

# Vernieuwingsimpuls / Innovational Research

## Grant application form 2003

VIDI scheme

Please refer to Explanatory Notes when completing this form

### Registration form (basic details)

#### 1a. Details of applicant

-Name, title(s): Dr. A. van der Wouden

-Male/female: Male

-Address for correspondence: ATW, Postbus 9515, 2300 RA Leiden

-Preference for English correspondence: no

-Telephone: 071 5277983/030 2536172/071 5171089

-Fax: 071 5272125

-E-mail: vdwouden@let.rug.nl

-Website (optional): <http://www.let.rug.nl/~vdwouden>

-Doctorate (date): 23 06 1994

-Use of extension clause (see Notes): yes

after the birth of my daughter (July 17<sup>th</sup> 1995) I have taken up maternity leave ("ouderschapsverlof") for three months net

#### 1b. Title of research proposal

Dutch as a construction language

#### 1c. Summary of research proposal

(max. 300 words, plus max. 5 KEYWORDS)

The research focuses on the description and implementation of extended lexical units (ELUs), i.e. all linguistic building blocks, compositional or not, larger than words, in spoken Dutch. It is innovative in at least four ways, firstly as it concentrates on spoken rather than written language, secondly as it is corpus-based, thirdly as it concentrates on ELUs, the area between grammar and lexicon, and fourthly as it takes as its grammatical framework Construction Grammar, one of the few theories taking seriously the fact that a considerable part of every day language consists of ELUs.

This grammatical description is important by itself, as it furthers our knowledge both of the Dutch language and language in general. But this kind of information is also crucial for didactic purposes in the realm of Dutch as a second or other language, on the one hand, and for Natural Language Processing (NLP) of Dutch, on the other, where ELUs are an important stumbling block for successful applications. Moreover, the implementation of ELUs is a challenge for any computational model of language, especially if it needs to be modular, flexible, extensible and robust.

Grammatical description of Dutch (and other languages) traditionally concentrates on written language; spoken language has its own properties that seldom receive the attention they deserve. Traditional grammars describe the phenomena traditionally described by grammars, but corpus based grammars allow one to pay attention to all and only those facts of language actually occurring.

Corpus based grammatical description is only possible if there is a serious, rich corpus which is as representative as possible for the language described. The project will seek to formally model the spontaneous parts of the Spoken Dutch Corpus (Corpus Gesproken Nederlands), to be completed by the end of 2003.

#### Key words

Constructions, spoken Dutch, computational linguistics, corpus linguistics, construction grammar

#### 1d. NWO Council area

GW Geesteswetenschappen

#### 1e. Host institution (if applicable)

ULCL, Leiden University

# Vernieuwingsimpuls / Innovational Research

## Grant application form 2003

VIDI scheme

Please refer to Explanatory Notes when completing this form

---

### Research proposal

#### 2. Description of the proposed research

(max. 4000 words on max. 8 pages, excl. literature references. A description of sub-projects is required. Include details of:)

#### Background

##### a. Spoken language

Spoken language usually precedes language represented in writing. Children know how to speak and listen years before they learn to read and write. The history of language is estimated to be in the order of magnitude of hundreds of thousands of years, the history of writing in thousands of years. There are many language communities without writing, but only in the case of dead languages such as Latin the number of readers will perhaps outnumber the number of speakers.

The fact that writing originates as a representation of spoken language does not imply that the only difference between written and spoken language is the modality (Bolinger & Sears 1981). For example, intonation in spoken language is replaced, to a certain extent, by punctuation in written language, but there is no direct mapping between the two. Miller & Weinert (1998) even go so far as to claim that spoken and written languages are (partially) different systems in almost every aspect, that is to say, of morphology, syntax, vocabulary, and the organization of texts.

The time pressure inherent to spontaneous speech situations leaves less time for complex mental computations than the scrupulous composition of written text. One of the strategies for speakers to overcome this problem is to employ ELUs (for Extended Language Units), which we will use as a pretheoretical term to refer to all linguistic building blocks larger than words, be they compositional or not, to construct their utterances rather than being completely original. Kuiper (1996) observes that in certain high-pressure situations, most of a speaker's utterances consist of stored (or, to use another metaphor, pre-compiled) linguistic material, with very little syntactic computation going on. We may assume that this is one extreme of a continuum, the other extreme being Chomsky's idealization of a creative language user with an infallible memory and infinite processing power, with enough time to verbalize new ideas in an original way. As a starting hypothesis (to be investigated in the project) we assume that the various subtypes of spontaneous speech in the Corpus will be on different positions on the scale between those extremes.

The primacy of spoken language notwithstanding, the grammatical description of Dutch (as opposed to recent developments in the English speaking community, cf. e.g. Biber et al. (1999)) usually takes written language as a starting point, paying attention to spoken language phenomena only occasionally, if at all (Haeseryn et al. 1997, De Vries 2001). Comparatively little research has been carried out into the properties of spoken Dutch, although there is a certain tradition of research into properties of spoken variants of Dutch starting with the studies by the Groningen high school teacher Wobbe de Vries around 1910 (De Vries 1910, 1911, 1914) through and the dissertations of Bertha Uijlings (1956) and Frank Jansen (1981) up to the recent book by Jelle de Vries (2001), plus a considerable amount of smaller papers. However, these studies dealing with oral Dutch are either in-depth studies of one or two spoken language constructions, or anecdotal overviews of a great number of peculiar phenomena, with shallow analyses at best. Since Jansen (1981), this will be the first corpus-based project investigating syntactic properties of spoken Dutch at a larger scale. As the Spoken Dutch Corpus (CGN), however, is much larger and much diverse than Jansen's one, it allows for comparison of variants of spoken Dutch from various regions and situations for the first time.

##### b. CGN

Although Dutch is among the world's best-studied languages, relatively little is known about spoken Dutch, as most linguistic studies deal with written variants of the language. The Spoken Dutch Corpus ("Corpus Gesproken Nederlands" or CGN) is meant to change this. It is a Dutch-Flemish project (1998-2003) aiming at the collection,

# Vernieuwingsimpuls / Innovational Research

## Grant application form 2003

## VIDI scheme

Please refer to Explanatory Notes when completing this form

---

description and annotation of ten million words of spoken Dutch, two thirds from the Netherlands, one third from the Dutch speaking part of Belgium (Oostdijk 2000a, 2000b; Oostdijk et al. 2002). The data are recorded in a variety of situations: from informal coffee table conversations to parliamentary speeches, and from literature read aloud to telephone calls; data are recorded from speakers of both sexes, from various regions, and with various sociolinguistic backgrounds.

The speech data are enriched with various levels of linguistic annotation, including orthographic transcription, Part-of-Speech tagging and lemmatization (the assignment of base forms and morphosyntactic tags (Van Eynde et al. 2000), and syntactic analysis (Hoekstra et al. 2001, Van der Wouden et al. 2002a, 2002b). The corpus allows for new types of investigation into the intricacies of spoken Dutch by researchers from a broad range of backgrounds (sociolinguistics, psycholinguistics, theoretical linguistics, phonetics, cognitive science, forensic linguistics, speech technology, etc.).

### c. Extended lexical units (ELUs), or the organization of the language faculty

According to traditional generative thinking (Chomsky 1981), the form of well-formed utterances in natural languages is determined by the interaction of two subsystems of the language system, viz. the grammar and the lexicon. The grammar is the computational system, the set of general rules governing the structures of words, clauses, utterances, texts and whatever other levels of linguistic structure one wishes to distinguish. The lexicon, on the other hand, is the mental dictionary, the repository containing the unpredictable properties of words, or lexical items in a broader sense, i.e., not just words, but also morphemes, idioms, stock phrases, formulae, etcetera (di Sciullo & Williams 1987).

The division of labor between grammar and lexicon is an open question. Among other things, it is the topic of what one might want to call an ideological discussion. Radically lexicalist theories such as Combinational Categorical Grammar (Steedman 2000), Type Logical Grammar (Moortgat 1997, Bernardi 2001) and Head Driven Phrase Structure Grammar (HPSG, Pollard & Sag (1994)) have, on theoretical grounds, other answers to this question than theories in which the autonomy of syntax is one of the corner stones (Chomsky (1981), Barbiers (1995), etc.).

The question of the division of labor, however, may also be approached as an empirical one (cf. Jackendoff (1997)). It then concerns the division of labor between grammar and lexicon in actual language use. What part of real life utterances is computed, what part is taken from the lexicon in one way or another?

In the literature, estimates concerning the quantitative importance of ELUs differ greatly: Altenberg (1998) reports that spoken English is very rich in “recurrent strings”, by which he means “any continuous string of words recurring more than once in identical form” (p. 101): his corpus of nearly half a million running words contains over 201000 recurrent word-combinations. “A rough estimation indicates that over 80 per cent of the words in the corpus form part of a recurrent word-combination in one way or another.” (p. 102). Sprenger (2003) reports figures for Dutch that are an order of magnitude lower, but 1. she looked at written rather than spoken language 2. she only looked at content words (and not at function words) and 3. she only counted fixed expressions that were explicitly marked that way in dictionaries. Finally, Jackendoff (1997: 157) quotes Weinreich (1969) citing estimates of over 25,000 fixed expressions stored in the lexicon of an average speaker of English. Anyway, the conclusion in Jackendoff (1997: 157) that “[t]here are too many idioms and other fixed expressions for us to simply disregard them as phenomena ‘on the margin of language’” seems entirely warranted, and is a major motivation for this project.

Perhaps the exact amount of ELUs in spoken language is less interesting than a more qualitative approach: it is an open question in what kind of configurations we may find ELU-like phenomena. Elsewhere I have argued that collocation may be found everywhere, both below the word level (*spin* ‘spider’ functions as an intensifier in the compound *spinnijdig* ‘hopping mad, livid’ only (Van der Wouden 1997)) and between ELUs (the combination of particles *maar eens* has a strong preference for weak directive constructions, such as imperatives *laat maar eens wat zien* ‘please show us something’ and suggestions *ik zou maar eens naar bed gaan* ‘you’d better go to bed now’ (Van der Wouden 2001)); Booij (2002) argues convincingly that constructions function in inflectional paradigms: certain Dutch verb tenses are expressed by morphologically complex words (e.g. imperfect tense *liep* ‘walked’ and *speelde* ‘played’), other verb tenses are expressed by means of constructions, i.e. ELUs (e.g. perfect tense *heeft gelopen* ‘has walked’ and *heeft gespeeld* ‘has played’). The question then becomes: is there any kind of construction or configuration where we do NOT find ELU-like behavior?

# Vernieuwingsimpuls / Innovational Research

## Grant application form 2003

## VIDI scheme

Please refer to Explanatory Notes when completing this form

---

In order to try to begin to find an answer to this type of question, the project will have to look systematically at the various kinds of ELUs that have been proposed in the literature. Below is an incomplete overview (cf. also Moon (1998)).

### Collocations

Recently, various researchers have suggested that the role of the lexicon may be more important and interesting than traditional generative thinking along the lines of Chomsky (1991) and subsequent literature suggests. For instance, collocation, i.e. idiosyncratic distributional behavior of lexical items, abounds in language. Traditional text book examples of collocations involve, among other things (Hausmann 1989-1991), prepositions with verbs and adjectives with nouns. E.g., it is unpredictable beforehand that the Dutch verb *luisteren* 'listen' takes a complement with the preposition *naar* 'to, towards'. Another example: the English adjectives *strong* and *powerful* are near-synonyms (or plesionyms Hirst 1995). Still, the words show different distributions, e.g., one says *strong tea* and not *powerful tea*, but if the noun modified is *computer*, the other adjective is preferred: *powerful computer* rather than *strong computer*. As we don't know of any principle of grammar from which the preference of the noun *tea* for the adjective *strong* may follow, we have to assume that this information is stored in the lexicon, and the same of course holds for the collocational bond between *computer* and *powerful*.

Van der Wouden (1997) argues at length that the distribution of negative polarity items, which is usually seen as a grammatical phenomenon, be it syntactic (e.g. Progovac (1994)), semantic (e.g. Ladusaw (1979)) or pragmatic (e.g. Linebarger (1980)) in nature, has a collocational component as well. For example Dutch *ook maar* 'at all'<sup>1</sup> needs a stronger negative context (Zwarts 1981, 1986; Van der Wouden 1997) than *hoeven* 'need':

- (1) a. *Niemand heeft ook maar iets gezien*  
Nobody has at-all anything seen  
'Nobody saw anything at all'
- b. *Niemand hoeft water te drinken*  
Nobody need water to drink  
'Nobody need drink water'
- c. *\*Weinigen hebben ook maar iets gezien*  
Few have at-all anything seen
- d. *Weinigen hoeven water te drinken*  
Few need water to drink  
'Few (people) need drink water'

There are, however, certain contexts that allow *ook maar* while disallowing *hoeven*, e.g. conditional clauses:

- (2) a. *Als je ook maar iets gezien hebt moet je bellen*  
If you at-all anything seen have must you call  
'Please call if you saw anything'
- b. *\*Als je water hoeft te drinken moet je bellen*  
If you water need to drink must you call

As this type of distributional effect does not, as far as we know, follow from any principle of grammar (and cannot even be accounted for in standard versions of the grammar, as Sailer & Richter (2002) show for HPSG), they have to be stipulated somewhere in the lexicon.

### Idioms

According to the Cambridge International Dictionary of English, an idiom is "a group of words in a fixed order having a particular meaning, different from the meanings of each word understood on its own", in other words, idioms are non-compositional ELUs. More often than not, the syntactic flexibility of idioms is restricted.

Consider the idiom *het loodje leggen* 'to come off badly', literally 'to lay the piece of lead'. The meaning of the whole has nothing to do with either laying or pieces of metal. But it doesn't work to attribute the

---

<sup>1</sup> We note in passing that the meaning and other properties of the ELU *ook maar* do not follow from its constituting parts.

# Vernieuwingsimpuls / Innovational Research

## Grant application form 2003

## VIDI scheme

Please refer to Explanatory Notes when completing this form

---

idiomatic meaning to the literal string *het loodje leggen*, as the verb inflects just as any other, and the idiom behaves syntactically just like any other combination of verb plus direct object (3a-3b). If we try to put the sentence in passive form, however, only the literal meaning is available (3c).

- (3)
- a. *Hij legt het loodje*  
He lays the piece of lead  
'He comes off badly'
  - b. *Hij heeft het loodje gelegd*  
He has the piece of lead laid  
'He has come off badly'
  - c. *Het loodje is door hem gelegd*  
The piece of lead is by him laid  
'The piece of lead has been put down by him'

Idioms, especially those that show syntactic flexibility, are problematic for all linguistic theories where the meaning of language constructs is the result of somehow combining the meanings of the constituting parts (Chafe 1968, Van der Linden 1993, Schenk 1994).

### Constructions

Fillmore et al. (1998) argue in favor of "the view that the realm of idiomaticity in a language includes a great deal that is productive, highly structured, and worthy of serious grammatical investigation." They further suggest "that an explanatory model of grammar will include principles whereby a language can associate semantic and pragmatic interpretation principles with syntactic configurations larger and more complex than those definable by means of single phrase structure rules." This paper is usually taken to be the beginning of a revival of the notion *construction* in linguistics (Goldberg 1995: 1). The central notion in this type of approaches is *construction* or *constructional idiom* (Langacker (1987), Booij (2002), etc.), "syntactic constructions with a (partially or fully) non-compositional meaning contributed by the construction, in which - unlike idioms in the traditional sense - only a subset (possibly empty) of the terminal elements is fixed" (Booij). Traditional examples of such constructional idioms are the *een schat van een kind* construction (Paardekooper 1956, Everaert 1992, Booij 2002) and the *way* construction (Israel 1996, Verhagen 2002):

- (4)
- a. *Een schat van een kind*  
A sweetheart of a child  
'a sweet child'
  - b. *Een kast van een huis*  
A cupboard of a house  
'a big house'
  - c. *De priesters wurmen zich een weg door de gelovigen*  
The priests squeeze REFL a way through the faithful  
'the priests squeeze their way through the faithful'

(4a) and (4b) look like ordinary noun phrases with preposition phrase modifiers, but their semantics is completely unexpected. This therefore has to be stored in the lexicon, together with information such as the fact that the righthand noun phrase (*kind*, resp. *huis*) is head of the construction. Comparable reasoning holds for the *way* construction: the sentence in (4c) has a resultative meaning that does not follow from the meanings of the verb *wurmen* or the noun *weg*.

### Formulae

Another type of ELU one finds in the literature is *formula*, "a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar" (Wray 2002: 9).

A considerable proportion of our everyday language is "formulaic" (Wray): it is predictable in form, idiomatic, and seems to be stored lexically in fixed, or semi-fixed, chunks. Native speakers of Dutch know what to say when they start dinner, viz. *smakelijk eten* (or a variant, depending on geographical and social factors (Pauw

# Vernieuwingsimpuls / Innovational Research

## Grant application form 2003

VIDI scheme

Please refer to Explanatory Notes when completing this form

---

van Wieldrecht 1985, Van Ditzhuyzen & Groskamp-ten Have 2001)), and they can choose from a range of possible things to say when they want to finish a phone call, e.g. *ik ga maar eens ophangen* I go PART PART up-hang 'I think I'll hang up' or something. These speakers do not have to construct these utterances, they retrieve them directly.

Some conversations appear to consist of such formulae, cliché's and stock phrases only (Van Eijk 1987); the only purpose of the exchange seems to be the constitution and/or confirmation of the social relationship between the conversation partners. Of course, this is an extreme case, but hardly any conversation can do without formulae. It is still an open question what the amount of formulaic language is within spontaneous spoken language.

### Fixed phrases/Vaste verbindingen

Lexicographers have their own ideas on the ELUs in language that deserve treatment in the dictionary (Verstraten 1992). Psycholinguistic research shows that many of these fixed phrases in the lexicographic sense ("vaste verbindingen" is the traditional Dutch term) must be stored as units in the mental lexicon (Cutting & Bock 1997, Sprenger 2003).

## 2a. Research topic: The modelling of spoken language, or Dutch as a construction language

Notwithstanding their important role in (spoken) language, ELUs are difficult to tackle in most current grammar formalisms (Van der Wouden 1997). Efforts to describe such phenomena in a serious way in existing frameworks usually lead to extensions of the theory (cf. e.g. Van Gestel (1995) for Chomsky (1981) style generative grammar and Sailer & Richter for HPSG (Pollard & Sag 1994)).

The grammatical framework currently taking ELUs most seriously is Construction Grammar (Fillmore et al. (1988) and subsequent literature). Construction Grammar is a generative grammar framework in the broad sense (Kay 1998), in that "it tries to account for the infinite number of expressions that are allowed by grammar while attempting to account for the fact that an infinite number of other expressions are ruled out or disallowed" (Goldberg 1995: 7); it belongs to the same family of formal, monostratal, radically lexical grammar types as HPSG (Kay 1998). Under a strong interpretation of the theory, "[t]he mechanism for shaping sentences is an interconnected repertory of grammatical constructions; a few of them are quite general and abstract, but most of them identify specific lexical constituents, in most cases just one (these are the ordinary lexical items) but in many cases more than one" (Fillmore 1997: ix).

The ultimate proof of the usefulness of a linguistic theory is a working implementation. One of the goals of the project is to test the grammatical descriptions in a working system, to wit, the Leiden Delilah parser (Cremers 2002). As this system is about the only working parsing system for Dutch that has a serious semantic component, it appears to be especially suited to function as a test bed for a formalism such as Construction Grammar where syntax and semantics go hand in hand.

However, as far as we know, there is no functioning implementation of Construction Grammar yet. The problems it takes to enrich HPSG-like frameworks with a mechanism to handle collocations (Sailer & Richter 2002) suggest that it will be non-trivial to extend any lexically driven parsing system, including Delilah, with a module to deal with ELUs. The implementation of (partially) non-compositional ELUs with a certain amount of morphosyntactic flexibility is a major challenge for at least two reasons:

- on the one hand, it offers insight into the grammar needed to drive a structured dynamic lexicon, and into the amount of computation needed in the lexicon;
- on the other hand, it tests the architecture (especially the modularity and the flexibility) and the data structures of semantically relevant lexica for advanced applications.

Both aspects are crucial for the Delilah system, one of the very few NLP applications with a semantic focus. To put it strongly: the formal characterization of ELUs is one of the core problems of semantically driven parsing and generation. If real life language abounds with fixed expressions with a semantics that is not completely compositional, it is more than necessary for a semantic parser to deal with them in a way that is both efficient and robust. The Delilah lexicon in its current state scratches only the surface of this problem; the proposed project will give the Delilah system a major boost. The systematic implementation of non-

# Vernieuwingsimpuls / Innovational Research

## Grant application form 2003

VIDI scheme

Please refer to Explanatory Notes when completing this form

---

compositional constructions built out of lexical items that also occur in compositional constructs as well may ask for a complete rethinking of the parser's architecture.<sup>2</sup>

### 2b. Approach

The project will try to find an answer to the following questions:

- what types of extended lexical units (ELUs) occur in spontaneous spoken Dutch?
- is it possible to describe them in (a version of) Construction Grammar?
- what part of the utterances in various types of spontaneous spoken Dutch is constructional in nature? is there any relation between the type of speech and the type(s) of constructions?
- are there any restrictions on possible ELUs; can they be understood from the theory?
- is it possible to implement all naturally occurring types of ELU in a lexically driven parser such as Delilah?

### 2c. Innovation

The research proposed is innovative in at least four ways, as it concentrates on spoken rather than written language, as it is corpus-based, as it concentrates on ELUs, and fourthly as it takes as its grammatical framework Construction Grammar, one of the few theories taking seriously the fact that a considerable part of every day language consists of ELUs.

1. Since Jansen (1981), this will be the first project investigating syntactic properties of spoken Dutch at a larger scale. As the Corpus used is much larger and much diverse than Jansen's one, it allows for comparison of variants of spoken Dutch from various regions and situations for the first time.

2. In the Netherlands (and Belgium), there is hardly a tradition of corpus-based research into the grammatical properties of the spoken vernacular (*pace* De Jong (1979) and Jansen (1981)). The new Spoken Dutch Corpus (CGN) offers new and exciting possibilities for this kind of research.

3. Although quite some research into Dutch idioms, fixed phrases, etc. has been carried out from psycholinguistic, theoretical and lexicological perspectives, the proposal is new in taking into consideration all lexical units larger than words.

4. The research proposed uses uses grammatical framework of Construction Grammar (CG), one of the very few theories taking seriously the fact that a considerable part of every day language consists of ELUs, rather than assuming that all speech consists of new sentences generated 'from scratch', out of words and grammar rules. CG is a relatively new theory; it will be the first time that it will be applied to Dutch on such a large scale. The project also aims at implementing CG, which is completely new, as far as we can tell.<sup>3</sup>

### 2d. Plan of work

The research lends itself for a natural division into two parts: one part mainly combines exploration (of the corpus) and description (of the constructions found), whereas the other part deals more with modelling and implementation.

#### A. Researcher 1 (postdoc - the applicant)

The approach of the first researcher will mainly be bottom-up. This part of the project combines exploration and description. Starting with standard computational linguistics techniques (Manning & Schütze 1999, Van der Wouden 2001), all recurrent strings in the sense of will be extracted from the spontaneous parts of the Spoken Dutch Corpus. As far as it is feasible within the project, they will all be inspected manually (which does not

---

<sup>2</sup> Cf. e.g. Sailer & Richter (2002) and the Call for Papers for the ACL 2003 Workshop on Multiword Expressions (<http://www.cl.cam.ac.uk/users/alk23/mwe/mwe.html>).

<sup>3</sup> "I am attempting to find other computational work which is motivated by this theory, but so far have not had much success." (Karen Verspoor, message to CXG (construction grammar discussion) list, February 6<sup>th</sup>, 2003).

# Vernieuwingsimpuls / Innovational Research

## Grant application form 2003

VIDI scheme

Please refer to Explanatory Notes when completing this form

---

exclude the possibility of automatic or statistic filtering). It is to be expected that a considerable part of the combinations can be explained by statistics - there is a fair chance that frequent items occur in each other's environment - or by well-known very general principles of grammar - it is e.g. not unlikely that a noun follows a determiner. Still, statistics and grammar alone cannot explain all of the reported 80 % of words occurring in such recurrent strings (Altenberg 1998). So what is left over after filtering are potential ELUs. On the basis of earlier research, it is to be expected that many new types of constructions will be found. E.g. function words such as Dutch modal particles turned out to show collocational behavior as well (Van der Wouden 2001); in certain cases, the combination of two (or more) such particles may develop a specialized semantics (Hoeksema 2002) and turn into a word (e.g. the untranslatable Dutch *weleens*, which has been analyzed as an imperfective marker (Zwarts et al. 2002), which derives from the particle combination *wel eens*). Moreover, lexicalized particle combinations may develop collocational behavior themselves, as can be seen in the preference of the particle combination *maar eens* for weak directives (Van der Wouden 2002a, 2002b)).

The recurrent strings found to be constructions will be classified and grammatically described. The end product will be a monograph with the provisional title *Constructions in Dutch, a corpus based approach*.

Another task of researcher 1 is to oversee (together with dr. Cremers and prof. Verhagen) the work of Researcher 2.

### B. Researcher 2 (graduate student)

The approach of researcher 2 is mainly top-down. It involves (theoretical) modelling on the one hand and (practical) implementation on the other. After study of the relevant literature (augmented by input from researcher 1) he (or she, of course) will develop a typology of ELUs. He will propose a description of each of the types within the Construction Grammar framework (which may need extensions for certain ELUs the Construction Grammar people never thought of). He will propose an extension of the Leiden Delilah parser (Cremers 2002) to deal with Construction Grammar descriptions. He will put the validity of the constructions' descriptions to test by implementing them in the parser.

End goal of the student's research is a dissertation on the representation of Dutch constructions, supervised, in principle, by prof. dr. A. Verhagen, dr. C.L.J.M. Cremers and dr. A. van der Wouden.

### Work Schedule

**year 1** The PhD student is trained in the Construction Grammar framework and in the technical details of the Delilah parser. Literature study. Statistical extraction of recurrent combinations from the corpus. Participation in the 3rd International Conference on Construction Grammar (ICCG 3), Marseille, France, 01-04 July 2004. Preparation of at least one publication.

**year 2** Extension of the Delilah parser to handle constructions and other fixed expressions. Selection of relevant recurrent combinations; development of descriptive framework. Presentation of first results at an international conference. A longer visit is planned to one of the centers of Construction Grammar, e.g., the University of California at Berkeley. We choose for this moment, as this visit may further the student's training given in the first year.

**year 3** Grammatical description of a number of constructions. Implementation of a number of example construction in Delilah, testing, redesign. Presentation of first results at an international conference.

**year 4** Continuation of the activities in year 3. The PhD student writes his or her thesis.

**year 5** The applicant writes his monograph.

The research will be carried out in the Leiden Centre for Linguistics (ULCL). Among the ULCL staff members directly relevant for this research proposal are dr. C.L.J.M. Cremers (computational linguistics), prof.dr. V.J. van Heuven (empirical linguistics) and prof.dr. A. Verhagen (Dutch linguistics).

Experience has learnt that the automatic extraction of possible ELUs from a corpus of considerable size (such as the CGN) is unfeasible on a standard state of the art Personal Computer (Pentium 4, 1,7 GHZ, 256 MB). Therefore, the project needs either a real powerful PC, with a lot of memory, or facilities on a large computer somewhere else, e.g. in the faculty of sciences.

### 2e. Literature references

- Altenberg, Bengt. 1998. On the phraseology of spoken English: The evidence of recurrent word-combinations. In *Phraseology. Theory, Analysis, and Applications*, ed. by A.P. Cowie, 101-22. Oxford: Clarendon Press.
- Barbiers, Sjef. 1995. *The Syntax of Interpretation*. University of Leiden dissertation.
- Bernardi, Raffaella. 2002. *Reasoning with polarity in categorial type logic*. Utrecht dissertation.
- Biber, Douglas, and others. 1999. *Longman grammar of spoken and written English*. Harlow: Longman.
- Bolinger, Dwight, Donald A. Sears. 1981. *Aspects of Language*. New York: Harcourt Brace Jovanovitch, 3rd edition. (1st edition 1968).
- Booij, Geert. 2002. Constructional idioms, morphology, and the Dutch lexicon. To appear in the *Journal of Germanic Linguistics* 13, 4.
- Cacciari, C., P. Tabossi. 1988. The comprehension of idioms. *Journal of Memory and Language* 27, 668-683.
- Chafe, Wallace L. 1968. Idiomaticity as an anomaly in the Chomskyan paradigm. *Foundations of Language* 4, 109-127.
- Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.
- Cremers, Crit. 2002. (‘n) Betekenis berekend. *Nederlandse Taalkunde* 7, 375-395.
- Cutting, J.C., K. Bock. 1997. That’s the way the cookie bounces. syntactic and semantic components of experimentally elicited idiom blends. *Memory and Cognition* 25, 57-71.
- van Ditzhuyzen, Reinildis, Amy Groskamp-ten Have. 2001. *Hoe hoort het eigenlijk?*. Haarlem: Becht, 28th edition.
- van Eijk, Inez. 1987. *Dooddoeners en stoplappen*. Utrecht [etc.]: Het Spectrum.
- Everaert, Martin. 1992. Nogmaals: een schat van een kind. In *De binnenbouw van het Nederlands. Een bundel artikelen voor Piet Paardekooper*, ed. by Hans Bennis and Jan W. de Vries, 45-54. Dordrecht: ICG Publications.
- Van Eynde, Frank, Jakub Zavrel, Walter Daelemans. 2000. Lemmatisation and morphosyntactic annotation for the Spoken Dutch Corpus. In *Computational Linguistics in the Netherlands 1999. Selected Papers from the Tenth CLIN Meeting*, ed. by Paola Monachesi, 53-62. Utrecht: Utrecht University, Utrecht Institute of Linguistics OTS.
- Fillmore, Charles J. 1997. Foreword. In *Words and the grammar of context*, ed. by Paul Kay, ix-xviii. Stanford, CA: CSLI.
- Fillmore, Charles J., Paul Kay, Mary Catherine O’Connor. 1988. Regularity and idiomaticity in grammatical constructions: The case of *let alone*. *Language* 64, 501-38.
- van Gestel, Frank. 1995. En Bloc Insertion. In *Idioms: Structural and psychological perspectives*, ed. by Martin Everaert, Erik-Jan van der Linden, André Schenk, and Rob Schreuder, 75-96. Hillsdale, NJ and Hove, UK: Lawrence Erlbaum Associates.
- Goldberg, Adele E. 1995. *Constructions. A Construction Grammar Approach to Argument structure*. Chicago: University of Chicago Press.
- Haeseryn, Walter, and others (eds.). 1997. *Algemene Nederlandse Spraakkunst*. Groningen and Deurne: Martinus Nijhoff and Wolters Plantijn. 2e, geheel herz. dr.
- Hausmann, Franz Josef. 1989-1991. Collocations. In *Wörterbücher: ein internationales Handbuch zur Lexikographie / Dictionaries: an international encyclopedia of lexicography / Dictionnaires: encyclopedie internationale de lexicographie*, ed. by Franz Josef Hausmann et al., I: 1010-19. Berlin [etc.]: De Gruyter. (Handbücher zur Sprach- und Kommunikationswissenschaft; Bd. 5).
- Hirst, Graeme. 1995. Near-synonymy and the structure of lexical knowledge. In *Representation and Acquisition of Lexical Knowledge: Polysemy, Ambiguity, and Generativity*, ed. by Judith Klavans, 51-56. Stanford: AAI Spring Symposium Series.
- Hoeksema, Jack. 2002. Polarity-sensitive scalar particles in early modern and present-day Dutch: distributional differences and diachronic developments. to appear in a special issue of the *Belgian Journal of Linguistics* on particles, edited by Ton van der Wouden, Ad Foolen and Piet Van de Craen.
- Hoekstra, Heleen, Michael Moortgat, Ineke Schuurman, Ton van der Wouden. 2001. Syntactic Annotation for the Spoken Dutch Corpus Project (CGN). In *Computational Linguistics in the Netherlands 2000*, ed. by W. Daelemans, K. Sima’an, J. Veenstra, and J. Zavrel, 73-87. Amsterdam/New York: Rodopi.
- Israel, Michael. 1996. The way constructions grow. In *Conceptual Structure, Discourse and Language*, ed. by

# Vernieuwingsimpuls / Innovational Research

## Grant application form 2003

## VIDI scheme

Please refer to Explanatory Notes when completing this form

---

- Adele Goldberg, 217-230. Stanford: CSLI.
- Jackendoff, Ray. 1997. *The Architecture of the Language Faculty*. Cambridge, Mass.: The MIT Press.
- Jansen, Frank. 1981. *Syntaktische konstrukties in gesproken taal*. Leiden dissertation.
- de Jong, Eveline D. (ed.) 1979. *Spreektaal*. Woordfrequenties in gesproken Nederlands, Utrecht: Bohn, Scheltema & Holkema.
- Kay, Paul. 1998. An informal sketch of a formal architecture for construction grammar. Ms. UC Berkeley.
- Kuiper, Koenraad. 1996. *Smooth talkers: the linguistic performance of auctioneers and sportscasters*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Ladusaw, William A. 1979. *Polarity sensitivity as inherent scope relations*. University of Texas at Austin dissertation. Distributed by IULC, Bloomington, Indiana, 1980.
- Langacker, Ronald W. 1987. *Foundations of cognitive grammar*. Stanford: Stanford University Press. Volume 1: Theoretical prerequisites.
- van der Linden, Erik-Jan. 1993. *A categorial, computational theory of idioms*. ITK, Tilburg dissertation.
- Linebarger, Marcia. 1980. *The grammar of negative polarity*. MIT dissertation. Distributed by IULC, Bloomington, Indiana, 1981.
- Manning, Christopher D., Hinrich Schütze. 1999. *Foundations of statistical natural language processing*. Cambridge, Mass.: The MIT Press.
- Miller, Jim, Regina Weinert. 1998. *Spontaneous spoken speech. Syntax and Discourse*. Oxford: Clarendon.
- Moon, R. 1998. *Fixed Expressions and Idioms in English: A Corpus-based Approach*. Oxford: Clarendon Press.
- Moortgat, Michael. 1997. Categorical type logics. In *Handbook of logic and language*, ed. by Johan van Benthem and Alice ter Meulen, 93-177. Amsterdam [etc.] and Cambridge, Mass.: Elsevier and MIT Press.
- Oostdijk, Nelleke. 2000a. Building a corpus of spoken Dutch. In *Computational Linguistics in the Netherlands 1999. Selected Papers from the Tenth CLIN Meeting*, ed. by Paola Monachesi, 147-157. Utrecht: Utrecht University, Utrecht Institute of Linguistics OTS.
- Oostdijk, Nelleke. 2000b. The Spoken Dutch Corpus. Overview and first evaluation. In *Proceedings of the second International Conference on Language Resources and Evaluation*, ed. by M. Gavralidou, G. Carayannis, S. Markantonatou, S. Piperidis, and G. Stainhaouer, 887-893. Paris: ELRA.
- Oostdijk, Nelleke, Wim Goedertier, Frank Van Eynde, Louis Boves, Jean-Pierre Martens, Michael Moortgat, Harald Baayen, 2002. Experiences from the Spoken Dutch Corpus Project. *Proceedings LREC 2002*.
- Overdiep, G.S. 1937. *Stilistische grammatica van het moderne Nederlandsch*. Zwolle: Tjeenk Willink.
- Paardekooper, P.C. 1956. Een schat van een kind. *De Nieuwe Taalgids* 49, 93-9.
- Pauw van Wieldrecht, Agnies. 1985. *Het dialect van de adel*. Utrecht: HES Uitgevers.
- Pollard, Carl, Ivan Sag. 1994. *Head-Driven Phrase Structure Grammar*. Stanford and Chicago: CSLI and University of Chicago Press.
- Progovac, Ljiljana. 1994. *Negative and positive polarity. A binding approach*. Cambridge: Cambridge University Press.
- Sailer, Manfred, Frank Richter. 2002. Not for love or money: Collocations! In *Proceedings of Formal Grammar 2002*, ed. by Gerhard Jäger, Paola Monachesi, Gerald Penn, and Shuly Wintner, 149-160.
- Schenk, André. 1994. *Idioms and collocations in compositional grammars*. Utrecht dissertation.
- di Sciullo, Anna-Maria, Edwin S. Williams. 1987. *On the definition of word*. Cambridge, Mass.: MIT Press.
- Sprenger, Simone A. 2003. Fixed expressions and the production of idioms. Dissertation Nijmegen, in progress.
- Steedman, Mark. 2000. *The syntactic process*. Cambridge, Mass.[etc.]: The MIT Press.
- Uijlings, B.J. 1956. *Praat op heterdaad*. Assen: Van Gorcum [etc.]. ook als diss. Utrecht onder de titel *Syntactische verschijnselen bij onvoorbereid spreken*.
- de Vries, Jelle. 2001. *Onze Nederlandse Spreektaal*. Den Haag: SDU Uitgevers.
- de Vries, Wobbe, 1910. Dysmelie. Opmerkingen over syntaxis. Verhandeling behorende bij het programma van het gymnasium der gemeente Groningen voor het jaar 1910-1911.
- de Vries, Wobbe, 1911. Dysmelie. Opmerkingen over syntaxis (vervolg). Verhandeling behorende bij het programma van het gymnasium der gemeente Groningen voor het jaar 1911-1912.
- de Vries, Wobbe, 1914. De typen der mededeeling. Opmerkingen over syntaxis. In: Verhandeling behorende bij het programma van het gymnasium der gemeente Groningen voor het jaar 1914-1915.
- Verhagen, Arie. 2002. From parts to wholes and back again. *Cognitive Linguistics* 13, 403-39.
- Verstraten, Linda. 1992. *Vaste verbindingen. Een lexiconologische studie vanuit cognitief-semantisch perspectief naar fraseologismen in het Nederlands*. Leiden dissertation.

# Vernieuwingsimpuls / Innovational Research

## Grant application form 2003

VIDI scheme

Please refer to Explanatory Notes when completing this form

---

- Weinert, Regina. 1995. The role of formulaic language in second language acquisition: A review. *Applied Linguistics* 16, 180-205.
- Weinreich, Uriel. 1969. Problems in the analysis of idioms. In *Substance and structure of language*, ed. by Joan Puhvel, 23-81. Berkeley and Los Angeles: University of California Press (reprinted in *On Semantics*, 1980). Van der Wouden, Ton. 1997. *Negative Contexts. Collocation, polarity, and multiple negation*. London and New York: Routledge.
- van der Wouden, Ton. 2001. Collocational behaviour in non content words. In *COLLOCATION: Computational Extraction, Analysis and Exploitation. Proceedings of a Workshop during the 39th Annual Meeting of the Association for Computational Linguistics and the 10th Conference of the European Chapter, Toulouse, France, July 7th*, ed. by Béatrice Daille and Geoffrey Williams, 16-23. Toulouse, France: CNRS - Institut de Recherche en Informatique de Toulouse, and Université de Sciences Sociales.
- van der Wouden, Ton. 2002a. Particle research meets corpus linguistics: on the collocational behavior of particles. to appear in a special issue of the *Belgian Journal of Linguistics* on particles, edited by Ton van der Wouden, Ad Foolen and Piet Van de Craen.
- van der Wouden, Ton. 2002b. Partikels: naar een partikelwoordenboek voor het Nederlands. *Nederlandse Taalkunde* 7, 20-43.
- van der Wouden, Ton, Heleen Hoekstra, Michael Moortgat, Bram Renmans, Ineke Schuurman. 2002a. Syntactic Analysis in the Spoken Dutch Corpus (CGN). In *Proceedings of the third International Conference on Language Resources and Evaluation*, ed. by Manuel González Rodríguez and Carmen Paz Suárez Araujo, 768-773. Paris: ELRA.
- van der Wouden, Ton, Heleen Hoekstra, Michael Moortgat, Ineke Schuurman, Bram Renmans. 2002b. Syntactische annotatie voor het Corpus Gesproken Nederlands (CGN). *Nederlandse Taalkunde* 7, 335-352.
- Wray, Alison. 2002. *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.
- Zwarts, Frans. 1981. Negatief polaire uitdrukkingen 1. *Glott* 4, 35-132.
- Zwarts, Frans. 1986. *Categoriale grammatica en algebraïsche semantiek. Een studie naar negatie en polariteit in het Nederlands*. Groningen dissertation.
- Zwarts, Frans, Ton van der Wouden, Víctor Sánchez Valencia. 2002. The aspectual significance of event particles. to appear in a special issue of the *Belgian Journal of Linguistics* on particles, edited by Ton van der Wouden, Ad Foolen and Piet Van de Craen.

[2f. **Utilisation paragraph:** Only required for proposals to be submitted to Technical Sciences, see Notes]

# Vernieuwingsimpuls / Innovational Research

## Grant application form 2003

VIDI scheme

Please refer to Explanatory Notes when completing this form

### Cost estimates

#### 3a. Budget

	2004	2005	2006	2007	2008	TOTAL
<b>Staff costs: (in k€)</b>	108.172	112.179	116.503	122.449	87.700	547.002
Applicant	83.824	86.000	87.700	87.700	87.700	432.922
Post-doc						
Ph D student (AIO)	24.349	26.179	28.804	34.749		114.080
Support staff						
<b>Non staff costs: (k€)</b>						
Equipment	3.000					3.000
Consumables						
Travel and subsistence	3.600	13.600	3.600	3.600	3.600	26.200
Other						
<b>TOTAL</b>	<b>114.772</b>	<b>125.779</b>	<b>120.103</b>	<b>126.049</b>	<b>89.500</b>	<b>576.202</b>

**3b. Have you requested any additional grants for this project either from NWO or from any other institution?** yes/no (If 'yes', see Notes) no

# Vernieuwingsimpuls / Innovational Research

## Grant application form 2003

VIDI scheme

Please refer to Explanatory Notes when completing this form

### Curriculum vitae

#### 4a. Personal details

Title(s), initial(s), first name, surname: Dr. A. (Ton) van der Wouden

Male/female: Male

Date and place of birth: 20 september 1958, Gouda

Nationality: Dutch

#### 4b. Master's ('Doctoraal')

University/College of Higher Education: Universiteit Leiden, Dutch Department

Date: september 1985 (cum laude)

Main subject: Dutch linguistics

#### 4c. Doctorate

University/College of Higher Education: Rijksuniversiteit Groningen

Date: 23 juni 1994

Supervisor ('Promotor'): Prof. dr. F. Zwarts, prof. dr. H.J. Verkuyl, dr. J. Hoeksema

Title of thesis: *Negative Contexts*

#### 4d. Work experience since graduating

(per appointment: fte, tenured term ('vast') / fixed-term ('tijdelijk'))

6/2001 – now	Postdoctoral researcher Utrecht University/NWO, Syntactic Annotation of the Corpus Gesproken Nederlands	fixed-term, 1.0 fte
6/2000-6/2001	Postdoctoral researcher Utrecht University/NWO, Syntactic Annotation of the Corpus Gesproken Nederlands	fixed-term, 0.5 fte
6/2000-6/2001	Postdoctoral researcher NWO/RijksUniversiteit Groningen, VNC-project "Partikelgebruik in Nederland en Vlaanderen"	fixed-term, 0.5 fte
1/1998-6/2000	Postdoctoral researcher NWO/RijksUniversiteit Groningen, VNC-project "Partikelgebruik in Nederland en Vlaanderen"	fixed-term, 1.0 fte
6/1995-now	Guest researcher Leiden University, General Linguistics	p.m.
spring 1999	Visiting professor ("Lehrbeauftragter"), Carl von Ossietzky University, Oldenburg, Germany	p.m.
spring 1996	Visiting professor ("Lehrbeauftragter"), Carl von Ossietzky University, Oldenburg, Germany	p.m.
7/1996-now	Guest researcher Groningen University, Dutch Department	p.m.
11/1995-3/1996	Post-doctoral fellow, Linguistics Department, Yale University, on a NWO TALENT grant	fixed-term, 1.0 fte
6/1994-6/1995	Post-doctoral researcher Groningen University, Dutch Department.	fixed-term, 1.0 fte

#### 4e. Man-years of research

(see Notes)

ca. 7 years

#### 4f. Brief summary of research over last five years

(max. 350 words)

Since 2000, I have been involved in the Syntactic Annotation of the "Corpus Gesproken Nederlands" (Spoken Dutch Corpus); I participated in setting up the annotation schemes and protocols, and in overseeing, checking and correcting the job students's work. Together with colleagues from Utrecht and Louvain, Belgium I did some first corpus explorations into the syntactic and other linguistic properties of spoken Dutch. Before that, I was the postdoc researcher in the VNC-project "Partikelgebruik in Nederland en Vlaanderen" (Particle Use in the Netherland and Flanders), together with researchers from the Free University in Brussels, aiming at the description of the particles of Dutch and the variation in their usage.

# Vernieuwingsimpuls / Innovational Research

## Grant application form 2003

VIDI scheme

Please refer to Explanatory Notes when completing this form

---

### 4g. International activities

In 1985 I participated in the Generative Linguistic Summerschool in Salzburg, Austria

In 1987 I participated in the Linguistic Institute in Stanford, CA, USA

In 1998 I participated in a Computational Linguistics Summerschool in Lugano, Switzerland

In 1989 I participated in the Linguistic Institute in Tucson, AZ, USA

Winter 1995-96 I was a postdoctoral fellow in the Linguistics Department of Yale University on a NWO TALENT grant.

Over the years, I participated in numerous international conferences.

### 4h. Other academic activities

1999-now board member (chair since 2001) of the Algemene Vereniging voor Taalwetenschap (Linguistic Society of the Netherlands)

1994-now editor of the Dutch linguistics journal TABU

1995-now manuscript reviewer for *Linguistics and Philosophy*, *Journal of East Asian Languages*, *Natural Language and Linguistic Theory*, *Journal of Semantics*, *Linguistics*, *Nederlandse Taalkunde*, *Linguistics in the Netherlands*, *Syncom*, *neerlandistiek.nl*; Cambridge University Press, Oxford University Press, etc.

2001-now member of the editorial board of *neerlandistiek.nl*

2001-2002 editor of a special volume of the *Belgian Journal of Linguistics* (together with Piet Van de Craen and Ad Foolen) (in press).

2000 organiser of "Discourse Particles, Modal And Focal Particles And All That Stuff . . . , An international conference on Particles, Brussels, 8-9 December 2000 (under auspices of the Belgian Linguistic Circle, together with Inge Callebaut, Piet Van de Craen, Ad Foolen and Frans Zwarts).

1999-now co-organiser of the yearly conference "TIN-dag" (Taalkunde in Nederland).

2000, 2001 editor of *Linguistic in the Netherlands* (together with Helen de Hoop and Hans Broekhuis, respectively)

1996-2001 co-organiser of the Groningen PIONIER-conference *Perspectives on Negation* and co-editor of the Proceedings.

### 4i. Scholarships and prizes

### List of publications

#### 5. Publications:

##### -International (refereed) journals

- S Ton van der Wouden. Particle research meets corpus linguistics: on the collocational behavior of particles. *Belgian Journal of Linguistics*, 16, 2003, 151-174 (in press).  
Frans Zwarts, Ton van der Wouden en Víctor Sánchez Valencia. The aspectual significance of event particles. *Belgian Journal of Linguistics*, 16, 2003, 175-208 (in press).  
Ad Foolen and Ton van der Wouden, Introduction [to the special issue on particles], *Belgian Journal of Linguistics*, 16, 2003, 1-6 (in press).  
Ton van der Wouden. Indefinite pronouns. *Linguistic Typology* 4 (2000), 281-291.  
Ton van der Wouden. Not everything is possible. *GLOT International*, 4(3):17-18, 1999.  
Ton van der Wouden. Negative contexts. *GLOT International*, 1(5):9-10, 1995.

##### -National (refereed) journals

- S Ton van der Wouden, Heleen Hoekstra, Michael Moortgat, Ineke Schuurman en Bram Renmans. Syntactische annotatie voor het Corpus Gesproken Nederlands (CGN). *Nederlandse Taalkunde* 7, 4 (2002), 335-352.
- S Ton van der Wouden. Partikels: Naar een partikelwoordenboek voor het Nederlands. *Nederlandse Taalkunde* 7, 1 (2002), 20-43.  
Ton van der Wouden. Over *zeker*, zeker? zeker! *TABU* 30 (2000), 63-89.  
Ton van der Wouden. [signalement van] Dirk Geeraerts, Stefan Grondelaers en Dirk Speelman, Convergentie en divergentie in de Nederlandse woordenschat. Een onderzoek naar kleding- en voetbaltermen, Amsterdam, Meertens Instituut, 1999. *TABU* 30 (2000), 94-95.  
Ton van der Wouden. Smeermiddelen van de taal. Partikels in het Nederlands. *Onze Taal*, 68(11) (1999), 292-95.  
Olga van Marion en Ton van der Wouden. Graag scherpe critiek! *Musaeus, tijdschrift voor Renaissance-letterkunde*, 6(4) (1999), 5-8.
- S Marc van Oostendorp en Ton van der Wouden. Corpus internet. *Nederlandse Taalkunde*, 3(4) (1998), 347-361.
- S Ton van der Wouden. Dat had niet zo hoeven: Modaliteit en negatie in de nieuwe ANS. *Nederlandse Taalkunde*, 3(3) (1998), 237-252.
- S Ton van der Wouden. Partikels. *Nederlandse Taalkunde*, 3(1) (1998), 76-77.  
Ton van der Wouden. Waar Machteld nou? *TABU*, 28(4):159-161, 1998.  
Ton van der Wouden. Verboden op het werk te komen. *Onze Taal*, 67(1) (1998), 14.  
Ton van der Wouden. Hoeven. *TABU*, 26(3) (1996), 164-182.  
Ton van der Wouden. De reflectie van logische relaties in woordstructuur: de distributie van het prefix *on*. *TABU*, 25(1) (1995):61-71.  
Ton van der Wouden. Het partitieve *er*. *Tabu*, 25(4) (1995):205-208.  
Ton van der Wouden. Lijftelwoorden. *Onze Taal*, 64(2/3) (1995):36.  
Ton van der Wouden. Dubbele ontkenningen. *Onze Taal*, 64(7/8) (1995), 186-87.
- S Ton van der Wouden. Negatieve omgevingen. *Onze Taal*, 64(11) (1995), 284-285.
- S Ton van der Wouden. Beperkingen op het optreden van lexicaal elementen. *De Nieuwe Taalgids*, 85(6) (1992), 513-38.  
Ton van der Wouden. Oorlogsmetaforen van alledag. *Onze Taal*, 61 (1992), 7/8:149.  
Ton van der Wouden. Een mondvol woordgrappen. *Onze Taal* 61 (1992), 7/8:166.  
Ton van der Wouden. Een land van brood en honing. *Onze Taal* 60 (1991), 5: 10-11.  
Dirk Heylen, Michael Moortgat, and Ton van der Wouden. Categoriale ontleding: Theorie en praktijk. *Spektator*, 18(1) (1988).  
Ton van der Wouden. Positief polaire uitdrukkingen 1. *Glott*, 11 (1988), 165-90.  
Ton van der Wouden. Dubbele ontkenningen. *Glott*, 10 (1987), 485-501.

# Vernieuwingsimpuls / Innovational Research

## Grant application form 2003

VIDI scheme

Please refer to Explanatory Notes when completing this form

### -Books, or contributions to books

- S Ton van der Wouden, Heleen Hoekstra, Michael Moortgat, Bram Renmans and Ineke Schuurman. Syntactic Analysis in the Spoken Dutch Corpus (CGN). In Manuel González Rodríguez and Carmen Paz Suárez Araujo (eds.), *Proceedings of the third International Conference on Language Resources and Evaluation*, Paris, ELRA, 2002, 768-773.
- Jack Hoeksema, Hotze Rullmann, Víctor Sánchez Valencia, en Ton van der Wouden (red.). *Perspectives on Negation*. Amsterdam: John Benjamins, 2001.
- Ton van der Wouden en Hans Broekhuis (red.) *Linguistics in the Netherlands 2001*. Amsterdam/Philadelphia: John Benjamins, 2001.
- S Ton van der Wouden. Collocational behaviour in non content words. In *COLLOCATION: Computational Extraction, Analysis and Exploitation. Proceedings of a Workshop during the 39th Annual Meeting of the Association for Computational Linguistics and the 10th Conference of the European Chapter, Toulouse, France, July 7th*, ed. by Béatrice Daille and Geoffrey Williams, 16-23. Toulouse, France: CNRS - Institut de Recherche en Informatique de Toulouse, and Université de Sciences Sociales, 2001.
- Ton van der Wouden. Three modal verbs. In *Zur Verbmorphologie germanischer Sprachen*, ed. by Sheila Watts, Jonathan West and Hans-Joachim Solms, 189-210. Tübingen, Niemeyer, 2001.
- S Heleen Hoekstra, Michael Moortgat, Ineke Schuurman en Ton van der Wouden. Syntactic Annotation for the Spoken Dutch Corpus Project (CGN). In *Computational Linguistics in the Netherlands 2000*, ed. by W. Daelemans, K. Sima'an, J. Veenstra and J. Zavrel, 73-87. Amsterdam/New York, Rodopi, 2001.
- Helen de Hoop and Ton van der Wouden (red.). *Linguistics in the Netherlands 2000*. Amsterdam [etc.]: John Benjamins, 2000.
- S Ton van der Wouden. Focus on appendices in Dutch. In *Linguistics in the Netherlands 2000*, red. Helen de Hoop en Ton van der Wouden, 233-245. Amsterdam/Philadelphia: John Benjamins, 2000.
- Ton van der Wouden. *Verboden op het werk te komen. Klein woordenboek van Vlaamse taal- en andere eigenaardigheden*. SIWU, Enschede, 1998.
- Ton van der Wouden. On the development of marked negation systems: The Dutch situation in the seventeenth century. In Richard M. Hogg and Linda van Bergen, editors, *Historical Linguistics 1995, Volume 2*, pages 311-329. Amsterdam [etc.]: John Benjamins, 1998.
- S Ton van der Wouden. Niet huilen maar: negation, directives and particles. In Sjef Barbiers, Johan Rooryck, and Jeroen van de Weijer, editors, *Small words in the big picture. Squibs for Hans Bennis*, number 2 in HIL Occasional Papers, pages 125-128. HIL, Leiden, 1998.
- S Ton van der Wouden. *Negative Contexts. Collocation, polarity, and multiple negation*. Routledge, London and New York, 1997.
- Ton van der Wouden. Litotes and downward monotonicity. In Heinrich Wansing, editor, *Negation: a notion in focus*, 145-168. Walter de Gruyter, Berlin, 1996.
- S Ton van der Wouden. *Moelijk* is (not) difficult. In Marcel den Dikken and Kees Hengeveld, editors, *Linguistics in the Netherlands 1995*, pages 203-14. Benjamins, Amsterdam, 1995.
- S Ton van der Wouden. *Negative Contexts*. Proefschrift Groningen, 1994.
- S Ton van der Wouden. Polarity and 'illogical negation'. In Makoto Kanazawa and Christopher J. Piñón, editors, *Dynamics, polarity, and quantification*, volume 48 of *CSLI Lecture Notes*, pp 17-45. Center for the Study of Language and Information, Stanford, 1994.
- Víctor Sánchez Valencia, Ton van der Wouden and Frans Zwarts. Polarity, Veridicality, and Temporal Connectives. In Paul Dekker and Martin Stokhof, editors, *Proceedings of the Ninth Amsterdam Colloquium, December 14-17, 1993*, pages 587-606. ILLC / Department of Philosophy, University of Amsterdam, 1994.
- S Ton van der Wouden and Frans Zwarts. A semantic analysis of negative concord. In Utpal Lahiri and Adam Zachary Wyner, editors, *SALT III: Proceedings of the third conference on semantics and linguistic theory*, pages 202-219. Cornell University Department of Modern Languages and Linguistics, 1993.
- S Ton van der Wouden en Frans Zwarts. Negative concord. In Dicky Gilbers en Sietze Looyenga, red., *Language and cognition 2. Yearbook 1992 of the research group for linguistic theory and knowledge*

# Vernieuwingsimpuls / Innovational Research

## Grant application form 2003

VIDI scheme

Please refer to Explanatory Notes when completing this form

- representation of the University of Groningen, pp. 317-331. TENK, Groningen, 1992.
- S Ton van der Wouden. Prolegomena for a multilingual description of collocations. In *EURALEX '92 Proceedings I-II. Papers submitted to the 5th EURALEX International Congress on Lexicography in Tampere, Finland*, ed. by Hannu Tommola, Krista Varantola, Tarja Salmi-Tolonen, and Jürgen Schopp, 449-456, Tampere. Department of Translation Studies, university of Tampere, 1992.
- Ton van der Wouden. Celex: Building a multilingual multifunctional polytheoretical lexical database. In T. Magay and J. Zsigány, editors, *BudaLEX '88 Proceedings. Papers from the EURALEX Third International Congress, Budapest, 4-9 September 1988*. Akadémiai Kiadó, Budapest, 1990.
- Pim van der Eijk and Ton van der Wouden. A modular lexicon environment for NLP. In U. Zernik, editor, *Proceedings of the First International Lexical Acquisition Workshop*, Detroit, 1989.
- Dirk Heylen, Michael Moortgat, and Ton van der Wouden. Categoriale ontleding: Theorie en praktijk. *Spektator*, in Anneke Neijt and Dik Bakker (red.): *Computerlinguïstiek, een overzicht in artikelen*, Dordrecht, Foris, 1990.

### -Other

- Michael Moortgat, Ineke Schuurman en Ton van der Wouden. CGN Syntactische Annotatie. Versie januari 2002. Nota UiL-OTS Utrecht/CCL Leuven, CGN-annotatie-CD's, Den Haag, Taalunie (november/december 2002) (gedistribueerd door ELDA, Paris).
- Michael Moortgat, Ineke Schuurman en Ton van der Wouden. CGN Syntactische Annotatie. Nota UiL-OTS Utrecht/CCL Leuven, CGN-annotatie-CD's, Den Haag, Taalunie (april 2001).
- Ton van der Wouden. Partikelgebruik. *De Woordenaar. Nieuwsbrief van het Matthias de Vries-Genootschap*, 1(2) (oktober 1997), 17, 1997.
- Laurence R. Horn. Flaubert-triggers, squatatieve negatie, en andere grillen van de grammatica. *TABU* 26, 183-205, 1996. (vert. Ton van der Wouden)
- William A. Ladusaw. Een compositionaliteitspuzzel in dubbele ontkenningen. *TABU* 26, 213-234, 1996. (vert. Ton van der Wouden)
- Ton van der Wouden. [gepromoveerd]. *Dokumentaal*, 23(3):122-4, 1994.
- Ton van der Wouden. Logische patronen in natuurlijke taal. *Dokumentaal*, 22(1):2-3, 1993.
- Ton van der Wouden and Wietske Sijtsma, red. *Computational Linguistics in The Netherlands. Papers from the First CLIN-Meeting*. OTS Working Papers nr. OTS-WP-CL-91-001, 1991.
- Ton van der Wouden. Een computerlinguïstische vingeroefening. *Tabu*, 15, 2, 1985.
- Ton van der Wouden. If Nederlands then go to informatica. *META*, 19,5:16-19, 1985.
- John van Lit en Ton van der Wouden. Comparatieve syntaxis. *META*, 19,4:12-16, 1985.
- Ton van der Wouden. Positief polaire uitdrukkingen in het Nederlands. Doctoraalscriptie, Vakgroep Nederlands Leiden, 1985.

### Signature

I hereby declare that I have completed this form truthfully:

Name:

Place:

Date:

Please submit the application to NWO in electronic form (pdf format is required!) using the IRIS system, which can be accessed via the NWO website ([www.nwo.nl/vernieuwingsimpuls](http://www.nwo.nl/vernieuwingsimpuls)).

# Vernieuwingsimpuls / Innovational Research

## Grant application form 2003

VIDI scheme

*Please refer to Explanatory Notes when completing this form*

---

### Post to NWO / Vernieuwingsimpuls

**To streamline the processing of applications, please complete the form below and post a print-out of this page together with any relevant documents to NWO.**

---

**I the undersigned declare that I have today posted (tick relevant documents):**

**Institutional guarantee from Board ('Inbeddingsgarantie College van Bestuur')**

(to be sent by post, strongly desired for VIDI) will be/is sent separately, ref. **916 - 2 - 2003**

**Address list of 'non-referees'**

(to be sent by post before submission DEADLINE, optional for all applicants, max. 3, see Notes)

**Name of applicant: dr. A. (Ton) van der Wouden**

**Place: Leiden**

**Date: February 18<sup>th</sup>, 2003**

**Postal address: ATW, PO Box 9515, 2300 RA Leiden**

**NWO Council area: GW**

---

Send the documents to:

NWO/Vernieuwingsimpuls

Council area: **GW**

P.O. Box 93138

2509 AC The Hague

(The Netherlands)