The CLARIN-NL Project

Jan Odijk
UiL-OTS, Utrecht University
Janskerkhof 13
3512 BL Utrecht
E-mail: j.odijk@uu.nl

Abstract

In this paper I present the CLARIN-NL project, the Dutch national project that aims to play a central role in the European CLARIN infrastructure, not only for the preparatory phase, but also for the implementation and exploitation phases. I argue that the way the CLARIN-NL project has been set-up can serve as an excellent example for other national CLARIN projects, for the following reasons: (1) it is a mix between a programme and a project; (2) it offers opportunities to seriously test standards and protocols currently proposed by CLARIN, thus providing evidence-based requirements and desiderata for the CLARIN infrastructure and ensuring compatibility of CLARIN with national data and tools; (3) it brings the intended users (humanities researchers) and the technology providers (infrastructure specialists and language and speech technology researchers) together in concrete cooperation projects, with a central role for the user's research questions,, thus ensuring that the infrastructure will provide functionality that is needed by its intended users.

1. Introduction

In this paper I present the CLARIN-NL project, the Dutch national project that aims to play a central role in the European CLARIN infrastructure, not only for the preparatory phase, but also for the implementation and exploitation phases. I argue that the way the CLARIN-NL project has been set-up can serve as an excellent example for other national CLARIN projects, for the following reasons: (1) it is a mix between a programme and a project; (2) it offers opportunities to seriously test standards and protocols currently proposed by CLARIN, thus providing evidence-based requirements and desiderata for the CLARIN infrastructure and ensuring compatibility of CLARIN with national data and tools; (3) it brings the intended users (humanities researchers) and the technology providers (infrastructure specialists and language and speech technology researchers) together in concrete cooperation projects, with a central role for the user's research questions,, thus ensuring that the infrastructure will provide functionality that is needed by its intended users.

2. CLARIN-NL

The CLARIN-NL project is a large national project in the Netherlands which aims to play a central role in the Europe-wide CLARIN infrastructure. CLARIN-NL offers scholars the tools to allow computer-aided language processing, addressing one or more of the multiple roles language plays (i.e. carrier of cultural content and knowledge, instrument of communication, component of identity and object of study) in the Humanities and Social Sciences.

CLARIN-NL aims to design, construct, validate, and exploit a research infrastructure that is needed to provide a sustainable and persistent *e*Science working environment for researchers in the Humanities, and Linguistics in particular, who want to make use of language resources and the technology to use these resources for their research. This infrastructure will provide these researchers with a wide variety of resources and services, intelligent access methods for exploring the resources and innovative ways of combining different resources into virtual collections, so that information hidden in unstructured textual and multimedia documents can be disclosed. Inter-operability of independently developed resources and services will be key for a properly functioning infrastructure. The infrastructure will be easy to use for non-technical researchers. Targeted dissemination activities, educational programmes and training sessions will enable a whole generation of researchers and students to acquaint themselves with this new research methodology and the potential for groundbreaking research it offers, creating an advanced scientific environment in the Netherlands that will attract top-researchers and students from abroad.

CLARIN-NL forms the Netherlands national counterpart of the CLARIN enterprise on the European level (CLARIN-EU). It therefore resembles and complements the preparatory project that is currently being executed on the European level (CLARIN-prep). Many of the activities and sub-projects within CLARIN-NL implement activities in the Netherlands that in the programme of work for CLARIN-prep are envisaged to take place in every participating country and that will be funded through the national contributions to CLARIN. Such activities include (1) the design and implementation of the infrastructure technology; (2) application projects in which technology providers and the intended users integrate local repositories and set up local services for prototypical test installations as initial demonstrators, enabling evidence-based contributions to the discussion on standards and best practices for inter-operability, and to contribute to the survey of requirements for the infrastructure technology; (3) the preparation of an essential data collection and service set for the locally relevant languages (ideally on the basis of existing tools and data) that allows for testing and validation of proposed standards, services and tools in the experimental prototype; and (4) the integration of advanced infrastructure services.

Since it is not possible to assign all these tasks to specific

participants right from the start, CLARIN-NL has been set up as a mixture between a programme and a project.
CLARIN-NL also contains a range of activities that aim to further strengthen the leading position the Netherlands currently has in CLARIN-EU (both the principal coordinator and the technical coordinator for infrastructure technology are based in the Netherlands). It has a separate line of activities aimed to position the Netherlands prominently in CLARIN-EU also beyond CLARIN-prep, and to extend its leading position further by initiating, in an early stage, projects with selected international partners to develop, in a multilingual setting, showcase demonstrators of the infrastructure and the services it offers, as well as by setting up at least two centres of expertise

The CLARIN-NL proposal covers a period of 6 years, partitioned in three phases of two years: the preparation phase, the construction phase, and the first two years of the exploitation phase. Though the infrastructure is primarily aimed at language and humanities researchers, it offers various opportunities for usage in other domains and by other users, both for commercial applications as well as for important developments in society.

CLARIN-NL has effectively started on April 1, 2009 and will run for six years (2009-2014). Its budget is 9.01 million euro. CLARIN-NL is the first and so far the only CLARIN-related national project that has been awarded money not only for the preparatory phase but also for the implementation and exploitation phases. I submit that these plans and experiences can serve as an example that other national CLARIN project proposals can profit from.

3. Participants and Organization

Currently CLARIN-NL has 23 participants from linguistics and the humanities more broadly.[1] It includes all universities with a humanities faculty or with expertise in language and/or speech technology, and several institutes from the Royal Netherlands Academy of Arts and Sciences. The institutes carry out research in linguistics construed broadly (10), language technology (6), speech technology (2), culture (2), lexicography (2), social history (4), and literature (1).[2] The project thus covers a large part of humanities research and some social sciences research, though some subdisciplines are as yet insufficiently represented. Libraries are represented as well (2), and 5 institutes are data centres. Of these, 4 have expressed the intention to become a CLARIN centre (type A or B),[3] and some others are considering this.

The governance structure is as follows. The executive board (EB[4]) prepares the policy, and after approval, implements it. It consists of 4 persons: the programme director, a representative of the humanities, a technical director, and an education and awareness director. The EB reports to the board, which is responsible for all decisions concerning the project. The board[5] consists of 8 persons selected from the participating institutes with extensive scientific and managerial experience. The board and the EB are advised by the National Advisory Panel (NAP),[6] which consists of 17 persons representing a broad spectrum of the humanities field, and by the International Advisory Panel (IAP[7]) with internationally renowned experts in the relevant fields (humanities, infrastructures), which is currently being composed.

4. Infrastructure Implementation

A number of subprojects for the implementation of the infrastructure and the set-up of the CLARIN centres in the Netherlands have been initiated.

Metadata project A subproject has started up to carry out first tests of the Component Metadata Infrastructure (CMDI[8]) as developed in CLARIN-prep (WP2 and WP5) against a representative sample of data residing in Dutch research organizations, and to make tools for working with CMDI. This project has been started up very early, because it will lead to concrete guidelines and recommendations for the participants in the demonstrator and data curation subprojects described in section 6, and the tools will provide a user friendly interface to CMDI.

Infrastructure implementation A project to realize large parts of the implementation of the technical infrastructure has been defined and initiated. This project will set up the various infrastructure services, an open archiving service, various small registries, a federation of centres, both nationally and internationally by joining the European federation being set up; for metadata it will build the CMDI infrastructure, set up a schema registry, experiment with profile matching, carry out maintenance of the ISOCAT data category registry[9] and add a registry of relations between data categories (relation registry RELCAT). It will coordinate and give guidance for work on web services; it will deal with wrapper and service bus specification and implementation, select work flow tools and carry out experiments with them. It will run training courses and advise others on all kinds of infrastructure related issues. It will run for three years and involves institutes that have explicitly expressed their intention to become a CLARIN data centre. It is led by MPI, Nijmegen.[10] By involving all candidate CLARIN centres in the Netherlands the knowledge and expertise in these matters that is currently concentrated at MPI can be distributed and replicated to (employees of) other centres, thus creating a more robust configuration and minimizing

[1] http://www.clarin.nl/node/7
[2] The numbers given add up to more than 23 since some institutes are active in multiple disciplines
[3] See the CLARIN short guide on centres and references given there: http://www.clarin.eu/files/centres-CLARIN-ShortGuide.pdf
[4] http://www.clarin.nl/node/22
[5] http://www.clarin.nl/node/15
[6] http://www.clarin.nl/node/16
[7] http://www.clarin.nl/node/52
[8] http://www.clarin.eu/files/metadata-CLARIN-ShortGuide.pdf
[9] http://www.isocat.org/
[10] http://www.mpi.nl

the dependency on a single partner's knowledge and expertise.

Search functionality The infrastructure implementation project is complemented by a concrete demonstration project (called Search&Develop) in which centralized metadata search and distributed content search functionality will be implemented, with the same partners, the current candidate CLARIN centres in the Netherlands (INL[11], MPI, Meertens[12], and DANS[13]). This project has just started and will also run for three years. It is an ideal project to achieve the required national cooperation between CLARIN centres and the resulting functionality can become a showcase on the European level, thus strengthening the position of the Netherlands in the European CLARIN endeavour.

5. User Survey and Base Line Measurement

CLARIN-NL aims to create an infrastructure for humanities researchers and, to a lesser extent, social sciences researchers that work with language. It is essential that the infrastructure offers functionality that is needed and desired by its intended users. In order to get a better picture of the functionality that is needed a user survey is being carried out.

Some humanities researchers already work with digital data and tools, but many of them do not. It will be very difficult if not impossible for the latter to specify their needs, since they have no experience with it and no good overview of what is currently technically feasible. Therefore, the user survey has been set up as a set of interactive interviews, in which the user surveyor gets into a dialogue with the researcher to get an overview of his/her research questions and suggest possible tools and data that might facilitate the research.

It is also important to determine to what extent digital data and tools are currently being used, and by whom, or, if they are not used, to find out causes for this. This can serve as a base line measurement and it will make it possible to compare the current situation with the situation in a later stage when the CLARIN infrastructure has been implemented and has been operational. In this way it will become possible to determine the impact of the CLARIN infrastructure on the humanities research. For this reason, such a base line measurement is being carried out in parallel with the user survey.

6. Data Curation and Demonstrator projects

A call has been launched in 2009 for data curation and demonstrator projects. 17 proposals were submitted, and 11 of them could be awarded funding.

The goal of a **data curation project** is to adapt an existing resource so that they are visible, uniquely referable and accessible via the web, and properly documented.

The goal of a **demonstrator project** is to create a documented web application starting from an existing tool or application that can be used as a demonstrator and function as a showcase of the type of functionality CLARIN will incorporate and support.

Important goals **common** to both types of projects are (1) apply standards and best practices[14] and make use of the suggested CLARIN architecture and agreements to understand their limitations and the requirements for extensions; and (2) establish requirements and desiderata for the CLARIN infrastructure.

In particular, all projects will have to make metadata for their resources in accordance with the CMDI, and contribute to semantic interoperability by mapping the data categories used to data categories in ISOCAT, or by creating new data categories in ISOCAT.

In this way, curated resources and a range of showcases of functionality CLARIN will become available, and at the same time evidence-based requirements and desiderata for the CLARIN infrastructure and supported standards and best practices will be obtained. This will make it possible to influence the final selection of standards and best practices that will be promoted by CLARIN. These data curation and demonstrator projects neatly complement the European CLARIN preparatory project, in which the budget for a work package on these issues was cut by the European Commission.

In addition, each project has to involve intended users of the CLARIN infrastructure, which contributes to bringing the communities of the humanities researchers and technology providers (infrastructure specialists and language and speech technology researchers) together in concrete projects in which they are cooperating. This is essential, since only in this way there can be guarantees that the infrastructure will provide the functionality that is actually needed by the researchers.

An overview of the awarded projects is provided here. For each project, it consists of the project acronym and title, the coordinating institute, a very short description of the project goal, and a very short description of the scientific impact of the project. More detailed description of the projects can be found on the CLARIN-NL website[15]

AAM-LR - Automatic Annotation of Multi-modal Language Resources (Radboud University Nijmegen)
The AAM-LR project aims at building a demonstrator of a web service that will help field researchers to annotate audio- and video-recordings. It will facilitate all field research in which digital audio or video recordings are being made.
Adelheid - A Distributed Lemmatizer for Historical Dutch (Radboud University Nijmegen). This project aims

[11] http://www.inl.nl
[12] http://www.meertens.knaw.nl
[13] http://www.dans.knaw.nl

[14] See http://www.clarin.eu/recommendations for a list of standards and best practices currently promoted by CLARIN.
[15] http://www.clarin.nl/node/70

at providing a web-application with which an end user can have historical Dutch text tokenized, lemmatized and part-of-speech tagged. It will facilitate research on historical texts in fields such as historical linguistics, literary and historical studies.

ADEPT - Assaying Differences via Edit-Distance of Pronunciation Transcriptions (University of Groningen). The goal of the project is to provide a web application capable of measuring the differences in sets of phonetic (or phonemic) transcriptions via edit distance. It is based on existing software, which, however, is too complex for many potential users. The tool facilitates research in areas such as phonetics, phonology, dialectology, comparative linguistics and second-language learning.

En Garde - Converting DUELME into LMF format (Utrecht University). The goal of the project is develop converters between the DUELME database of multiword expressions[16] and LMF format, and to create a curated DUELME resource fully compliant with standards supported by CLARIN. The DUELME database makes it possible to address a variety of research questions related to Dutch multi-word expressions in research areas such as computational linguistics and natural language processing, theoretical linguistics and psycholinguistics.

INTER-VIEWs - Curation of Interview Data (Radboud University Nijmegen). The INTER-VIEWS project will make a corpus of interview data available to the community of researchers in the humanities. This project will contribute to facilitating all research that makes use of interviews, and more specifically research into the Second World War.

MIMORE - Microcomparative Morphosyntax Research Tool (Meertens Institute). The demonstrator tool MIMORE will create a common search engine for the databases DynaSAND, DiDDD and MAND. The tool is especially relevant to research in variation linguistics. It will allow investigating theoretical questions concerning the language system and language variation, and the geographic distribution of (morpho-)syntactic variables.

Sign-LinC - Linking lexical databases and annotated corpora of signed languages (Radboud University Nijmegen). This project aims to link two independently evolved data sets for a signed language: the Corpus NGT [17]and the lexical database of the Dutch Sign Centre.[18] The linked databases and corpora will facilitate research in the area of sign language linguistics

TICCLops - Text-Induced Corpus Clean-up online processing system (Tilburg University). This demonstrator project will develop a tool that allows CLARIN users to submit their corpora for fully automatic spelling correction and normalization. The tool will facilitate all research that makes use of text corpora.

TDS Curator - A web-services architecture to curate the Typological Database System (Utrecht University). TDS Curator will make the Typological Database System (TDS) [19] into a sustainable service that conforms to CLARIN infrastructural requirements. The TDS facilitates typological linguistic research by providing integrated access to multiple independently developed typological databases through a common web interface

TQE - Transcription Quality Evaluation (Radboud University Nijmegen). The TQE project aims to create a completely automatic Transcription Quality Evaluation (TQE) tool. The tool will be useful for validating, obtaining, and selecting phone transcriptions, for detecting phone strings (e.g. words) with deviating pronunciation, and, in general, it can be usefully applied in all research - in various (sub-)fields of humanities and language and speech technology - in which audio and phonetic transcriptions are involved.

WFT-GTB (Fryske Akademy, Leeuwarden). This project carries out data curation of the *Wurdboek fan de Fryske Taal* database (WFT)[20] 'Dictionary of the Modern West Frisian language' database using the TEI encoding scheme and demonstration of the data in the *Geïntegreerde Taalbank* 'Integrated Language Bank' (GTB) dictionary web application.[21] The search layer for retrieval in scholarly dictionaries will also be made available as a web service. This resource and application will facilitate the study of formal, semantic and idiomatic aspects of Modern West Frisian as well as comparative studies thanks to the matching with entries from two Dutch dictionaries in the GTB.

Finally, one project has a somewhat special origin:

CKCC project The CKCC project (*Circulation of Knowledge and Learned Practices in the 17th-century Dutch Republic*, 'Geleerdenbrievenproject') is an independently financed (NWO) project coordinated by the Descartes Institute (Utrecht University) that investigates, on the basis of a corpus of 20,000 letters of scientists from the 17th century in the Dutch Republic and using language technology, the research question of how knowledge circulated in the 17th century. It was selected in the CLARIN-EU call for humanities and social sciences projects as the project proposal that "[would] best demonstrate the use of LRT and would show the potential of a research infrastructure in the humanities" (CLARIN Newsletter 6, p. 3).[22] This project has been assigned additional funding to extend its use of language technology and to make adaptations to apply CLARIN-recommended standards and best practices.

7. Cooperation between the Netherlands and Flanders

A project for cooperation between the Netherlands and Flanders in CLARIN has started. It has always been part of the CLARIN-NL plan to include cooperation with Flanders, especially concerning data and tools specific to the Dutch language (which is shared by the Netherlands and Flanders).[23] The cooperation project consists of multiple aspects, but the core is a project in which existing

[16] Grégoire (2009) and http://www.inl.nl/nl/lexica/duelme

[17] http://www.ru.nl/sign-lang/projects/corpus_ngt/

[18] http://www.gebarencentrum.nl/

[19] http://languagelink.let.uu.nl/tds/main.html

[20] http://www.fa.knaw.nl/fa/3fakgroepen-en-dissiplinen/fakgroep-taalkunde/leksikografy-terminology/wurdboek-fan-e-fryske-taal

[21] http://www.inl.nl/nl/lopende-projecten/geintegreerde-taalbank?task=view

[22] See the CLARIN Newsletter 6: http://www.clarin.eu/files/cnl06_web_0.pdf

[23] See (Odijk 2008:14),

tools for Dutch, mainly developed in the STEVIN[24] programme, are adapted to become web services that can be used in a work flow system. This will be done for two modalities: text and speech. The users involved in this project are researchers from literary studies and from archaeology on the text side, and social history researchers on the speech side. Offering web services in a work flow will become one of the most important functionalities that the CLARIN infrastructure will offer, and several research groups are working on such systems (e.g. Germany[25] and Spain[26]). In the CLARIN-NL project it was the intention from the start to contribute to these developments as well. The tools developed in the STEVIN programme are natural candidates for such web services, and there is added value in the cooperation with Flanders since the tools have originally been developed together and concern the shared Dutch language. This project will last approximately 3 years and involves a broad spectrum of partners (10 from the Netherlands and 7 from Flanders).

8. Education, Training and Awareness

The CLARIN-NL project includes a plan for creating awareness and for education and training. Several activities in these areas have already been undertaken and more are in the pipeline. Such activities include:

Attending relevant events such as workshops and conferences (11 in 2009) and giving presentations there (more than 25 in 2009).

Organizing events CLARIN-NL organizes many events itself, e.g. in the past period the kick-off meeting and first open call announcement, an information session on the first call, and a meeting in which all subprojects presented themselves.[27]

Having meetings with representatives of related projects, candidate partners, candidate CLARIN centres, ministries, funding agencies, etc.

Supporting events that are organized by others that are relevant to CLARIN-NL in order to increase CLARIN-NL's visibility.

Support for individual researchers so that they can be present at and contribute to CLARIN-relevant events

Organizing tutorials and lectures on topics that are central to CLARIN and that are not generally known in the research community. At this moment three tutorials are planned, one on ISOCAT (March 25), one on CMDI metadata (end of May), and one on persistent identifiers. CLARIN-NL also contribute to summer and winter schools such as the LOT winter school (Jan 2010, Amsterdam)[28] and the CLARA summer schools.[29]

Setting up the website for the project which contains all information on the project aimed at the research community at large, and which includes special facilities for registered researchers.[30] It also includes Web2.0 functionality such as a forum for discussion and (restricted) sections for the governance bodies and work spaces for individual subprojects.

Other forms of dissemination such as regularly sending out newsflashes to researchers registered at the website[31] and publications and announcements in existing magazines aimed at the targeted research community and via relevant mailing lists.

9. Relation with CLARIN Europe

The CLARIN preparatory project is preparing the set-up of a European Research Infrastructure Consortium (ERIC), a European legal entity specifically created for research infrastructures. The Dutch minister of Education, Culture and Sciences, Ronald Plasterk, has announced that the Netherlands is prepared to host the CLARIN ERIC, and he has invited his colleague ministers in the other European countries earlier this year to join this ERIC. The Netherlands thus plays a leading role in this process, and within the CLARIN-NL project special budget has been reserved to consolidate this leading position. At this moment it is still too early to judge how these matters will develop, but CLARIN-NL is ready to support the CLARIN ERIC.

The CLARIN-NL project is complementary to the European CLARIN preparatory project in several respects: First, CLARIN-NL does not only cover the preparatory phase, but also the implementation phase and part of the exploitation phase of the CLARIN infrastructure. Second, CLARIN-NL carries out activities in the preparatory phase, such as the metadata project, the data curation and demonstrator projects, for which no funds are available in the European preparatory project. Other national CLARIN projects could do similar things with a focus on tools and resources available in their countries.

10. Future

The CLARIN-NL EB is currently analyzing the situation in the Netherlands. It attempts to identify gaps in topics and disciplines covered that are as yet underrepresented, as well as essential infrastructure functionality that is not covered yet. It will – based on this analysis, and the results of the user survey, and after consultation with the NAP and IAP - work out a proposal for a new call for subprojects with clearly identified priorities. It will also determine the best organisation of this call, both in terms of character (open call, tender, direct assignments), budget, and timing. It is expected that the plans for a new call (or even multiple calls) will be available in June 2010.

[24] http://taalunieversum.org/taal/technologie/stevin/
[25] http://weblicht.sfs.uni-tuebingen.de/Weblicht.pdf
[26] http://gilmere.upf.edu/WS/
[27] See http://www.clarin.nl/node/10 for more details on these events
[28] http://www.lotschool.nl/link.php?url=http://www.lotschool.nl/files/schools/2010%20winterschool%20Amsterdam/
[29] http://www.mpi.nl/research/research-projects/language-archiving-technology/events/clara-summerschool
[30] http://www.clarin.nl.
[31] http://www.clarin.nl/node/82

The CLARIN-NL EB is also developing a plan for supporting centres of expertise in the Netherlands. Centres of expertise are physical or virtual centres that possess a specific type of knowledge and expertise on a topic that is relevant to CLARIN and that have sufficient mass to guarantee the sustainability of this knowledge and expertise. The CLARIN-NL EB is identifying candidates for such centres of expertise and will propose priorities for supporting them. This plan is also expected in June 2010.
CLARIN-NL will continue to undertake all kinds of activities in the areas of creating awareness, education and training. For some targeted researchers (e.g. linguists), one can observe a gradual shift in emphasis from creating awareness to education and training, whereas for other researchers (e.g. in literature and history) creation of awareness is still the primary concern, and it will be a challenge to get these research communities more involved than they currently are. The user survey, which will be accompanied by events to create more awareness of CLARIN-NL and its potential benefits, will contribute to this as well. CLARIN-NL will also start working on activities to get the CLARIN infrastructure and working in a CLARIN-compatible manner into the regular curricula of universities.

CLARIN-NL aims to create an infrastructure that is supposed to have a long term existence. However, CLARIN-NL is just a project, which will finish in 2014. It is therefore of utmost importance to already start working on embedding the CLARIN infrastructure in the normal research activities and preparing both a governance structure and structural financing for the CLARIN infrastructure to ensure the longer term existence of the CLARIN infrastructure. Of course, for this reason it is also closely tracking the developments in Europe, especially with regard to the CLARIN ERIC, and assessing how to best play a role in this organisation.

11. Conclusions

I submit that the CLARIN-NL project has a number of characteristics that make it an excellent example for other national CLARIN projects. I summarize these characteristics here:

First, the structure of the project as a mix between a programme and a project provides both the flexibility to adapt the contents to new developments, and to new players in the field (e.g. humanities researchers not reached yet), and at the same time it offers opportunities for defining a few longer term projects in selected areas so that knowledge and expertise built up will be sustained in the participating institutes.

Second, the data curation and demonstrator projects offer opportunities to seriously test the standards and best practices promoted by CLARIN, e.g. various proposed data formats, but also the newly developed CMDI metadata framework, interoperability via the ISOCAT data registry, etc. The results of these projects will strengthen these standards and best practices, and provide evidence-based arguments for modifications or extensions; more generally, these projects will yield evidence-based requirements and desiderata for the CLARIN infrastructure. This is the best way to ensure possibilities for influencing a selection of standards and best practices in CLARIN that is compatible with the existing national data. In addition the projects will yield curated data, and, via the demonstrators, a range of showcases that can be used to explain and demonstrate the advantages of the CLARIN infrastructure and the new possibilities it will offer to researchers.

Third, the project requires cooperation between the intended users and the technology providers (infrastructure specialists and language and speech technology providers), with a central role for the users' research questions and thus contributes to bringing these different communities together in concrete cooperation projects.

12. Acknowledgements

This work was funded by the NWO CLARIN-NL project (http:///www.clarin.nl).

13. References

Grégoire, N. (2009), 'DuELME: A Dutch Electronic Lexicon of Multiword Expressions', *Journal of Language Resources and Evaluation*, special issue on Multiword Expressions. (as yet only available on-line: http://www.springerlink.com/content/7308605442w17698/fulltext.pdf)

Odijk, J., (2008). 'CLARIN-NL Proposal', UiL-OTS, Utrecht. http://www.clarin.nl/system/files/clarin-nl.pdf