabulary List

V			N
abgeben (3)	gehören (3)	stammen (3)	Abfall (5)
achten (10)	gelten (6)	stecken (5)	Aktenkoffer (323)
anklagen (1)	gelingen (3)	stehen (5)	Amt (5)
ankommen (5)	genügen (8)	steigen (2)	Angebot (2)
ärgern (75)	glauben (279)	stimmen (31)	Art (2)
bedanken (2)	grauen (16)	streben (10)	Aufgabe (7)
bedürfen (6)	haben (151)	streiten (3)	Bedarf (2)
befestigen (5)	halten (15)	tanzen (2)	Berater (65)
begrüßen (75)	hängen (5)	testen (2)	Bericht (34)
berücksichtigen (3)	heißen (6)	treten (2)	Besuch (1)
bestehen (21)	hindern (14)	tun (11)	Beteiligten (12)
betrachten (14)	hoffen (5)	überlassen (5)	Beweis (2)
betrauen (5)	hören (6)	überraschen (7)	Computer (25)
betreten (1)	imponieren (3)	überreden (6)	Diebstahl (9)
betrügen (2)	interessieren (29)	untersagen (5)	Dutzend (2)
beurteilen (5)	kennen (21)	verdächtigen (2)	Erbsen (2)
bewahren (9)	kommen (180)	verdanken (5)	Forschung (4)
bezeichnen (8)	können (5)	verkaufen (377)	Frage (12)
bezichtigen (12)	lassen (12)	verlangen (3)	Geschäftsleitung (6)
bilden (2)	laufen (2)	vermuten (9)	Getreide (5)
bitten (9)	legen (14)	versprechen (6)	Gewinner (15)
bleiben (6)	lehren (13)	vertrauen (12)	Handvoll (4)
blicken (5)	lesen (5)	wagen (3)	Hans (5)
braten (1)	lieben (29)	wählen (7)	Idiot (6)
bringen (16)	liegen (23)	werden (548)	Jahre (2)
dauern (9)	machen (16)	wissen (8)	Kilogramm (6)
denken (4)	mitteilen (5)	wohnen (15)	Kind (20)
dienen (7)	mögen (5) (möcht_ (4))	wollen (9)	Kommilitone (432)
dürfen (60)	müssen (5)	ziehen (5)	Länge (3)
eilen (6)	nehmen (3)	zurücktreten (1)	Lehre (8)
einstellen (2)	nennen (5)	zusehen (2)	Manager (989)
empfehlen (6)	programmieren (2)	zwingen (10)	Mann (731)
entfallen (7)	reden (4)		Material (4)
entgehen (7)	regnen (23)		Mehrzahl (2)
entheben (5)	sagen (12)		Milch (5)
erinnern (14)	schenken (346)		Morgen (2)
fahren (24)	schimpfen (7)		Mutter (5)
fallen (3)	schlafen (574)		Nachbereitung (2)
fehlen (26)	schliessen (20)		Paar (2)
fliegen (2)	schmerzen (26)		Peter (38)
folgen (6)	schreiben (2)		Plan (38)
freuen (5)	schwindeln (25)		Platz (4)
frieren (6)	segeln (2)		Präsident (838)
funktionieren (6)	sehen (8)		Professor (318)
geben (36)	sein (326)		Projekt (1028)
gedenken (6)	sollen (5)		Rasen (1)
gefallen (16)	spielen (31)		Roman (432)
gehen (40)	sprechen (5)		Röslein (2)

N	ADJ	ADV
Saal (19)	alt (1914)	besonders (4)
Saar (2)	ernst (9)	dafür (12)
Saarbrücken (24)	falsch (3)	dagegen (2)
Storch (1)	freundlich (2)	damit (2)
Student (472)	gut (71)	danach (7)
Tag (6)	jung (7)	daran (21)
Test (269)	kompetent (5)	darauf (23)
Verhältnisse (2)	kühl (3)	darin (6)
Vorbereitung (2)	lang (2)	darum (2)
Vorfall (2)	leicht (9)	davor (9)
Wand (456)	müde (18)	dazu (7)
Wein (47)	notwendig (5)	dort (5)
Wert (3)	offen (5)	dorthin (2)
Zucker (6)	praktisch (4)	ebenfalls (2)
	rot (2)	ganz (2)
	schlecht (5)	genug (4)
	schön (2)	gestern (4)
	theoretisch (4)	morgen (7)
	trocken (21)	nie (29)
	viel (10)	niemals (50)
	wichtig (53)	soeben (3)
		später (4)
		wann (5)
		warum (3)
		weshalb (3)
		wie (40)
		wielange (3)
		wieso (3)
		wo (5)
		wofür (3)
		wogegen (1)
		womit (1)
		wonach (1)
		woran (3)
		worauf (1)
		woraus (1)
		worin (1)
		worüber (1)
		worum (1)
		wovon (1)
		wovor (1)
		wozu (1)

DET	P	PRO
all (37)	an (16)	denen (6)
beide (45)	auf (22)	deren (2)
dasjenige (6)	aus (18)	derer (6)
dasselbe (5)	bis (4)	dessen (4)
derjenige (7)	durch (2)	du (66)
derselbe (2)	für (27)	eins (2)
die/der/das (5909)	gegen (3)	er (204)
diejenige (4)	gegenüber (2)	es (184)
diese (485)	in (40)	etwas (2)
dieselbe (4)	mit (12)	ich (282)
ein (586)	nach (18)	ihm (23)
ein paar (2)	über (7)	ihr (34)
einige (34)	um (3)	irgendein (3)
etliche (38)	von (10)	irgendjemand (2)
jede (41)	vor (6)	jedermann (2)
jene (22)	während (1)	jemand (6)
kein (521)	zwischen (2)	man (2)
manch (42)		mir (1)
mehrer (28)		nichts (15)
mein (76)		niemand (6)
sein (1)		sich (18)
sämtliche (29)		sie (13)
solch (26)		was (27)
unsere (3)		wem (4)
viele (2)		wen (4)
welchem (1)		wer (5)
wenige (2)		wessen (2)
wessen (4)		wir (31)
wieviel (2)		
zwei (4)		

CPL	CONJ	PART
da (2)	nicht ... sondern (4)	als (24)
dann (2)	oder (46)	auch (4)
daß (392)	sowie (4)	denn (5)
je (2)	sowohl ...als auch (2)	doch (8)
ob (26)	und (607)	eigentlich (5)
wenn (9)	weder ...noch (2)	ja (5)
		mal (5)
		nicht (133)
		nur (3)
		schon (2)
		wohl (5)
		zu (119)

Bibliography

[Alvey (4),] Lexicon (online).

[Arnold *et al.*, 1994] Doug Arnold, Martin Rondell, and Frederik Fouvry. Design and Implementation of Test Suite Tools. Report to LRE 62-089 D-WP5.1, University of Essex, UK, 1994.

[Balkan *et al.*, 1991] L. Balkan, C. Chambers, J. Lindop, and al. Gb implementation report 1. Technical report, January - 30 November 1991.

[Balkan *et al.*, 1994a] Lorna Balkan, Siety Meijer, Kirsten Falkedal (principal authors), Doug Arnold, Dominique Estival, Sabine Lehmann, Sylvie Regnier-Prost, and Eva Dauphin. Test Suite Design — Guidelines and Methodology. Report to LRE 62-089 D-WP2.1a, University of Essex, UK, 1994.

[Balkan *et al.*, 1994b] Lorna Balkan, Siety Meijer, Kirsten Falkedal (principal authors), Doug Arnold, Dominique Estival, Sabine Lehmann, Sylvie Regnier-Prost, Eva Dauphin, and Klaus Netter. Issues in Test Suite Design. Report to LRE 62-089 D-WP2.1b, University of Essex, UK, 1994.

[Balkan *et al.*, 1995] Lorna Balkan, Frederik Fouvry, Doug Arnold, Sylvie Regnier-Prost, Eva Dauphin, and Veronika Lux. TSNLP User Manual. Volume 1: Test Suite Design and Methodology. Technical report, University of Essex, UK, 1995.

[Batori, 1975] I. Batori. Ein transformationelles modell für die koordination im deutschen. In I. Batori and al., editors, *Syntaktische und semantische Studien zur Koordination.* 1975.

[Bhatt, 1990] C. Bhatt. *Die syntaktische Struktur der Nominalphrase im Deutschen.* 1990.

[C. and Duttweiler, 1987] C. and G. Duttweiler. *Die 20000 Sätze und Ausdrücke der Handels- und Privatkorrespondenz.* Cosmos Verlag, Muri/Bern, 1987.

[Colliander, 1983] P. Colliander. *Das Korrelat und die obligatorische Extraposition.* 1983.

[Comrie, 1976] Bernard Comrie. *Aspect.* Cambridge University Press, 1976.

[Dauphin *et al.*, 1995] Eva Dauphin, Veronika Lux, Sylvie Regnier-Prost (principal authors), Doug Arnold, Lorna Balkan, Frederik Fouvry, Judith Klein, Klaus Netter, Stephan Oepen, Dominique Estival, Kirsten Falkedal, and Sabine Lehmann. Checking Coverage against Corpora. Report to LRE 62-089 D-WP3.2, University of Essex, UK, 1995.

[Drosdowski, 1984] G. Drosdowski, editor. *Duden. Grammatik der deutschen Gegenwartssprache.* 1984.

[Eisenberg, 1989] P. Eisenberg. *Grundriss der deutschen Grammatik.* 1989.

[Engel and Schumacher, 1976] U. Engel and H. Schumacher. *Kleines Valenzwörterbuch deutscher Verben.* 1976.

[Engel, 1970] U. Engel. Die deutschen satzbaupläne. In *Wirkendes Wort*, number 20, 1970.

[Engelen, 1975] B. Engelen. *Untersuchungen zu Satzbauplan und Wortfeld in der geschriebenen deutschen Sprache der Gegenwart. Verblisten.* 1975.

[Estival *et al.*, 1994a] Dominique Estival, Kirsten Falkedal, Lorna Balkan, Siety Meijer, Sylvie Regnier-Prost, Klaus Netter, and Stephan Oepen. Survey of Existing Test Suites. Report to LRE 62-089 D-WP1, University of Essex, UK, 1994.

[Estival *et al.*, 1994b] Dominique Estival, Kirsten Falkedal, Sabine Lehmann (principal authors), Lorna Balkan, Siety Meijer, Doug Arnold, Sylvie Regnier-Prost, Eva Dauphin, Klaus Netter, and Stephan Oepen. Test Suite Design — Annotation Scheme. Report to LRE 62-089 D-WP2.2, University of Essex, UK, 1994.

[Estival *et al.*, 1994b] Dominique Estival, Kirsten Falkedal, Sabine Lehmann, Hervé Compagnion (principal authors), Lorna Balkan, Doug Arnold, Frederik Fouvry, Eva Dauphin, Sylvie Regnier-Prost, Veronika Lux, Judith Klein, Judith Baur, Klaus Netter, and Stephan Oepen. The Construction of Test Material. Report to LRE 62-089 D-WP3.1, ISSCO, Université de Genève, Switzerland, 1995.

[Goetze, 1979] L. Goetze. *Valenzstrukturen deutscher Verben und Adjektive.* 1979.

[Grevisse, 1991] Maurice Grevisse. *Le Bon Usage. Grammaire française avec des Remarques sur la langue française d'aujourd'hui.* Duculot, Paris, 12ème edition, 1991.

[Gross, 1975] Maurice Gross. *Méthodes en syntaxe. Régime des constructions complétives.* Hermann, Paris, 1975.

[Gru, 1987 and 1990] Deutsch-english and deutsch-französisch, 1987 and 1990.

[Hammond *et al.*, 1988] L. Hammond, J. Lindop, E. Pollard, and al. Gb implementation report 1. Technical report, July - 31 December 1988.

[Helbig and Buscha, 1988] G. Helbig and J. Buscha. *Deutsche Grammatik. Ein Handbuch für den Ausländerunterricht.* 1988.

[Helbig and Schenkel, 1980] G. Helbig and W. Schenkel. *Wörterbuch zur Valenz und Distribution deutscher Verben.* Max Niemeyer Verlag, Tübingen, 1980.

[Höhle, 1991] T. Höhle. On reconstruction and coordination. 1991.

[Hornby *et al.*, 1985] A. S. Hornby, A. P. Cowie, and A. C. Gimson. *Oxford Advanced Learner's Dictionary of Current English (OALD).* Oxford or OALD 3rd edition on line, 18th edition, 1985.

[Huddleston, 1984] R. Huddleston. *Introduction to the Grammar of English.* Cambridge University Press, 1984.

[Kohrt, 1976] M. Kohrt. *Koordinationsreduktion und Verbstellung in einer generativen Grammatik des Deutschen.* 1976.

[Maier and Steffens, 1989] Petra Maier and Petra Steffens. Zur syntax pränominaler elemente in einer kategorialen unifikationsgrammatik des deutschen. IWBS Report 70, Institut für Wissensbasierte Systeme, Stuttgart, 1989.

[McCawley, 1988] James D. McCawley. *The Syntactic Phenomena of English*, volume 1 and 2. University of Chicago Press, Chicago, 1988.

[Oepen *et al.*, 1995a] Stephan Oepen, Klaus Netter, Judith Baur, Tom Fettig, Judith Klein, and Fred Oberhauser. The TSNLP Database. From tsct to tsdb. Report to LRE 62-089 D-WP6.1, DFKI Saarbrücken, Germany, 1995.

[Oepen *et al.*, 1995b] Stephan Oepen, Klaus Netter, Tom Fettig, Fred Oberhauser, Judith Baur, and Judith Klein. TSNLP User Manual. Volume 2: Test Suite Technology. Technical report, DFKI Saarbrücken, Germany, 1995.

[Oepen *et al.*, 1996] Stephan Oepen, Klaus Netter, and Judith Klein. TSNLP — Test Suites for Natural Language Processing. In John Nerbonne, editor, *Linguistic Databases*, CSLI Lecture Notes. Center for the Study of Language and Information, 1996. forthcoming.

[Quirk *et al.*, 1991] R. Quirk, S. Greenbaum, G. Leech, and J. Svartvik. *A Comprehensive Grammar of the English Language.* Longman, 1991.

[Reis and Vater, 1980] M. Reis and H. Vater. “beide”. In G. Brettschneider and C. Lehmann, editors, *Wege zur Universalienforschung.* 1980.

[Reis, 1982] Marga Reis. “zum subjektbegriff im deutschen”. In W. Abraham, editor, *Satzglieder im Deutschen.* 1982.

[Renz, 1989] I. Renz. Koordination von nichtverbalen satzkonstituenten. Technical report, IBM-IWBS Report 74, 1989.

[Soames and Perlmutter, 1979] Scott Soames and David Perlmutter. *Syntactic Argumentation and the structure of English.* University of California Press, Berkely – Los Angeles, 1979.

[Sommerfeldt and Schreiber, 1983a] Sommerfeldt and Schreiber. *Wörterbuch zur Valenz und Distribution deutscher Adjektive.* Max Niemeyer Verlag, Tübingen, 1983.

[Sommerfeldt and Schreiber, 1983b] Sommerfeldt and Schreiber. *Wörterbuch zur Valenz und Distribution deutscher Substantive.* Max Niemeyer Verlag, Tübingen, 1983.

[Svartvik, 1985] J. Svartvik. *A Comprehensive Grammar of the English Language.* Longman, 1985.

[van Oirsouw, 1985] R. van Oirsouw. A linear approach to coordinate deletion. In *Linguistics*, number 23, 1985.

[Vater, 1986] H. Vater. *Zur Syntax der Determinatien.* 1986.

[Weisgerber, 1983] M. Weisgerber. *Valenz und Kongruenzbeziehungen.* 1983.