

Background

Dialogue corpus annotation is fundamental for many different purposes in language engineering, such as in:

- training speech recognisers;
- constructing lexicons and grammars for spoken language dialogue systems;
- extracting dialogue control structures;
- comparing and evaluating annotation schemes with humans and/or machines.

Moreover, the increasing variety and sophistication of language engineering products is accompanied by needs for novel, standardised annotation schemes and tools.

A major problem in dialogue annotation is the lack of annotation standards which makes it hard for, e.g., spoken language dialogue system developers to re-use corpora and tools for handling annotated corpora in different development or customisation projects. The usual procedure in language engineering projects so far has been either to develop the needed resources from scratch or to acquire resources from previous projects and painstakingly adapt them to novel purposes.

Objectives

The MATE consortium will draw on results from projects world-wide on spoken dialogue annotation and tools. MATE will review these results and use them as background for proposing a standard for the annotation of spoken dialogue corpora, covering the levels of prosody, (morpho-)syntax, co-reference, dialogue acts, and communication aspects, with particular emphasis on their cross-level interaction. In parallel, MATE will build a workbench, i.e. a set of integrated tools which will support the use of the standard and enable annotation, different presentation formats, information extraction, statistical analysis, and mapping between different formats.

Expected results

Future success of applying language technologies depends on the production of language resources and their continuous development and re-use. The standard to be proposed by MATE should enable the re-use of spoken dialogue corpora. The MATE workbench should provide the means to produce and exploit corpora more efficiently and with greater accuracy and consistency. In turn, these benefits will make it more cost-effective to produce products for the market place, particularly spoken language dialogue systems but also other language engineering products.

The annotation standard proposed by MATE will:

- allow multiple annotation levels and cross-level annotation;
- allow coexistence of a variety of coding schemes and standards;
- allow multilinguality;
- integrate standardisation efforts in the US, Europe and Japan; and
- be open-ended with respect to information levels and within-level categories.

The key aspects of the workbench are:

- acquisition and manual/semi-automatic annotation/modification of data using the MATE standard;
- presentation/visualisation of spoken dialogue corpora and annotations at different levels, according to user-defined partial views;
- extraction/retrieval from annotated corpora according to any combination of constraints from both the dialogue text and any type of annotation;
- statistical procedures to determine inter-coder consistency; frequency of object language phenomena, etc.
- import/export of annotated material and easy integration of results of external (possibly automatic) tools.

MATE needs your collaboration

Focusing on standardisation for efficiency, MATE needs and welcomes the widest possible collaboration in order to maximise the relevance and usability of its results. You are cordially invited to join the MATE Advisory Panel if you are interested in collaborating on standards and tools for handling annotated spoken dialogue corpora. Your contribution could include commenting on intermediate results, providing access to spoken dialogue corpora whether annotated or not, pointing to existing coding schemes and corpus handling tools, and listing needs you have experienced in your own work as regards corpus annotation and tools. Advisory Panel Members will have access to the MATE workshops, early access to MATE documents, and the opportunity to test drive the MATE workbench early on.

How to join the Advisory Panel

You can choose any of the following options: (a) fill in and return the form attached to this leaflet; (b) complete the Advisory Panel form at <http://grindsted.mip.ou.dk/mate/>; or (c) send an email to mate@mip.ou.dk containing the same information.

More information about MATE

More detailed, continuously updated information on MATE, its progress, results and needs for interaction is available at <http://grindsted.mip.ou.dk/mate/>

.....

I would like to join the MATE Advisory Panel

Please complete and return this form to Odense University by mail (or by fax +45 63 15 72 24)

Name of contact person

Department

Organisation

Address

Postal Code/City

Country

Telephone

Fax

Email

URL

DateSignature.....

MATE partners and people

- The Maersk Mc-Kinney Moller Institute for Production Technology (MIP) (coordinating partner)
Odense University, Science Park 10
DK-5230 Odense M, Denmark
Phone: +45 65 57 35 53
Fax: +45 63 15 72 24
Laila Dybkjær, contact person (laila@mip.ou.dk)
Niels Ole Bernsen (nob@mip.ou.dk)
Henrik Kasch (kasch@mip.ou.dk)
Morten Baun Møller (baun@mip.ou.dk)

.....

The Maersk Institute
Odense University
Forskerparken 10
5230 Odense M
Denmark

Centro Studi e Laboratori Telecomunicazioni S.p.A. (CSELT)
Via Reiss Romoli 274
I-10148 Torino, Italy
Phone: +39 11 22 86 247
Fax: +39 11 22 86 207
Morena Danieli, contact person (morena.danieli@cse.lt.it)
Silvia Quazza (silvia.quazza@cse.lt.it)

- Departament de Filologia Espanyola (DFE)
Universitat Autònoma de Barcelona, Bellaterra,
E-08193 Barcelona, Spain
Phone: +34 93 581 12 16
Fax: +34 93 581 16 86
Joaquim Llisterra, contact person (joaquim.llisterra@cc.uab.es)
Juanma Garrido (juanma@liceu.uab.es)
- Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI)
Stuhlsatzenhausweg 3
D-66123 Saarbrücken, Germany
Phone: +49 681 302 53 46
Fax: +49 681 302 53 41
Norbert Reithinger, contact person (norbert.reithinger@dfki.de)
Marion Klein (mklein@dfki.de)
- Human Communication Research Centre (HCRC)
2, Buccleuch Place
Edinburgh EH8 9LW, UK
Phone: +44 131 650 4665
Fax: +44 131 650 4587
Marc Moens, contact person (marc@cogsci.ed.ac.uk)
Jean Carletta (j.carletta@edinburgh.ac.uk)
Amy Isard (amyi@cogsci.ed.ac.uk)
David McKelvie (dmck@cogsci.ed.ac.uk)

Massimo Poesio (poesio@cogsci.ed.ac.uk)

- Institut für Maschinelle Sprachverarbeitung (IMS)
Universität Stuttgart
Azenbergstraße 12
D-70174 Stuttgart, Germany
Phone: +49 711 121 1373
Fax: +49 711 121 1366
Ulrich Heid, contact person (heid@ims.uni-stuttgart.de)
Andreas Mengel (mengel@ims.uni-stuttgart.de)
- Istituto Linguistica Computazionale (ILC)
Via della Faggiola 32
I-56126 Pisa, Italy
Phone: +39 50 560481
Fax: +39 50 556285
Vito Pirrelli, contact person (vito@ilc.pi.cnr.it)
Claudia Soria (soria@ling.unipi.it)
- Telefónica Investigación y Desarrollo (TID)
Emilio Vargas, 6
E-28043 Madrid, Spain
Phone: +34 91 337-42-52
Fax: +34 91 337-42-02
Daniel Tapias, contact person (daniel@craso.tid.es)

Multilevel
Annotation
Tools
Engineering

MATE

Multilevel Annotation, Tools Engineering

LE Telematics Project LE4-8370

1 March 1998 - 31 October 1999

