CLCG Research Program 2005-2009

CLCG, FdL, RuG

May 2005

1 Purpose

This document outlines the research program of the Center for Language and Cognition, Groningen, a research institute in the Faculty of Arts. We wish in particular to identify research priorities and goals for the period 2005-2009.

2 CLCG

The Center for Language and Cognition, Groningen (CLCG), is one of three research institutes within the Faculty of Arts. CLCG is the organizational home of all of the linguistic research of the faculty, and all staff members with recognized research time for linguistic research are members of CLCG, including permanent staff in several different departments, postdoctoral fellows financed by various research agencies, and graduate students striving for a Ph.D. in Linguistics. It is CLCG's corporate goal to conduct linguistic research both of a fundamental and of an applied nature, the latter taking various forms, including contract research, cooperative research with industry and government, and consultancy. Finally, the cognitive neuroscience work in Groningen provides a special opportunity for linguistic contributions, particularly for researchers in neurolinguistics (including aphasiology and dyslexia) and to a lesser extent in language acquisition, computational linguistics and linguistic theory.

CLCG recognizes the special responsibility of the University of Groningen and its Arts Faculty in promoting research in the Germanic, Romance, Slavic and Finno-Ugric languages; in the Dutch languages, including especially Frisian and Lower Saxon; in the many applications of linguistics, in particular including not only educational applications in literacy education and in foreign language learning and teaching, but also applications in communications, in the computational processing of language, and in dyslexia and aphasiology.

2.1 Research Groups

There are at present six research groups in CLCG. These groups are defined by the linguistic subdiscipline the group focuses on, and the groups are the motors producing ongoing research. Every group meets regularly, on average at least twice a month (normally with varying participation). In some groups the meetings are focused along

particular research lines, and not every meeting involves the entire group, not even potentially. New research is presented at these meetings, especially research from group members, but also important external developments. Because we view regular scientific discussion as the primary purpose of the group, some members may prefer groups in which their theoretical preferences are shared. But we proceed from group definitions in terms of subdisciplines rather than in terms of theoretical preferences or theoretical frameworks.

These are listed below with the number of permanent staff members in the group (each with research time between 30% and 40% of his of her appointment). The groups are the primary initiators of research directions and research activities. They are the source of the program below.

Group	perm.staff	%
Syntax & Semantics	11	19.6
Discourse & Communication	8	14.3
Variation & Change	14	25.0
Linguistic and Literacy Development	12	21.4
Computational Linguistics	6	10.7
Neurolinguistics	5	8.9
Total	56	≈100

2.2 Research Groups and Programs

We turn to a presentation of the research themes of the different groups and the permanent staff members who make up those groups.

Syntax & Semantics The syntactic research examines the syntactic structure of the human language capacity and the way it is embedded in our general cognitive capacity. The 2005 – 2009 research program is motivated, among other things, by the so-called logical problem of language acquisition and the evidence comes from comparative and typological research on the synchronic and diachronic variation among natural languages. The semantic research focuses on the restrictions on patterns related to quantification, negation and inference, and the semantics of negative polarity, degree adverbs. Further research includes topics on discourse semantics, in particular information structure and the dynamic properties of tense and aspect.

Members (05/2005): Reineke Bok-Bennema, Ger de Haan, Petra Hendriks, Jack Hoeksema, Brigitte Kampers-Mahne, Jan Koster (coörd.), Alice ter Meulen, Arie Molendijk, Ron van Zonneveld, Jan-Wouter Zwart, Frans Zwarts (currently rector).

Discourse & Communication The goal of this group's research is a deeper understanding of face-to-face and mediated discourse. We see discourse as a situated, culture- and context-sensitive joint activity of participants using language and non-verbal semiotic systems. A comprehensive account of discourse calls for a variety of approaches which analyze its semantics and pragmatics, its textual and

interactional structures, as well as the cognitive, social, and historical processes in which it is embedded. This challenge is addressed in this group by bringing together a variety of backgrounds and expertise, and by actively seeking collaboration with other groups.

Members (05/2005): Marcel Bax, Jeanine Deen, Markus Egg, Titus Ensink, John Hoeks, Harrie Mazeland, Gisela Redeker (coörd.), Christoph Sauer.

Language Variation and Change is concerned with synchronic and diachronic variation in a great variety of Indo-European and Finno-Ugric languages and dialects. From their different language backgrounds, members of this group address their research questions using a "bottom-up" approach, i.e. starting with the collection and analysis of a solid body of data and examining its implications for theoretical claims. While aiming for contributions in dialectology, historical linguistics and language contact, they thus likewise add to a better understanding of the language families they specialize in and to a number of current theoretical issues, which are specified below in § 3.

Members (05/2005): Agnes Bie-Kerekjarto, Helene Brijnen, Bram ten Cate, Maria Czibere, Cornelius Hasselblatt, Tette Hofstra, Peter Houtzagers, Bob de Jonge, Geart van der Meer, Hermann Niebaum, Muriel Norde (coörd), Simon Reker, Nanne Streekstra, Gerry Wakker.

Computational Linguistics focuses on natural language processing by computer. Areas of special interest are wide-coverage grammars (typically for Dutch), finite automata, machine learning of natural language, corpus-based methods, computational dialectology, computational modeling of communication and discourse, and the interface of language and speech technology.

Members (05/2005): Leonie Bosveld, Gosse Bouma, Dicky Gilbers, Charlotte Gooskens, John Nerbonne, Gertjan van Noord (coörd.)

Language and Literacy Development across the Life Span focuses on empirical research into the development of language and literacy across the life span. The group's research deals with development in both first and second or foreign languages and addresses various aspects of knowledge and use of language and literacy. Development is studied across the lifetime, in various contexts (home, school) and with a focus on individual and situational variables that affect learning.

Members (05/2005): Heike Behrens, Jan Berenst, Nanette Bienfait, Kees de Bot, Pieter Breuker, Kees de Glopper (coörd.), Hilde Hacquebord, Angeliek van Hout, Erik Kwakernaak, Wander Lowie, Aart Pouw, Marjolijn Verspoor

Neurolinguistics Neurolinguistics studies the neural basis of language with a focus on impaired language (including especially aphasiology, but also developmental disorders such as dyslexia) and on the use of neuro-imaging techniques.

Members (05/2005): Roelien Bastiaanse (coörd.), Pieter Been, Gerard Bol, Roel Jonkers, Laurie Stowe.

2.3 Foci, Commonalities, Cross-Talk

The six research groups involve a large number of research foci, methodologies, and theoretical frameworks. Nonetheless there are clear areas of concentration. Grammar — syntax and semantics — is central not only to the syntax and semantics group, which focuses on the theoretical analysis of syntax and semantics, but is quite central to one or more research lines in **every** other group. Perhaps distinctive to Groningen is the unusually high level of interest in applications of linguistics. It is common to find interest in foreign language acquisition in linguistic research institutes, which is likewise strong in Groningen, but CLCG also boasts of demonstrated, successful work in applications such as language-based information systems, information search and extraction (computational linguistics); aphasiology and dyslexia (neurolinguistics); literacy education and language loss (language development); and computer-assisted language learning (computational linguistics and language development).

Rather than note each point of contact between groups separately here, we note them as they arise in the programs of the research groups (immediately below).

3 Program

CLCG seeks to continue its research while respecting the responsibilities in § 1. While Computational Linguistics, Neurolinguistics, Syntax and Semantics and Linguistic and Literacy Development have been most successful in attracting the external funding that is sorely needed for concentrated innovative efforts, all of the groups recognize the need for efforts in this area.

As we sketch the research lines in the six research groups below, please bear in mind that individuals and research lines may not be strictly identified. Several researchers are active in more than one research line.

Syntax & Semantics This group focuses on general, theoretical issues in the analysis of grammar.

Foundations of syntactic theory The theoretical orientation of the syntactic research program is within the general theory known as Minimalism. Within this general framework, the Groningen approach has its own character in a number of respects: (a) asymmetrical Merge, (b) strict locality of syntactic relations (sisterhood) and (c) Pied Piping as the condition allowing this strict locality in the first place and as the main dimension of parameterization across languages. An issue that will be further explored is the desirability of dynamic, strictly derivational grammars versus the more "cartographic" models of grammar. These matters will be explored against a deeper question, namely whether grammars are about abstract objects (Platonism) or about the mental states of an idealized speaker-hearer (Conceptualism).

Specific syntactic topics Empirical foci are to be derived from theoretical concerns, but if we may extrapolate from the current issues, research will focus on verb movements and morphology, verb chains (including auxiliary elements, tense markers and complementizers), agreement phenomena and case marking, coordination and parataxis. Increasingly, the group also studies these phenomena from a typological point of view.

Growth and change of the lexicon The semantically oriented work of the group will continue its corpus-based research on growth and change of the lexicon. Among other topics, special attention will be given to negative polarity items and to lexical domains in with rapid growth and specialization, such as in the case of degree adverbs and minimalizing expressions.

Semantics of tense and aspect This research line will continue, particularly in Romance, albeit with less energy due to personnel developments. Some of the research will be focused on temporal reasoning and on aspectual adverbs, connecting their polarity behavior to conditionals and counterfactuals in dynamic semantics.

Further integration Next to the standard orientations in syntactic and semantic theory (Minimalism, model theory), part of the discourse-oriented work of the group will also seek embedding in the general framework of Optimality Theory. Apart from its theoretical concerns, the group generally aims at further integration of its results with current work in psycholinguistics and neurolinguistics (ERP and fMRI studies).

Discourse and Communication The joint focus of the group is the study of communication as discourse. We investigate the structures and situated practices of interactions, texts, and audiovisual communication. Acknowledging the complexity of this objective, our research builds on theories and methods from a wide range of approaches to language and communication, including semantics, pragmatics (incl. historical pragmatics), and semiotics; sociolinguistics, psycho- and neurolinguistics; systemic-functional, cognitive, computational, and interactional linguistics; frame theory and critical discourse analysis.

Interactive discourse We study the situated joint construction of discourse by analyzing corpora of institutional interactions and everyday conversations, including historical data. We focus on lexical, grammatical, prosodic, and non-verbal information and patterns of sequential organization. For a systematic account of the linguistic resources and their use in interaction, we envisage close collaboration with the CLCG Computational Linguistics and Syntax and Semantics groups. Specific hypotheses arising from our analytical results will be tested with psycholinguistic and neurolinguistic (ERP and/or fMRI) experiments.

Text and Context We will investigate textual structures and the processing of spoken and written monologue discourse by analyzing corpora and performing experiments. We focus on contextually enriched interpretations, framing, and coherence to gain insight into the interaction of linguistic resources with world knowledge and context. On this basis we will formulate design principles and develop methods for the evaluation and optimization of texts (possibly in collaboration with the CLCG Language and Literacy Development over the Life Span (LANSPAN) group).

Mediated communication The semiotics and practices of mediated communication will be studied with discourse analysis and experimentation. We will investigate the multimodal composition of media representations, the role of the physical realization of messages, and the exploitation of media affordances. This research will contribute to a theory of multimodal document design. One of its applications will be the development of pictorial information materials for the language-impaired or illiterate (in collaboration with the Expertisecentrum voor Taal, Onderwijs en Communicatie).

Communication and culture Drawing on the ongoing research in interaction, text, and mediated communication, we study the ways in which cultural identities are constituted and negotiated. We focus on ethnicity, gender, and power, and investigate leadership, intercultural communication, institutional image management, and the historical and evolutionary roots of human symbolic interaction. We will seek collaboration with the CLCG Language and Literacy Development over the Life Span (LANSPAN) group, esp. when second- and foreign-language proficiency are at issue.

Language Variation and Change will concentrate on the following areas in 2005-2009:

Synchronic variation This line will focus on the encoding of tense, mood and aspect (Ancient Greek, German, Romance); markedness and definiteness (Ancient Greek); semantic structures (Hungarian); dialect geography (Slavic, German); lexicography (English, Balto-Finnic); communicative strategies (Romance); and language attitudes (Finno-Ugric, Slavic). An important question with respect to Romance is how synchronic variation leads to diachronic change. We are exploring the idea contextually non-triggered communicative strategies could lead to linguistic change. In Ancient Greek we are exploring semantic as opposed to pragmatic factors which determine the choice of aspect.

Diachronic variation We investigate morphosyntactic changes such as grammaticalization, deflexion, changes in the encoding of tense, mood and aspect (German, Scandinavian, Finno-Ugric and Romance), but also reconstruction of older stages of dialect groups and languages (Dutch, Slavic). Our grammaticalization work focuses on the hotly debated principle of unidirectionality. Though most historical linguists have now abandoned the idea that grammatical change is strictly unidirectional with respect to grammaticalization, it is unclear which changes are genuinely counterdirectional, i.e., "degrammaticalizations". Using data from several languages we will seek "degrammaticalization parameters", parallel to Lehmann's famous grammaticalization parameters. We will use data drawn from secondary sources (with the aid of specialists) and from a corpus-based analysis of the rise, maintenance and fall of suffixes in Continental Scandinavian from ca. 700 on.

Language contact in past and present times We are interested in contact-induced language change (Finnic, Scandinavian, Romance); the dialect-standard language continuum (in the province of Groningen and on both sides of the Dutch-German border); the influence of High and Low German on neighboring languages and language families (Sorbian, Kashubian, Scandinavian and Finnic). We focus on Germanic-Finnic interference as shown in the emergence of compound (or phrasal) verbs in Finnic, where the neighboring Germanic languages are perhaps responsible for this novel word formation. Research has to be morphosyntactic and pragmatic instead of only morphological (and historical-phonological) as in the past. Only then we will be able to answer the question whether the changes are contact-induced, internal or merely apparent, i.e. dependent on description. We will also try to clarify the linguistic history of the north and east of the Netherlands and adjoining areas in Germany, particularly where vis-à-vis the Low German-Dutch dialect continuum. Scholars specializing in Dutch generally regard these dialects as Dutch, a result of the gradually increasing Dutch cultural influence since the Middle Ages, the influence of the standard language, etc. Linguistically, however, the development of these dialects should be viewed in their Low German context.

Linguistic and Literacy Development across the Life Span examines a range of issues concerning primary language acquisition, the development of literacy, the learning of second and foreign languages, and language loss. The main streams of research focus on

Language acquisition Research on language acquisition takes an emergentist or usage-based approach and studies how the grammatical and semantic structure of the target language are derived from the input language. The focus is on the communicative interaction with speakers of the target language and on the cognitive skills and social-communicative mechanisms that support language acquisition.

Development of language use in preschool and kindergarten Research on the development of language use in preschool and kindergarten studies how different situations or participation frames contribute to language development. The focus is on the communicative interaction that takes place in situations where young children are engaged in play, work or book reading and on the learning opportunities and learning effects these situations provide.

Second language learning, bilingualism and dynamic systems theory Research on second language learning, bilingualism and dynamic systems theory aims to describe and explain processes of second or foreign language learning in various settings (school and non-school). Dynamic systems theory is used to account for language development as a process of self-organization under highly constrained inputs. Intra-individual variability is used as an indicator of developmental transitions. The focus of these studies is on syntactic and phonological development.

The development of bilingual education in the Netherlands In recent years there has been a tremendous growth in the number of schools offering different types of bilingual education. Research in these schools focuses on the development of proficiency in the target language and the mother tongue. In addition there are projects on quality assessment in these schools and teacher qualifications.

Language, literacy and learning Research on language, literacy and learning studies how reading, writing and oral communication contribute to learning in primary, secondary and higher education. The focus is on processes of communication and knowledge construction and on the way reading, writing and talking are learned and/or used by learners and teachers.

Language development and aging Research on language development and aging studies how aging individuals perform on second language learning tasks and how aging affects processes of language loss. The focus is on the cognitive and neurological factors that affect the processing of language.

Computational Linguistics The computational linguistics research in CLCG combines linguistic sophistication and corpus-based evaluation methodology. In particular, we foresee the following research activities.

Wide coverage computational analysis of Dutch Research in this area has culminated in the Alpino parsing system for Dutch. We expect to continue progress by focusing on problems of processing efficiency, processing robustness, and disambiguation. In addition, the Alpino grammar will need to be extended for semantic analysis. This in turn requires e.g. the integration of ontological information and discourse information (including pronoun resolution, etc.) We look forward to cooperation with the CLCG Discourse and Communication group in order to make progress on semantics and pragmatics.

We will also investigate potential applications for wide coverage computational analysis of Dutch. We foresee that we can improve question answering applications, and more generally information retrieval applications by exploiting the automatic fine-grained syntactic and semantic analysis tools.

The availability of large annotated corpora allows for various new opportunities in corpus-linguistics. A further exciting new possibility is the use of huge machine-annotated corpora. The CL group will also be involved in various corpus annotation efforts. Not only will we attempt to extend the amount of annotated corpus material, but we will also focus on more involved levels of annotation (such as discourse annotation, and semantic annotation). Improved semi-annotated annotation techniques is a closely related research objective. In this context we also focus on semi-automated techniques for lexical acquisition, for instance for the fine-grained lexical description of semi-fixed expressions, as well as the acquisition of subcategorization frames for infrequent words.

Computational studies of language variation We shall continue development and analysis of dialect distance metrics, including as novel elements lexical and syntactic distances, and we shall continue to seek to exploit the more exact measures in investigation of external determinants of variation such as geography and social contact. We also plan to investigate intelligibility (which is threatened by unimpeded variation) both computationally and by means of perception experiments.

Finally, the status of variation in constraint-based phonological models is an issue, in particular segmental and rhythmic variability. These may be studied via their acoustic and perceptive characteristics, and their patterning in acquisition. We plan to cooperate with the CLCG group focused on 'Language Variation and Change'.

Visual communication In the context of visual communication (involving diagrams, graphical user interfaces, and other information visualizations), we plan to focus on the modeling link, i.e., the link between the graphics and the application domain (what is represented), trying to understand the usability of (components of) visual representations.

Neurolinguistics studies how and where language is represented in the brain, working with studies of deficient language functioning and with neuro-imaging, where the latter takes place at the Neuro Imaging Center (NIC):

Aphasiology investigates grammatical disorders in aphasia, focusing on the comprehension and production of verbs and sentences in languages with a wide range of structural properties (Dutch, Italian, Turkish, Russian). A European research collaboration will be started to include more languages. This research aims to identify the role of syntactic and morphological properties and the linguistic specifics of the language behavior of aphasics with a grammatical deficit. In addition to the behavioral data, ERP and fMRI data of non-brain-damaged subjects are used to find out where word order and verb processing and production takes place. There is an additional research line on phonological disorders which investigates the relation between production and perception deficits in phonology. This project is a cooperation with Utrecht and Nijmegen, and there are plans to apply for a new project.

Dyslexia The research traces the developmental path in familial dyslexia by means of regular assessment of perception and language in dyslexic families and controls starting at two months and continuing until reading assessment if possible. The odds that children from dyslexic families will become dyslexic are approx. 50%. Until the age of 4 years assessment by Event Related brain Potentials (ERP) is emphasized. Later on the emphasis shifts to behavioral tests. Principal aims of the program are 1) to find early precursors of dyslexia to allow for early diagnosis and treatment in the future; 2) to explore the comorbidity of developmental dyslexia with SLI and ADHD; 3) to explore the possibilities of sub-typing by brain mapping techniques; and 4) to test the hypothesis that a magnocellular deficit in the brain may be causally linked to dyslexia.

Neuroimaging focuses on language processing in healthy speakers, esp. how various forms of potentially conflicting information are combined during comprehension, including (1) the role of the verb and its argument structure, a study that will use behavioral and neuro-imaging data; (2) the processing of ambiguities, idioms, and so called 'garden path sentences' both at the syntactic and the lexical level, a project using behavioral and fMRI data; (3) the interpretation of WH-questions, for which behavioral, fMRI and eye-tracking data will be used. Attention will be paid to brain organization as revealed in normal humans and in various patient populations.

Language Acquisition Disorders focuses on grammatical disorders in children with different etiologies: Specific Language Impairments (SLI); Down's syndrome, PDD-NOS and Hearing Impairments. The attention will shift from analysis of the spontaneous speech to experimental research. Also, an effort will be made to translate the findings of the fundamental research for clinical purposes, by developing a test for children with language deficits. The comorbidity of SLI and developmental dyslexia will be explored.