

# ACTIVITY REPORT 1980 - 1990 FOR EUROTRA-NL

(version 1509v2)

Steven Krauwer and Louis des Tombe

November 9, 1993

# Contents

<b>1</b>	<b>Genesis of Eurotra-NL</b>	<b>1</b>
1.1	Prehistory . . . . .	1
1.2	The Contract of Association . . . . .	2
1.3	The research center . . . . .	4
1.4	The team . . . . .	4
<b>2</b>	<b>Results</b>	<b>7</b>
2.1	Research . . . . .	7
2.1.1	General system design and specifications . . . . .	7
2.1.2	The MiMo systems . . . . .	8
2.1.3	Contributions to mainstream Eurotra research . . . . .	9
2.1.4	Lexical issues . . . . .	9
2.1.5	Language specific and contrastive research . . . . .	10
2.2	Implementation . . . . .	10
<b>3</b>	<b>Other actions</b>	<b>12</b>
3.1	Scientific and educational . . . . .	13
3.1.1	Publications . . . . .	13
3.1.2	Presentations . . . . .	13
3.1.3	Workshops . . . . .	13
3.1.4	Working papers and journals . . . . .	13
3.1.5	Special events . . . . .	14
3.1.6	Courses in Computational Linguistics . . . . .	14
3.1.7	Rosetta . . . . .	14
3.2	Organisational . . . . .	14
3.2.1	Local integration . . . . .	14
3.2.2	National . . . . .	15
3.2.3	At the project level . . . . .	15
<b>4</b>	<b>Future</b>	<b>15</b>
4.1	Contract of Association . . . . .	15
4.2	Shared cost activities under the 1991-92 programme . . . . .	16
4.3	Lexical issues . . . . .	16
4.4	Industrial contacts . . . . .	16
4.5	Collaboration with Canada . . . . .	16
<b>5</b>	<b>Lessons learned</b>	<b>16</b>
5.1	Centralised/planned vs decentralised/liberal . . . . .	16
5.2	Ambitions vs realism . . . . .	17
5.3	Continuity . . . . .	17
5.4	Expectations 1991 - 2000 . . . . .	18
5.5	Needs . . . . .	18
<b>6</b>	<b>Epilogue</b>	<b>19</b>
<b>7</b>	<b>References</b>	<b>20</b>

## Introduction

This Activity Report of the NL Language Group covers the whole period of Dutch involvement in Eurotra from late 1980 until the end of 1990.

The report is structured as follows:

The first section describes the genesis of the NL team, and contains some factual data concerning signature of contracts, financial provisions and size of the team.

The next section gives an overview of the research and implementation results obtained. Given the scope of this report no detailed information on the content of the various research activities is provided, but an effort has been made to give a clear picture of the research profile of the NL group, and to provide the interested reader with extensive bibliographic references to scientific publications or internal Eurotra Reports authored or co-authored by (ex-)members of the team. The results of implementation work are given in a brief overview.

Section 3 is dedicated to other activities than those following directly from the execution of the Contract, both at the scientific level (publicity, connections with the scientific community, etc), and at the organisational level.

In section 4 a short overview is given of activities in progress or in preparation for 1991, 1992 and later.

Section 5 has as its title 'Lessons learned', and tries to identify a number points to be kept in mind for future programmes and projects and for the creation of future EC and national MT platforms.

A short epilogue (section 6) presents a few final, evaluative remarks.

Section 7 is the bibliography.

Annex I lists all members of the NL team between September 1986 (signature of the first Contract of Association) and December 1990.

A full account of the grammatical and lexical coverage of the modules produced by the NL/B team is provided in Annex II, which is extracted from the Final Implementation Report of the NL/B Eurotra group.

## 1 Genesis of Eurotra-NL

### 1.1 Prehistory

When the Netherlands got involved in Eurotra around 1980, Computational Linguistics as such was already a well-established activity in research and education at Dutch universities.

Most literary faculties had regular courses in programming and literary and linguistic computing (e.g. indexes and concordances, statistical approaches, morphological and syntactic parsing).

Even Machine Translation was already under active exploration (Rosetta – Philips Research Labs, Eindhoven), or in preparation (DLT – BSO, Utrecht).

In 1980 the founding members of Eurotra approached researchers from the Technical University of Delft (A.G. Sciarone) and the University of Utrecht (S. Krauwer) in order to initiate Dutch participation. Until that moment the interests of the Dutch language had been taken care of by researchers from the Catholic University of Leuven.

During the preparatory phase of the project, before the Council Decision in 1982, the main instruments for the Commission were the so-called study contracts, made with the participating centers.

Some of the series of contracts were language based, and some were oriented towards specific linguistic topics or towards system design in general.

From 1981 until 1984 work on the study contracts concerning the Dutch language was done on a collaborative basis between Leuven, Delft and Utrecht. This series of contracts (ET-7, ET-10, ETL-4, ETL-7) provided the centers with the resources necessary to create and maintain a small core team, consisting of senior researchers (involved on a part-time basis) and advanced students. In addition this created the possibility to start building a modest infrastructure.

Participation in topic or design oriented contract work took place on a personal basis, and here the Utrecht team took part in semantic research (Des Tombe, contract ET-10SEM), software specifications (Krauwer, contracts ETS-1, ETS-3, ETS-6), linguistic specifications (Des Tombe, ETL-3, ETL-5, ETL-8), and framework design (Krauwer, contract ETS-9; Krauwer, Van Noord, Des Tombe, contract ETL-9).

Throughout this period efforts were made to establish a joint Eurotra center for Belgium and Holland in order to consolidate the excellent working relationships between the Dutch and Belgian groups, and to have a good starting point for collaborative work after signature of the Contracts of Association. Unfortunately this ideal had to be abandoned when it turned out that it would take the Dutch government much longer to prepare for the signature of the contract than the Belgian government (who signed in 1984).

After 1984, when most governments had signed the Contracts and most national groups had been established and had become operational, the language specific study contracts came to an end, and one of the consequences was that the Delft center became more and more detached from the project in the absence of any active involvement or financial resources. Utrecht suffered to a lesser extent from this discontinuity problem, because of their continuing involvement in the system specification and design activities, which led to a series of contracts allowing for a small core team to be maintained.

## **1.2 The Contract of Association**

In the course of 1986 the University of Utrecht -after consultation with the national funding authorities- established the Stichting Taaltechnologie (Foundation of Language Technology) which would act as the contract partner on behalf of the Dutch government.

National funding was channeled via NBBI, a foundation, dealing with Libraries and Infor-

mation Services, created and funded by the ministries of Education and Economic Affairs. When the Contract was signed, in September 1986, the following financial provisions were made:

- Since the Netherlands and Belgium share the Dutch language, work on the Dutch language and financial provisions would be divided at a 2:1 ratio, conforming to common practice in the Dutch Language Union, based on a bilateral agreement between Belgium and the Netherlands dealing with activities concerning Dutch language and literature.
- The Eurotra group in the Netherlands was cofinanced by the Commission and the Dutch government on a 60:40 basis.
- Originally the overall budget per language was 3000 KECU, resulting in 2000 KECU for the Netherlands and 1000 KECU for Belgium. Of the total amount of 2000 KECU the Commission would contribute 1200 KECU, and the Dutch government 800 KECU.

During the execution of the work some adaptations were made to the original budget. Spain and Portugal joined the EC, and the project was extended. Furthermore special arrangements were made between EC and the Dutch government in order fund the first phase of the LEXIC project (cf. section 2.1.4. below) within the framework of the Contract of Association. And finally, some extra payments were made by the CEC for involvement of members of the NL team in central operations, and for the organisation of the Annual Eurotra Workshop 1990 in Noordwijkerhout.

The following table summarizes the financial provisions made in the Contract of Association and the two addenda.

<i>Contract</i>	<i>Duration</i>	<i>Amount</i>	<i>CEC</i>	<i>NL</i>
ET-CA-NL	Sept 86 – June 88	1200	720	480
ET-CA-NL/Addendum 1	July 88 – June 89	600	360	240
Central Operations	July 88 – June 89	30	30	
ET-CA-NL/Addendum 2	July 89 – Dec 90	945	567	378
Annual Workshop	May 90	60	60	

The task of the group (to be carried out together with the Leuven group) was the following:

- Production of the Dutch monolingual modules of the system (i.e. analysis and synthesis and dictionaries).
- Production of all transfer components from all other EC languages into Dutch.
- Monolingual and contrastive research.
- The creation of a powerful and competent center for Natural Language Processing.

### 1.3 The research center

The Stichting Taaltechnologie, which acted as the contract partner, was established by the University of Utrecht in order to further the interests of natural language processing in general, and machine translation in particular.

A "Eurotra research unit" was established as a subpart of the Research Institute for Language and Speech of the faculty of humanities of the University of Utrecht, and contractual arrangements were made between the Stichting and the University.

The unit was headed by S. Krauwer and L. des Tombe, both members of the permanent staff of the Linguistics Department and the Research Institute for Language and Speech.

Members of the team were recruited via informal circuits and open job advertisements, from all over the country. They were all employed by the University of Utrecht. Since work on the Dutch language was already distributed over two centers, it was not judged desirable to aim at further decentralization.

Since neither the faculty nor the university as a whole could offer any state-of-the-art infrastructure (UNIX equipment, networking facilities, electronic mail, on site library facilities), the center had to build up its own infrastructure. By the end of 1990 the center was well-equipped, with RISC based workstations, faculty network facilities (based on the initial ethernet structure created by the Eurotra group) connected to the national and international academic networks, and an up-to-date documentation center.

Although the creation of the Eurotra team in the Netherlands was the direct motive to create the research center, it was the intention from the very outset the intention to set up a structure (comprising both people and facilities) that could serve as a basis for a variety of future actions in the field of natural language processing.

### 1.4 The team

The initial team was formed by the core team that had been created and maintained on the basis of the study contracts. Gradually, the team was extended, and people from Utrecht and elsewhere were attracted. There were a number of reasons why the growth of the team was rather gradual, and the center was not fully staffed until the end of 1989:

- First of all it seemed unwise from a management point of view to employ too many people who were basically untrained at the same time.
- Secondly, around the same time when Eurotra started in the Netherlands public funding had been made available to not only Eurotra, but also to Rosetta (Philips) and DLT (BSO). Furthermore a number of chairs in computational linguistics were created at various Dutch universities, and as a consequence it turned out far from easy to attract new staff.
- A third factor was that, due to the late signature of the Dutch contract, the Leuven group had already selected English and German from the set of foreign languages to deal with, which left the Utrecht group with the lesser known languages (as seen from a Dutch perspective, at least): French, Danish, Italian, Greek, Spanish and

Portuguese. Traditionally Dutch computational linguists have as their background Dutch, English or general linguistics, whereas most students of the other languages tend to have a strong preference for literary studies. As a result it was not always easy (and sometimes even impossible) to recruit people with sufficient knowledge of both the various languages and of computational or even general linguistics.

Over the years staff turnover has been very modest.

When Spain and Portugal joined the project two other projects were started: LEXIC (in collaboration with Philips and Van Dale, cf. 2.1.4), which was incorporated in the Contract of Association, and GRAMMAR (with Tilburg University, cf. 2.1.5).

Below an overview of the composition of the team over the years expressed in terms of man months per quarter. Figures for LEXIC and GRAMMAR are included. Support staff includes management, administration, computer officers.

Quarter	MM Linguistic staff			MM Support staff			MM TOTAL
	ET	GRAMM	LEXIC	ET	GRAMM	LEXIC	
1986-4	5			3			8
1987-1	5			4			9
-2	15			4			19
-3	19			6			25
-4	23			7			30
1988-1	23			8			31
-2	27			8			35
-3	32			8			40
-4	46			7			53
1989-1	40	4		10	1		55
-2	47	7		10	1		65
-3	48	7		9	1		65
-4	43	7		10	1		61
1990-1	60	8		9	1		78
-2	59	8		9	1	1	78
-3	51	9	12	9	1	2	84
-4	46	10	14	8	1	2	81
<b>Total</b>	<b>589</b>	<b>60</b>	<b>26</b>	<b>129</b>	<b>8</b>	<b>5</b>	<b>817</b>

Projection for the first quarter of 1991:

1991-1	15	10	14	3	1	2	45
--------	----	----	----	---	---	---	----

See Annex I for a list of members of the NL team over the years.

After expiration of the Contract of Association, work on LEXIC and GRAMMAR will continue in 1991. The new Contract of Association will be of a much more modest size than before, and lead to a reduction of the Eurotra team working under this contract to ca 30% of its current size (and the whole team to ca 50%).

As a consequence of this reduction ca 12 members of the team had to find other employment. Most of them have succeeded and have found jobs in private companies or universities (e.g. in Germany and Switzerland).



## 2 Results

The achievements of the NL group can be subdivided into two main subclasses: research and implementation.

### 2.1 Research

The NL group has carried out a number of research activities over the last few years, in part following directly from the requirements of the Eurotra programme as laid down in the Contract of Association, in part following from a more general interest in laying the grounds for the development of a proper technology for MT.

Research activities will be divided into 5 categories:

- i General system design and specifications
- ii The MiMo systems
- iii Contributions to mainstream research
- iv Lexical issues
- v Language specific and contrastive research

#### 2.1.1 General system design and specifications

##### 2.1.1.1. Software specifications

During the preparatory years the Utrecht group has taken an active part in the design of the system and the specifications of the software. The NL share of this work was done by S. Krauwer under special study contracts between EC and the University of Utrecht, and resulted in a number of technical reports for the EC, the last of which was Final Report ETS-6 on the Software Specifications in 1985 (Johnson et al. 1985b), and a number of scientific publications on software design for MT systems (e.g. Johnson et al. 1984).

##### 2.1.1.2. Linguistic specifications

At the same time the Utrecht group started to participate in the production of the linguistic specifications for Eurotra. This work was carried out by L. des Tombe, under special study contracts. The result was a series of technical reports, the last of which was Final Report ETL-8 on the Eurotra Linguistic Specifications (Arnold et al. 1985), and scientific publications (e.g. Johnson et al. 1985a, Van Eynde et al. 1985).

##### 2.1.1.3. Framework design

The period between 1984 and 1987 can best be described as a period of strong convergence between linguistic and software design activities. For a while there was a very close and fruitful cooperation involving people from Manchester, Essex, Geneva and Utrecht. This work resulted in the  $\langle C, A \rangle - T$  model, which has as its main characteristics (i) compositionality, (ii) declarativeness, and (iii) lexicalism. This model was based, amongst

other things, on research into the nature of transfer in MT systems (cf. Krauwer & Des Tombe 1984abc, 1985). The  $\langle C, A \rangle - T$  model is described in a number of internal reports to the EC (e.g. King & Krauwer 1987a), and in scientific publications (Arnold et al. 1985, Arnold & Des Tombe 1987, Arnold et al. 1986, Des Tombe 1986, Des Tombe et al. 1985).

Unfortunately, the resulting model was never actually used in Eurotra. Most of the people involved disappeared from the project; the institute in Geneva stopped cooperation altogether. The  $\langle C, A \rangle - T$  model had a restricted amount of influence on the development of what was to become the official Eurotra software. More directly, this work was taken up by some of the Eurotra centers as a basis for so-called 'sidelines': MiMo1 in Essex and Utrecht, CAT2 in Saarbruecken, and MiMo2 in Utrecht.

#### 2.1.1.4. Reference manual

The convergence between formalism and linguistics was consolidated by the creation of the Eurotra Reference Manual, appearing for the first time in January 1986 and edited (until its 6th release) by M. King and S. Krauwer (King & Krauwer 1986abcd, 1987ab).

### 2.1.2 The MiMo systems

#### 2.1.2.1. The MiMo translation system

Early 1987 Utrecht discontinued participation in the further development of the mainstream Eurotra framework, and concentrated (in collaboration with Eurotra-Essex and ISSCO (Geneva)) on a locally developed direct descendant of the  $\langle C, A \rangle - T$  framework, leading to the production of a small prototype translation system, called MiMo, later that year. The system was capable of translating between Dutch, English, German and Spanish (in all directions), at reasonable speed. Although the lexical coverage was small, its linguistic coverage was gradually extended to cover a fair subset of the languages involved. This research is documented in a number of technical reports and publications (Van Noord et al. 1989, Van der Eijk 1989, Van Noord 1989d, Dorrepaal & Van Noord 1990).

#### 2.1.2.2. The MiMo2 translation system

In 1988 a new prototype translation system was designed, the MiMo2 system. The main objective was to experiment with modern insights in Computational Linguistics (since the project showed a clear reluctance to import or even explore new developments in Computational Linguistics), and to build a demonstrable MT system. Like its predecessor it was based on declarativeness and compositionality. It was sign-based, and one of its formal features was reversibility. The underlying formalism belongs to the PATR family. Since on the basis of present knowledge MT systems are only feasible when their translation task is limited in a number of aspects, it was decided to implement a small system for a limited subdomain and a limited text type: international teletext news. The system has a small lexicon, but extensive grammatical coverage for Dutch, English and Spanish (translating in all directions). Since it has no world knowledge component it yields all linguistically possible translations.

This research has led to a variety of publications, focusing on a number of topics: Gen-

eration (Van Noord 1989ab, 1990a, Shieber et al. 1989, 1990), Decidability (Van Noord 1989d), and Reversibility (Van Noord 1990b). A general description of the design is given in Van Noord (1989c), and linguistic aspects in Dorrepaal & Van Noord (1990b), Dorrepaal (1990ab), Van Noord & Van der Eijk (1990), Van Noord et al. (1990b), Van Noord et al. (forthcoming).

### **2.1.3 Contributions to mainstream Eurotra research**

In the meantime Utrecht continued participation in mainstream linguistic research aiming at improvement or extension of the existing linguistic specifications. Main areas of interest: Negation (Hoekstra 1989, 1990ab), Morphology (Verheul in Ananiadou et al. 1989, Van Gaalen in Ananiadou et al. 1990a), Discourse (Dorrepaal et al. 1990), Interlevel Phenomena (Dijenborgh in Syea et al. 1990).

In addition some other exploratory studies took place: Preference (Petitpierre et al. 1987), and Sublanguages (Pohlmann in Ananiadou et al. 1990b).

The problem of word order and discontinuities was taken up by a special interest group on linear order, and a number of internal reports were produced and bundled in Van Noord et al. (1990a).

### **2.1.4 Lexical issues**

Over the years Utrecht has developed a strong interest in Lexical Issues.

Initially most lexical work was done in the context of the Dictionary Task Force, via membership of L. des Tombe and P. ten Hacken. This has led to a number of technical reports (e.g. Erlandsen et al. 1982), contributions to the reference manual (Ten Hacken in Valentini et al. 1990), and scientific publications (e.g. Ten Hacken 1989, 1990).

On the basis of this interest the group started to explore the possibility of designing a neutral dictionary. A first result was *Ndict*, a neutral Dutch dictionary that could serve as the basis for the Eurotra Dutch dictionaries, and that could remain invariant under modifications in the Eurotra linguistic specifications. This work has been reported upon in e.g. Van der Eijk & Van der Wouden (1989), Van der Eijk (1990) and Bloksma et al. (1990).

A second result was a strategy to extend an existing monolingual dictionary mechanically by application of closure operations (connected to the sublanguage) and filtering of the results via existing lexical databases (cf. report by Van der Wouden in Krauwer 1990, section 2.1.1.0).

A third result of *Ndict*, made possible by the extension of the overall project volume as a consequence of Spain and Portugal's joining the project was the LEXIC project, aiming at the production of a reusability methodology for lexical resources, to be implemented by means of production of multi purpose Spanish-Dutch and Dutch-Spanish dictionaries for Eurotra and Rosetta (the Philips MT project). A project definition study was financed by NBBI, and conducted in close collaboration between STT, Philips and Van Dale Lexicography. The resulting project was split up in a feasibility phase and a production phase.

The feasibility phase was started in July 1990, and financed by NBBI, Philips, Van Dale and EC. According to the original planning the production phase would have started in July 1991 (provided the feasibility study yields a positive result), but due to unforeseen delays in the startup of the 3rd Framework Programme (the most likely source of possible EC contributions) the second phase will have to be delayed, even though all other partners have already made their financial commitments.

### **2.1.5 Language specific and contrastive research**

During the preparatory phase of Eurotra (1980-1984) the Utrecht group has carried out a number of linguistic studies in close collaboration between Utrecht, Delft and Leuven.

A variety of topics were covered, and reported on in Technical Reports: Coordination and Comparison (Van Eynde et al. 1981), Transfer (Van Eynde et al. 1982, Belder et al. 1983), Time, Modality and Semantic Relations (Belder et al. 1984).

After signature of the Contract of Association language specific research continued, in part directly connected to implementation work (problem solving, the results of which are reflected in the implementation and in the general implementation reports rather than in explicit reports), and in part in a more general sense, e.g. 'er' or DCDs (cf. Dijenborgh, resp. Pohlmann, both in Krauwer (1990)).

Apart from these, another research activity was started, aiming at the creation of reusable grammatical resources, very similar to the objectives of the LEXIC project. Also in this case NBBI provided financial support for a project definition study, and as a result the GRAMMAR project was created, financed by NBBI and carried out in collaboration between STT and Tilburg University. The project will finish mid 1991. The objective is to design a prototype database system containing grammatical descriptions of a number of linguistic phenomena for a limited number of languages (Dutch, English, Spanish, Portuguese). A number of intermediate reports have been produced (Groos et al. 1989, 1990abc), and two scientific publications (Kester 1990, 1991). A larger scale follow-up project is in preparation. The general approach to reusability of grammatical resources can be found in Van Riemsdijk & Des Tombe (forthcoming).

## **2.2 Implementation**

One of the major tasks under the Contract of Association has been the production of analysis and synthesis modules for Dutch, and transfer modules from all other eight EC languages into Dutch. Throughout the duration of the project this part of the work has been carried out in close collaboration with the Leuven group.

Division of labour:

Dutch Analysis and Synthesis: Utrecht and Leuven. Dutch dictionary: Utrecht. Transfer from Danish: Utrecht. Transfer from German: Leuven. Transfer from English: Leuven. Transfer from Spanish: Utrecht. Transfer from French: Utrecht. Transfer from Greek: Utrecht. Transfer from Italian: Utrecht. Transfer from Portuguese: Utrecht.

The work has been reported on in joint Implementation Reports to the Commission (Des Tombe 1988, Krauwer 1988, Dijenborgh & Krauwer 1989 and 1990, and finally Krauwer 1990).

The modules and reports delivered have been critically reviewed by the Commission (e.g. Havenith & Groenez, 1990; Havenith 1991), with a very positive evaluation as a result.

Annex II contains a copy of the Coverage section of the Final Implementation Report (Krauwer 1990).

Below we present an impressionistic overview of the results.

Grosso modo one could say that the current (i.e. end of 1990) implementation (especially the monolingual part) covers a substantial part of the subset of the Dutch language that can be described within the limits imposed by the linguistic specifications and the formalism. This grammatical coverage is complemented by a large dictionary (20000 uninflected entries), and together they constitute a system capable of handling a fair subset of the Dutch language, provided it can be run in an appropriate soft- and hardware environment.

In this context it should be noted that the project has never succeeded in providing appropriate software. The official Eurotra software has been a problem throughout the project. It is unbearably slow, making it impossible for the linguists to properly test their grammars and dictionaries.

As for the transfer components it should be clear that they are necessarily much more modest in grammatical and lexical coverage, and in quality as well, for the following reasons:

- a Necessarily transfer components are always at least one step behind the monolingual components, since transfer work can only be done on the basis of analysis and synthesis components that remain unchanged throughout the implementation period.
- b The coverage of transfer components is necessarily based on the intersection of the coverage of source and target language components, only the latter of which is under the target language group's own control.

The result can be described as three fairly sized translation systems (in terms of grammatical coverage): English  $\rightarrow$  Dutch, German  $\rightarrow$  Dutch and Dutch  $\rightarrow$  English. It has to be added that a large investment in testing and debugging is still necessary. Given the state of the official software, testing is extremely time consuming. Porting to other software is impossible: the state of the art in Computational Linguistics is not yet such that there are standards for 'reusability' of linguistic knowledge.

For the other six languages there are small translation systems, all of them translating from text to text.

In order to give an impression of the current performance of the system (limited to the monolingual Dutch components) a summary of test results is presented below.

A 'test suite' was constructed. This is a list of sentences and ungrammatical word strings, especially designed to check whether a system does what it is expected to do. This particular one was based on a test suite used for parsing of English at Hewlett-Packard

(Flickinger et al, 1987), and adapted to the Eurotra coverage definition and to Dutch as the input language.

The object of the tests were all and only the Dutch modules (i.e. analysis and synthesis). Per sentence, the result of the analysis process was given to synthesis as input, so that a Dutch sentence was the final result.

Output was judged OK if the output sentences were grammatical and acceptable as paraphrases of the input sentence, or if ungrammatical input resulted in zero translations.

Output was judged bad if ungrammatical input resulted in any output (grammatical or ungrammatical), if grammatical input resulted in no output (for whatever reason), if the output was ungrammatical, or if the output was grammatical but not acceptable as a paraphrase of the input.

Output was judged OK and bad if some of the output was OK and the rest of the output was bad.

Overview of statistics:

- 701 sentences within the scope of the coverage were tested
- the output of 372 sentences was judged OK
- the output of 289 sentences was judged bad
- the output of 40 sentences was partly OK and partly bad

A few comments on the figures are in place here.

First of all, the official Eurotra software that should be used to perform these tests requires considerable computer and human resources: running this test set once takes nearly 3 days with 7 powerful (16 MIPS) workstations running in parallel, the total accumulated CPU time amounting to ca 500 hours (in an environment where the user is charged for the use of CPU time at e.g. 0.10 ECU per CPU second, the cost of one single test run would amount to 180000 ECU). This imposes heavy constraints on the frequency and the volume of testing one can afford.

Second, the test set is specifically constructed to spot weaknesses in the system, i.e. each sentence contains a real problem, as opposed to arbitrary sentences taken from running text, some of which may contain only problems, and some none. The obvious advantage is a reduction of the overall testing volume.

Third, it has to be kept in mind that the results presented here originate from the very first time the test suite as a whole was run through the system as it was delivered at the end of 1990, i.e. no debugging had been done yet. A first global analysis of the errors encountered suggests that a large proportion of the problems is relatively easy to solve.

The Final NL/B Implementation Report (December 1990, ca 800 pages) gives a full account of implementation work done and results obtained (Krauwier 1990).

### **3 Other actions**

Apart from the activities directly following from the tasks assigned to the NL group under the Contract of Association, a number of other activities have taken place, which for

various reasons deserve to be mentioned here explicitly.

## **3.1 Scientific and educational**

### **3.1.1 Publications**

In addition to the publications related directly to the execution of the Eurotra programme a number of other scientific papers have been published, such as Des Tombe & Warwick (1990), Van Gaalen (1990), Van der Wouden (1989) and Van der Linden & Van der Wouden (1990a) on various lexical issues, Van der Wouden (1990b) on semantics, and Ruessink (1991) on morphology.

General publications on machine translation include e.g. Neijt & Hoekstra (1986). An overview of the current Eurotra formal and linguistic framework is given in Raw et al. (1989).

### **3.1.2 Presentations**

Members of the NL team have on various occasions reported on their work at major CL conferences (COLING, American and European ACL), and other scientific conferences. At the national level numerous presentations of the project have been given for different expert and non-expert audiences.

### **3.1.3 Workshops**

From 1980 the NL group has been represented in all annual Eurotra workshops, and in various smaller scale workshops organised by the project. During all these years members of the NL group have played an active part, by presenting talks and chairing sessions.

In 1990 the NL group hosted the last annual workshop in Noordwijkerhout, with ca 160 participants from the Eurotra community, and three invited speakers from outside (J. Hobbs, H. Kamp, and W. Wahlster).

### **3.1.4 Working papers and journals**

The NL group has participated (and is still participating) in a number of publicity actions emerging from the project. S. Krauwer is one of the founding editors of the *Studies in Machine Translation and Natural Language processing* (published by the Office for Official Publications of the EC in Luxemburg). C. Verheul is one of the volume editors of the volume on morphology in this series. S. Krauwer is one of the guest editors of the special issue on Eurotra of the journal *Machine Translation* (Allegranza et al, forthcoming). T. van der Wouden is one of the editors of the linguistic journal *GLoT*. Leuven and Utrecht have created their joint *Working Papers in Natural Language Processing*, edited by F. Van Eynde and P. ten Hacken. Eight volumes have been published.

### **3.1.5 Special events**

The 1989 GLOW conference took place in Utrecht, and it was preceded by a 2 days special event dedicated to Computational Linguistics, organised by the Research Institute for Language and Speech. L. des Tombe was member of the programme committee. Speakers included prominent researchers in Natural Language Processing such as S. Pulman (SRI Cambridge, UK), B. Ingria (BBN, Cambridge MA, USA), P. Isabelle (CWARC, Montreal, Canada) and J. Tsujii (UMIST, Manchester, UK).

T. van der Wouden organised, together with E.J. van der Linden (Institute for Language Technology and Artificial Intelligence, University of Tilburg) a two day workshop on Computers and the Lexicon. There were ca 80 attendants from The Netherlands and Belgium. A selection of the papers read was published in a special issue of the journal TABU (Van der Linden & Van der Wouden 1990, Van der Wouden and Van der Linden 1990).

In October 1990 the Utrecht group organised a one day conference called Computational Linguistics in the Netherlands, with ca 180 participants from all over Holland and Belgium, with Mark Johnson (Brown University, visiting professor in Stuttgart) as invited speaker. This event was judged extremely successful by the participants, and it will be repeated this year. The proceedings will be published in the course of 1991.

### **3.1.6 Courses in Computational Linguistics**

Although the research team as such did not have any teaching obligations, many of the members taught courses that were part of the CL curriculum of the department of linguistics, via part-time appointments on a temporary basis.

### **3.1.7 Rosetta**

Throughout the duration of the programme there have been close connections between the Eurotra team and the members of the Rosetta team. Not only is part of the linguistic work for the Rosetta project contracted out by Philips to the STT (and carried out by employees of the Research Institute for Language and Speech), but also at the more personal level there have been many interconnections, notably the appointment of J. Landsbergen, leader of the Rosetta project, as part-time professor of CL in the research institute.

## **3.2 Organisational**

### **3.2.1 Local integration**

The underlying philosophy of the local organisation has been to aim at integration into the standard university environment, i.e. the Research Institute for Language and Speech and the Department of Linguistics. For this reason a number of personal unions were established (e.g. heads of the group only working for Eurotra on a part time basis, members of the group employed as teachers on part-time basis), and practical measures taken



(e.g. contractual arrangements concerning the physical location of the group in the direct neighbourhood of the Research Institute).

Organisationally the research team is part of the Research Institute, although the actual tasks to be executed by the team are determined by the contractual obligations rather than by the research programme of the Research Institute.

### **3.2.2 National**

Some exploratory activities have taken place in view of the establishment of a national MT (or even wider) platform, and although the research community shows a considerable amount of interest (cf. the success of the CLIN event above), there is very little response from possible funders of such a platform (industry, ministries).

### **3.2.3 At the project level**

#### **3.2.3.1. NL/B**

Since the Dutch language is shared between the Netherlands and (part of) Belgium, one language group was created, with two centers, one in Leuven and one in Utrecht. Work and finance were divided at a 2:1 ratio.

Although it was unfortunate that it took the Dutch authorities two years longer than the Belgian government to make proper financial provisions, the two centers have been working closely together from the very start.

After signature of the Contract of Association the two groups decided to act as one single group which was reflected by the submission of joint programmes of work and reports. This collaboration has been very successful and will be continued for the new contract period.

#### **3.2.3.2. Liaison group**

From the very beginning Utrecht has played an active role in the Liaison Group, the managing body of the project. S. Krauwer, permanent representative of the NL team in the LG since the signature of the Contract of Association, has been a member of various ad hoc working groups and more permanent committees, such as the Internal Assessment Committee, and the so-called KMPP committee dealing with the planning of future activities (1991-92, 3rd Framework Programme).

## **4 Future**

### **4.1 Contract of Association**

It is the intention of the NL group to continue to participate in the 1991-92 programme. Unfortunately the transition from the old to the new programme will reduce the funding level for the NL team by ca 60%.

## **4.2 Shared cost activities under the 1991-92 programme**

The NL group is preparing a number of proposals in reaction to the Call for Proposals for shared cost research under the 1991-92 programme, but given the extremely small amount of money available for this action there is little chance that this will contribute more than marginally to the continuity of the team. The proposed research areas are: collocations, NP-interpretation, discourse, sublanguage and ambiguity, MT theory, reusability of grammatical resources. The last mentioned project would be the continuation of the GRAMMAR project described in 2.1.5.

## **4.3 Lexical issues**

Even if it has to be expected that there will be a gap in time and finance between the first and the second phase of the LEXIC project (aiming at a reusable Spanish-Dutch and Dutch-Spanish dictionary), the group is still determined to find the funding necessary for the continuation, possibly under the 3rd Framework Programme, or under any other appropriate national or community programme.

## **4.4 Industrial contacts**

A number of contacts with industry have been established, in order to explore the possibility of collaborative MT actions. Apart from the notable exceptions Philips and BSO, both of which have been involved in MT for years now, there seems to be little inclination to get actively (and financially) involved at this moment. At the same time it has to be noted that many companies are commercially involved in smaller scale activities (i.e. dictionaries on CD ROM, pocket size bi- or even multilingual dictionaries, on-line dictionaries for PCs etc). This fact should not be ignored when establishing short and longer term national and EC policies for MT.

## **4.5 Collaboration with Canada**

The group has been approached by CWARC (Canada) in order to explore a possible joint project concerning robustness in MT systems, within the framework of Canadian-EC collaboration.

# **5 Lessons learned**

## **5.1 Centralised/planned vs decentralised/liberal**

The Eurotra project is unique in the world, both in size and ambition. The involvement of a large number of scientists from all member states has had a tremendous effect: it has encouraged those groups and people who were already involved in MT to share their knowledge and expertise with many others, and it has led to the creation of new centers of activity in all member states. This would never have happened if the project had not been

based on equal treatment of all official languages. Yet one can wonder -in retrospective- whether the interpretation that has been given to this notion of equality ("all language groups do the same work in parallel") was the most beneficial option for the project as a whole. Where the obvious advantage is that a large amount of convergence is ensured and that there can be no doubt as to the equal distribution of the workload, the obvious disadvantage is that the same problems are encountered, struggled with, solved and sometimes not solved by many teams in parallel. Especially where research activities are concerned better and more results might have been obtained by letting various small task forces tackling in parallel a whole range of problems, from a variety of angles. The monolithic approach adopted in the project might have been justified for a pure development project, but in the case of Eurotra it may have caused a large degree of underexploitation of intellectual resources, and a considerable degree of scientific frustration in the Eurotra community.

The lesson to be learned from this is that future large scale projects will have to pay more attention to the diversification of roles and tasks that is necessary to maximize the output, both quantitatively and qualitatively.

## **5.2 Ambitions vs realism**

No doubt the project has suffered from the strong tension between research and what is usually called development. The project may have been conceived as a development enterprise, with a certain research aspect, but already after a few years it became clear that a massive research effort was (and still is) necessary before fully automated high quality machine translation can be seen as a feasible enterprise. Eurotra is not alone in discovering this: many other MT system designers have arrived at the same conclusion. What is distressing is that it has turned out to be impossible for the project to face this reality and adapt the objectives of the project accordingly. Where the project -because of its size- had a unique opportunity to either concentrate on research or reorient itself towards more feasible variants of MT systems (controlled language, interactive, restricted domain and text type), thus exploiting the intellectual resources at its disposal, the rigidity of the goals has caused the project to spend much of its manpower on development type work based on immature research results, and therefore of no practical future use.

The lesson to be learned from this is that future R&D programmes in highly unexplored areas such as MT should have a large amount of built-in flexibility in order to allow for continuous adaptation to new insights and developments.

## **5.3 Continuity**

The NL group has suffered severely from the continuity problem. In the early years one of the participating groups (Delft) disappeared because of lack of continuity in funding, whereas Utrecht could survive the gap between startup of the project and signature of the contract on the basis of personal contracts that happened to exist. Right now the NL group is facing other delays, because of late signature of the Contract of Association (at least 4 months, possibly more), late startup of shared cost projects under the 91-92 programme (at least 9 months, probably 12 to 15), and late start of the 3rd Framework

Programme (possibly 1 or 2 years). All these delays may be natural, inevitable or even desirable from the Commission point of view, they are disastrous for the teams that are financially dependent on these programmes.

Since there is little reason to expect or even hope that this situation will ever change, it should be clear that a solution to this problem is needed in order to ensure some continuity in the area of MT basic and applied research. Here lies an important role for the national authorities, who could provide funding not only for specific projects, but also for preparation of new projects and bridging funding gaps between programmes.

#### **5.4 Expectations 1991 - 2000**

When looking at the results emerging from the Eurotra programme and other MT activities in the world around us, it cannot be denied that considerable progress has been made. Yet it has to be acknowledged that the more one knows about MT the more one realizes that there is still a long way to go to fully automated high quality translation. Care should be taken not to fall into the trap of confusing quality and cost-effectiveness: even if (in specific environments) MT systems can be operated on a cost-effective basis (and even more so in 5 or 10 years time when hardware price-performance ratios will become still more favourable), a massive research effort is needed in order to make MT systems qualitatively comparable to the human translator. In this respect there is little reason to expect any major breakthrough before the end of the century. In the meantime only a limited class of potential users will be interested in this type of MT (i.e. poor quality at low cost), whereas many others will find the quality unacceptably low for their purposes.

This should not be interpreted as to mean that the whole MT enterprise should be abandoned for the time being, or left to research institutions. On the contrary, on the basis of the present state of knowledge and technology it should be possible to make high quality MT systems for restricted applications, provided the notion of quality is not derived from a comparison with the human translator, but rather from comparison between a functional specification of the task at hand and the actual performance of the system. Such systems could be restricted in many ways, e.g. only applicable to small domains, functioning as a translator's aid, operating in interaction with the author of the text to be translated, operating as a multilingual text generator rather than a full translation system, etc.

#### **5.5 Needs**

For the longer term basic research in MT proper and related areas (varying from system architecture to knowledge representation) will have to continue. Where universities tend to derive their research actions from their teaching activities and industries from their profit expectations, it should be clear that special efforts at national and EC level are necessary in order to ensure funding for MT and related research. In order to make optimal use of these public funds, a transnational network of centers of expertise should be created and maintained, preferably coordinated by means of an EC agency.

For the shorter term, where the aim should be the design and production of restricted systems, activities should concentrate on the following areas:

- identification of restricted translation tasks, the functional specification and design of systems capable of executing these tasks, and methods to evaluate these systems;
- the definition and creation of generic lexical and grammatical resources;
- interfacing linguistic knowledge, domain specific knowledge, and the user.

Here again it should be stressed that public funding will be necessary, but in this case an additional effort should be made in order to obtain industrial support as well, since for restricted systems the gap between what is possible now and commercialisable products seems to be much smaller than for high quality fully automated systems.

Special attention should be paid to the position of the so-called *less favoured languages*, since without special measures it seems obvious that industrial interest (and investment) will focus on those languages that offer the best market perspectives.

## 6 Epilogue

In the preceding sections we have been rather critical about a number of aspects of the content and the organisation of the Eurotra project. This critical attitude should not be mistaken for a negative attitude.

We are critical because in our view Eurotra is not just an episode which is about to be concluded now, and which can be evaluated in terms of its own objectives and achievements alone.

Eurotra (the programme, not the system) should be seen as one out of a conglomerate of current and future actions, all directed towards the creation of a European Language Industry. Seen from this point of view the crucial question to ask when assessing Eurotra is not whether or not we have been able to build the prototype MT system that was announced in the Council Decision from November 1982 (as a matter of fact we haven't), but rather to what extent the project has helped to establish a better starting point for future actions.

And here we feel that Eurotra has indeed made an enormous contribution. We have scientific results (in the form of insights, publications) that we didn't have before, we have grammars and dictionaries that we didn't have before, and we have managed to create a European network of people from academia and industry with expertise in a large number of areas in the field of natural language processing.

But at the same time it cannot and should not be ignored that our assets for the future are not made up of positive insights and experiences alone. We have lost some illusions concerning the complexity of our task and this should help us in defining and delimiting better subtasks for future enterprises. We have learned how a large army of enthusiastic scientists can become ineffective or even totally paralysed when e.g. tasks are too complex, strategies not clear, or decision taking not efficient, and this experience should be used to help us in developing better plans and better organisational structures for the future.

## 7 References

Note: All internal (mostly technical) Eurotra reports are marked ETR (Eurotra Technical Report). If followed by a number this number refers to the Eurokom number. Other alphanumeric codes refer to contracts.

Allegranza, V., S. Krauwer, E. Steiner (eds) (forthcoming): Special Eurotra Issue of "Machine Translation" (to appear).

Ananiadou, S., L. De Wachter, A. Ralli, C. Verheul, A. Villalva (1989): Morphology, ETR, 50 p.

Ananiadou, S., A. Blatt, M. Carulla, L. De Wachter, M. van Gaalen, D. Maas, V. Pirrelli, J. Provoost, A. Ralli, A. Schoenmakers, A. Villalva (1990a): Word Structure, ETR, 220 p.

Ananiadou, S., R. Pohlmann, E. Quinlan, J. Simcoe-Shelton (1990b): Sublanguage, ETR, 113 p.

Arnold, D., L. des Tombe, L. Jaspaert, R. Johnson, S. Krauwer, M. Rosner, N. Varile, S. Warwick (1985a): 'A MU-1 View of the  $\langle C, A \rangle, T$  framework in Eurotra.' In: Proceedings of the Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages, pp. 1-14.

Arnold, D., Jaspaert, L., Des Tombe, A.L. (1985b): Eurotra Linguistic Specifications, ETR ETL-8, 300 p.

Arnold, D., Krauwer, S., Rosner, M., Des Tombe, L., & Varile, N. (1986): 'The  $\langle C, A \rangle, T$  framework in Eurotra: a theoretically committed notation for MT.' In: Proceedings of Coling, Bonn, 1986, pp. 297-303.

Arnold, D., & Des Tombe, A.L. (1987): 'Basic theory and methodology in Eurotra', in: S.Nirenburg (ed): Machine translation, Cambridge Univ. Press, pp. 114-135.

Arnold, D., Krauwer, S., Sadler, L., & Des Tombe, L. (1988): 'Relaxed Compositionality in MT' In: Proceedings of the 2nd International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages, Carnegie Mellon Univ, Pittsburgh, 1988, p. 61-81.

Belder, S., F. van Eynde, S. Krauwer, A. Neijt, B. Sciarone, L. des Tombe, H. Tonino (1984): Transfer: categorial labels, safety nets and strategies, ETR ETL-1-NL/B, 61 p.

Belder, S., F. van Eynde, L. Jaspaert, S. Krauwer, F. Maes, A. Neijt, F. Steurs, L. des Tombe, H. Tonino (1984): Time, modality and semantic relations, ETR ETL-4-NL/B, 151 p.

Bloksma, L., A. van Bolhuis, P. van der Eijk, P. ten Hacken, J. Herklots, D. Heylen, H. Pijnenburg, F. Sesink, A. Teeuw, L. des Tombe, T. van der Wouden (ed) (1990): Ndict Final Report, Internal report, July 1990, 105 p.

Dijenborgh, B.-J., & S. Krauwer (eds) (1989): Eurotra NL/B Implementation Report, ETR, Utrecht/Leuven, June 1989, 167 p.

Dijenborgh, B.-J., & S. Krauwer (eds) (1990): Eurotra NL/B Implementation Report,

- ETR, Utrecht/Leuven, January 1990, 200 p.
- Dorrepaal, J., E. Efthimiou, B. Ripplinger, W. Ramm, P. Schmidt, J. Schuetz (1990): Final Report on Research in Discourse Structure, ETR.
- Dorrepaal, J. & G.J. v. Noord (1990): Anaforische Relaties in het Automatisch-vertaalsysteem MiMo. In: A. Neijt en D. Bakker (eds.), Computerlinguïstiek, Foris Publications, Dordrecht, 1990.
- Dorrepaal, J. (1990a): A linguistic Account of Discourse Anaphora. Working Papers in Natural Language Processing, Katholieke Universiteit Leuven, Stichting Taaltechnologie Utrecht, 8, 1990
- Dorrepaal, J. (1990b): Discourse Anaphora. In: Proceedings of COLING-90, Helsinki, 1990, pp 95-99.
- Eijk, W.F. van der (1989): "Linguistic analysis in the MiMo translation system", Working Papers in Natural Language Processing, Katholieke Universiteit Leuven, Stichting Taaltechnologie Utrecht, 3, 1989, also in G. de Haan, W. Zonneveld (eds), Formal Parameters Yearbook IV, Utrecht 1989.
- Eijk, W.F. van der, & T. van der Wouden (1989): A Modular Lexicon Architecture for NLP. To appear in: U. Zernik, ed.: Proceedings of the First International Lexical Acquisition Workshop, AAAI Press.
- Eijk, W.F. van der (1990) "The Ndict formalism for Computational Lexicography". Paper presented at the Eurotra Annual Workshop, Noordwijkerhout, 1990.
- Erlandsen, J., F. van Eynde, J. McNaught, H. Somers, L. des Tombe, Y. Wilks (1982): Dictionary and semantics in Eurotra, ETR ET-10-SEM, 104 p.
- Eynde, F. van, Des Tombe, A.L., & Maes, F. (1985): 'The specification of time meaning for machine translation.' In: Proc. Second Conference of the European Chapter of the Ass. for Computational Linguistics, p. 35-40.
- Eynde, F. van, L. Jaspaert, S. Krauwer, A. Neijt, B. Sciarone, L. des Tombe (1981): Coordination and Comparison, ETR ET-7-NL/B, 196 p.
- Eynde, F. van, J. Honig, L. Jaspaert, S. Krauwer, A. Neijt, B. Sciarone, L. des Tombe, D. Vaassen (1982): The task of transfer vis-a-vis analysis and generation, ETR ET-10-NL/B, 108 p.
- Flickinger, Dan, John Nerbonne, Ivan Sag, and Tom Wasow (1987) "Toward evaluation of NLP systems." 25th Annual Meeting of the Association for Computational Linguistics, 1987.
- Gaalen, M. van (1990): "Een objectgericht georganiseerd lexicon" in TABU, jaargang 20, nummer 2, 1990, pp.95-106
- Groos, A., N. Corver, E.-P. Kester, R. Vos (1989): Eurotra-Grammar, First biannual report, Utrecht, August 1989.
- Groos, A., N. Corver, E.-P. Kester, R. Vos (1990a): Eurotra-Grammar, Second biannual report, Utrecht, February 1990.

- Groos, A., N. Corver, J. Kamperman, E.-P. Kester, R. Vos (1990b): Eurotra-Grammar, Third biannual report, Utrecht, June 1990.
- Groos, A., N. Corver, J. Kamperman, E.-P. Kester, R. Vos (1991): Eurotra-Grammar, Fourth biannual report, Utrecht, January 1991.
- Hacken, Pius ten, (1989): "Reading distinction in machine translation", Working Papers in Natural Language Processing, Katholieke Universiteit Leuven, Stichting Taaltechnologie Utrecht, 6, 1989
- Hacken, Pius ten (1990): Reading Distinction in MT, in Proceedings of COLING-90, Helsinki, 1990, pp 162-166.
- Havenith, R., D. Groenez (1990): Review of the NL/B Implementation Report and of the Actual Implementation, ETR, Luxemburg, January 1990.
- Havenith, R. (1991): Review of the B/NL end-of-year delivery, ETR, Luxemburg, April 1991.
- Hoekstra, H., (1989): 'Negation and Quantification'. A paper presented at the LSA Institute, Tucson (Arizona), August 1989.
- Hoekstra, H. (1990a): Eurotra Reference Manual 7.0, chapter B.I.3.4 ('Negation'), ETR.
- Hoekstra, H. (1990b): Final Report of the Research on Formal Semantics, ch. 3 ('Negation'), ETR.
- Johnson, R., Krauwer, S., Rosner, M., Varile, N. (1984): 'The design of the kernel architecture for the Eurotra software' In: Proceedings of COLING 84, Stanford University, Palo Alto, 1984, p. 226 - 235.
- Johnson, R., King, M., Des Tombe, A.L., (1985a): 'Eurotra: a multilingual system under development.' Computational Linguistics, vol 11, nr. 2-3, p. 155-169.
- Johnson, R., Krauwer, S., Rosner, M., Varile, N. (1985b): Software Specifications, ETR ETS-6, (136 p).
- Kester, E.-P. (1990) Adnominal participial constructions in Germanic and Romance languages, Recherches de Linguistique Francaise et Romane d'Utrecht, IX, p. 75-85.
- Kester, E.-P. (1991) Adnominal Infinitives in Germanic and Romance languages, Recherches de Linguistique Francaise et Romane d'Utrecht, X, p. 13-23.
- King, M., Krauwer, S. (eds)(1986a): The Eurotra Reference Manual 1.0, ETR, January 1986.
- King, M., Krauwer, S. (eds)(1986b): The Eurotra Reference Manual 1.1, ETR, March 1986.
- King, M., Krauwer, S. (eds)(1986c): The Eurotra Reference Manual 1.2, ETR, July 1986.
- King, M., Krauwer, S. (eds)(1986d): The Eurotra Reference Manual 2.0, ETR, October 1986.
- King, M., Krauwer, S. (eds)(1987a): The Eurotra Reference Manual 2.1, ETR, January 1987.



- King, M., Krauwer, S. (eds)(1987b): The Eurotra Reference Manual 3.0, ETR, January 1987.
- Krauwer, S., & Des Tombe, A.L. (1984a): 'Reflections on transfer in machine translation' In: *Contrastes, Revue de l'association pour le developement des etudes contrastives*, January 1984.
- Krauwer, S., & Des Tombe, A.L. (1984b): 'Automatisch vertalen' In: *Van periferie naar kern*, Utrecht, p. 155 - 165.
- Krauwer, S., & Des Tombe, A.L. (1984c): 'Transfer in a multilingual MT system' In: *Proceedings of COLING 84*, Stanford University, Palo Alto, 1984, p. 464 - 467.
- Krauwer, S., & Des Tombe, A.L. (1985): 'Compositionality.' In: *Formal parameters of generative grammar*, G.de Haan & W.Zonneveld (eds), p. 72-80.
- Krauwer, S. (ed) (1988): *NL/B Implementation Report*, ETR, Utrecht/Leuven, December 1988, 139 p.
- Krauwer, S. (ed) (1990): *Final Report Eurotra NL/B Utrecht/Leuven December 1990*, 397 p, with Appendix (394 p) *Final Report submitted to CEC DGXIII/B Luxemburg in fulfilment of Contract ET-CA-NL*, ETR.
- Linden, Erik-Jan van der, & Ton van der Wouden (1990): 'Computer en Lexicon'. In: *Van der Wouden & Van der Linden (1990)*, p. 61-74.
- Neijt, A. and H. Hoekstra (1986): 'Vertalen per computer'. In: *De Gids*, jrg.149, nr.7, 1986.
- Noord, G.J. v., Dorrepaal, J., Arnold, D., Krauwer, S., Sadler, L., Des Tombe, L. (1989): 'An approach to sentence level anaphora in machine translation' In: *Proceedings of the 4th conference of the European chapter of the Association for Computational Linguistics*, University of Manchester, 1989, p. 299-307.
- Noord, G.J. v. (1989a): "BUG: a Directed Bottom-up Generator for Unification Based Formalisms", *Working Papers in Natural Language Processing*, Katholieke Universiteit Leuven, Stichting Taaltechnologie Utrecht, 4, 1989.
- Noord, G.J. v. (1989b): "Bottom-up Generation in Unification-based Formalisms". Paper presented at the *Second European Natural Language Generation Workshop*. Edinburgh 1989.
- Noord, G.J. v. (1989c): "The MiMo2 Research System", paper presented at the *MT colloquium of the University of Stuttgart*. This paper was also presented at the *Eurotra Workshop 1989 in Vimeiro Portugal*.
- Noord, G.J. v. (1989d): "The decidability of Natural Language Grammars", *Working Papers in Natural Language Processing*, Katholieke Universiteit Leuven, Stichting Taaltechnologie Utrecht, 2, 1989
- Noord, van G.J. (1990a): "An Overview of Head-driven Bottom-up Generation", in: Dale, Mellish, Zock (ed), *Current Research in Natural Language Generation*. Academic Press.
- Noord, G.J. v. (1990b): "Reversible Unification Based Machine Translation". *Coling*

- 1990, pp 299-304. Also in *Working Papers in Natural Language Processing*, Katholieke Universiteit Leuven, Stichting Taaltechnologie Utrecht, 7, 1990.
- Noord, G.J. v. & P. van der Eijk (1990): *The MiMo2 Research System*. Paper presented at the Eurotra Annual Workshop, Noordwijkerhout, 1990.
- Noord, G.J. v., S. Balari, P. Schmidt, J. Schuetz (1990a): *Eurotra Papers on Word Order*, ETR, February 1990.
- Noord, G.J. v., J. Dorrepaal, P. v.d Eijk, M. Florenza, L. des Tombe (1990b): *The MiMo2 Research System*. In *Proceedings of The Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Language*, pp. 213-235, Austin, June 1990.
- Noord, G.J. v., J. Dorrepaal, P. van der Eijk, M. Florenza, H. Ruessink, L. des Tombe (forthcoming): *An overview of MiMo2*. To appear Allegranza et al. (forthcoming).
- Petitpierre, D., Krauwer, S., Des Tombe, A.L., Arnold, D., & Varile, N. (1987) : 'A model for preference' In: *proceedings of the ACL*, 1987.
- Raw, A., Frank Van Eynde, Pius ten Hacken, Heleen Hoekstra & Bart Vandecapelle (1989): " An introduction to the Eurotra machine translation system", *Working Papers in Natural Language Processing*, Katholieke Universiteit Leuven, Stichting Taaltechnologie Utrecht, 1, 1989.
- Riemsdijk, H. van & Des Tombe, A.L. (in press): 'Eurogrammar: towards a new standard for grammatical descriptions', in A.Zampolli (ed): *Papers to Honor Bernard Quemada*.
- Ruessink, H.A. (1991): "Two-level formalisms", in Coopmans, P., B. Schouten and W. Zonneveld (eds.) *OTS Yearbook 1990*. pp. 53-62. Also appeared as *Working Papers in Natural Language Processing*, Katholieke Universiteit Leuven, Stichting Taaltechnologie Utrecht, 5, 1989.
- Shieber, Stuart, Gertjan van Noord, Robert C. Moore, Fernando C.N. Pereira (1989): "A Semantic-head-driven Generation Algorithm for Unification Based Formalisms". in: *27th Annual Meeting of the Association for Computational Linguistics*. Vancouver 1989.
- Shieber, Stuart, Gertjan van Noord, Robert C. Moore, Fernando C.N. Pereira (1990): "Semantic-Head-Driven Generation". In: *Computational Linguistics* 16.1, March 1990, p. 30-42.
- Syea, A., B.-J. Dijenborgh, B. Daille, P. Monachesi, H. Grootendorst (1990): *Interlevel Syntax*, ETR, 116 p.
- Tombe, L. des, D. Arnold, L. Jaspaert, R. Johnson. S. Krauwer, M. Rosner, N. Varile, S. Warwick (1985): 'A preliminary linguistic framework for Eurotra, June 1985.' In: *Proceedings of the Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, p. 283-289.
- Tombe, L. des (1986): 'The Eurotra system', In: *Proceedings of the IBM conference on the mechanization of translation*, Copenhagen, 1986
- Tombe, L. des (ed.) (1988): *Dutch Implementation Report 2nd Phase*, ETR,

Utrecht/Leuven, July 1988, 96 p.

Tombe, L. des & Warwick, S (1990): 'The anatomy of ADMIT', paper read at the EURALEX conference, Malaga, Spain, August 1990.

Valentini, E., et al. (1990): The Eurotra Reference Manual 7.0, December 1990, ETR, ca 500 p.

Wouden, Ton van der (1989): De opmars van het lexicon. *Levende Talen* 438 (Maart 1989).

Wouden, Ton van der (1990): Positief Polaire Uitdrukkingen 1. *GLOT* 11.

Wouden, Ton van der, & Erik-Jan van der Linden (eds) (1990): Themanummer Computer en Lexicon, Special Issue of *TABU*, 20, 2.

## **ANNEX I: List of team members**

On the next page you will find an exhaustive list of members of the NL team during the period September 1986 (when the first Contract of Association was signed) until the end of 1990.

For each quarter the main occupation is indicated. *E* refers to Eurotra, *L* to LEXIC, and *G* to GRAMMAR linguistic staff. *S* refers to all sorts of support activities (e.g. management, administration, computer maintenance).

Note that a number of people were employed on a part-time basis.

Not listed here:

- people from Utrecht (and Delft) involved in Eurotra before signature of the contract;
- GRAMMAR staff working in Tilburg;
- replacement staff for S. Krauwer and L. des Tombe;
- members of boards and committees.

<i>year:</i>	86	87				88				89				90			
<i>quarter:</i>	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Abbing							S	S	S	S	S	S					
Bakker den										S	S	S	S	S	S	S	S
Bekker							S	S	S	S	S	S	S				
Bijwaard				E	E		E	E									
Bloksma													E	E	E	L	L
Bolhuis van							E	E	E	E	E	E	E	E	E	E	L
Boschker								E	E	E	E	E					
Broenink		E	E														
Cupcrus											E	E	E	E	E	E	E
Dijenborgh									E	E	E	E	E	E	E	E	E
Dorrepaal			E	E	E	E	E	E	E	E	E	E	E	E	E	E	E
Eijk van der				E	E	E	E	E	E	E	E	E	E	E	E	E	L
Florenza						E	E	E		E	E	E	E	E	E	E	E
Gaalen van														E	E	E	E
Geilen																	L
Groos			E	E	E	E	E	E	E	G	G	G	G	G	G	G	G
Hacken ten			E	E	E	E	E	E	E	E	E	E	E	E	E	E	E
Heylen											E	E	E	E	E	E	E
Herklots														E	E	L	L
Hoekstra	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E
Kamperman					E	E								G	G	G	G
Kester							E	E	E	G	G	G	G	G	G	G	G
Krauwcr	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
Leijten														E	E	E	E
Marsman												E	E	E	E	E	E
Ming																	S
Munster van																	L
Noord van				E	E	E	E	E	E	E	E	E	E	E	E	E	E
Pijnenburg									E	E	E	E	E	E	E	E	E
Pohlmann								E	E	E	E	E	E	E	E	E	E
Pozzo														E	E	E	E
Punsclie		S	S	S	S	S											
Roeland														E	E	E	E
Ross									E	E	E	E					
Ruessink														E	E	E	E
Schöttelndreier									E	E	E						
Schipper												E	E	E	E	E	E
Sesink			E	E	E	E	E	E	E	E	E	E	E	E	E	E	E
Slotman			E	E													
Teeuw								E	E	E	E	E	E	E	E	E	E
Tombe des	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
Vercouteren																S	S
Verheul	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E
Vos de					E	E	E										
Weerden van													S	S	S	S	S
Wouden van der											E	E	E	E	E	E	E
Zeeland van										E	E						

## **ANNEX II: Coverage**

On the next pages you will find an extract from the Final Implementation Report of Eurotra NL/B (December 1990).

The section is titled 'Coverage', and gives a full account of the grammatical and lexical coverage of all analysis, transfer and synthesis components produced by the NL/B group.