

# The NITE XML Toolkit: Demonstration from five corpora

Jonathan Kilgour and Jean Carletta

University of Edinburgh

United Kingdom

jonathan@inf.ed.ac.uk

## Abstract

The NITE XML Toolkit (NXT) is open source software for working with multi-modal, spoken, or text language corpora. It is specifically designed to support the tasks of human annotators and analysts of heavily cross-annotated data sets, and has been used successfully on a range of projects with varying needs. In this text to accompany a demonstration, we describe NXT along with four uses on different corpora that together span its most novel features. The examples involve the AMI and ICSI Meeting Corpora; a study of multimodal reference; a syntactic analysis of Genesis in classical Hebrew; and discourse annotation of Switchboard dialogues.

## 1 Introduction

Of the software packages that provide support for working with language corpora, the NITE XML Toolkit (NXT) is arguably the most mature, offering to combine multiple audio and video signals with crossing structures of linguistic annotation. It is currently in use on a range of corpora. Current users both create annotations in NXT natively and “up-translate” existing data from other tools into NXT’s storage format. Although its biggest strengths are for multimodal language research, some users have found it to be the right solution for work on text and speech corpora without video because of the way it handles annotation and analysis. NXT is open source software, available from Sourceforge and documented at <http://www.ltg.ed.ac.uk/NITE>. It is written in Java and uses the Java Media Framework (JMF) for its handling of signals.

In this text to accompany a demonstration, we summarize NXT’s functionality, comment on its use for four corpora that together showcase its most novel features, and describe funded future development.

## 2 The NITE XML Toolkit

At its core, NXT consists of three libraries: one for data handling, one for searching data, and one for building GUIs for working with data. The data handling libraries include support for loading, serialization using a stand-off XML format, navigation around loaded data, and changes to the data in line with a specific data model that is intended especially for data sets that contain both timing information and overlapping structural markup. The search facility implements a query language that is designed particularly for the data model and allows the user to find n-tuples of data objects that match a set of conditions based on types, temporal conditions, and structural relationships in the data. The GUI library defines signal players and data displays that update against loaded data and can highlight parts of the display that correspond to current time on the signals or that correspond to matches to a query typed into a standard search interface. This includes support for selection as required for building annotation tools and a specific transcription-oriented display.

NXT also contains a number of end user interfaces and utilities built on top of these libraries. These include a generic display that will work for any NXT format data, configurable GUIs for some common hand-annotation tasks such as markup of named entities, dialogue acts, and a tool for segmenting and labelling a signal as it plays. They also include command line utilities for common search tasks such as counting query results and

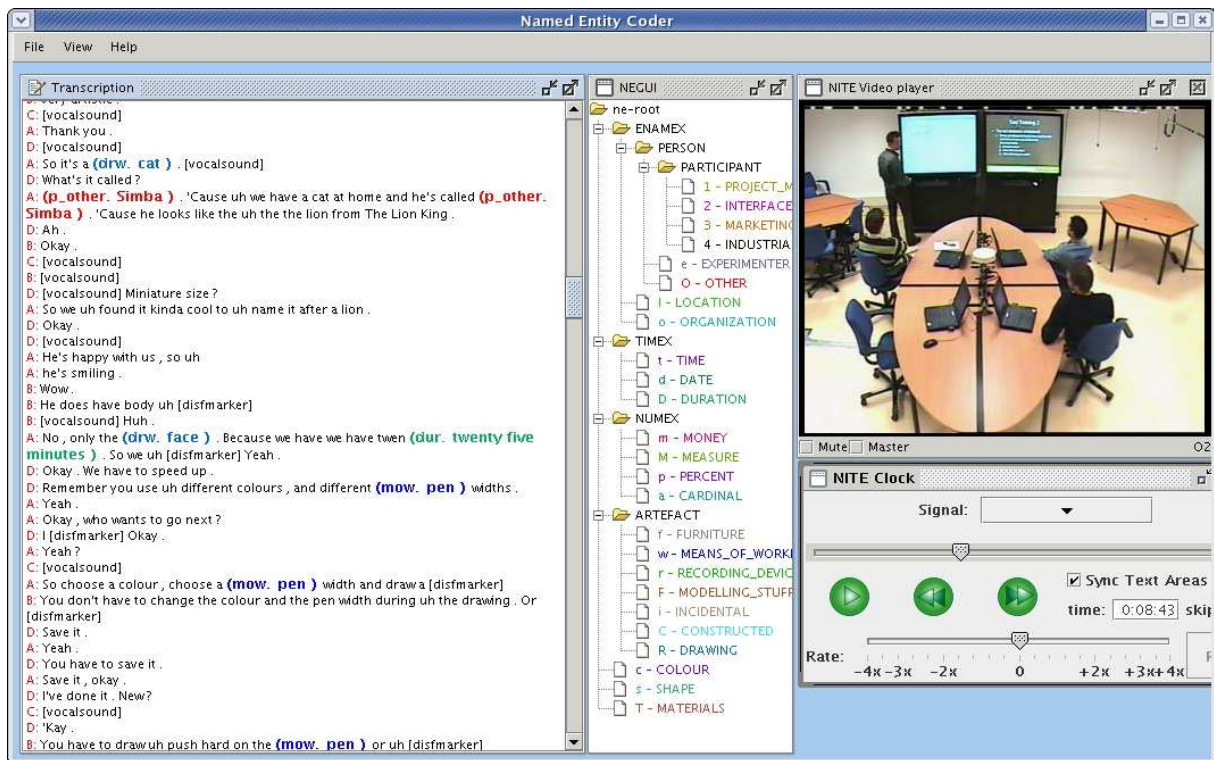


Figure 1: Named entity coding on the AMI corpus

some utilities for transforming data into, for instance, tab-delimited tables.

Finally, a number of projects have contributed sample data and annotation tools as well as mechanisms for transforming data to and from other formats. Writing and testing a new up-translation typically takes someone who understands NXT's format between one and three days. The actual time depends on the complexity of the structure represented in the input data and whether a parser for the data format must be written from scratch. Badly documented formats and ill-formed data take longer to transform.

### 3 Examples

#### Example 1: The AMI and ICSI Meeting Corpora

The AMI Project (<http://www.amiproject.org>) is currently NXT's biggest user, and is also its largest provider of financial support. AMI, which is collecting and transcribing 100 hours of meeting data (Carletta et al., 2005) and annotating part or all of it for a dozen different phenomena, is using NXT's data storage for its reference format, with data being generated natively using NXT GUIs as well as up-translated from other sources. The project uses ChannelTrans (ICSI, nd) for orthographic

transcription and Event Editor (Sumec, nd) for straightforward timestamped labelling of video; although NXT comes with an interface for the latter, Event Editor, which is Windows-only and not based on JMF, has better video control and was already familiar to some of the annotators. For annotation, AMI is using the configurable dialogue act and named entity tools as well as tailored GUIs for topic segmentation and extractive summarization that links extracted dialogue acts to the sentences of an abstractive summary they support. Figure 1 shows the named entity annotation tool as configured for the AMI project. Aside from the sheer scale of the exercise, the AMI effort is unique in requiring simultaneous annotation of different levels at different sites. NXT does not support data management, but its stand-off XML data format has made it relatively easy to manage the process using a combination of a CVS repository for version control, web forms for data upload, and wikis for work assignment and progress reports.

The AMI Project piloted many of their techniques on the ICSI Meeting Corpus (Janin et al., 2003), which shares some characteristics with the AMI corpus but is audio-only. More information about this closely related use of NXT can be found in (Carletta and Kilgour, 2005).

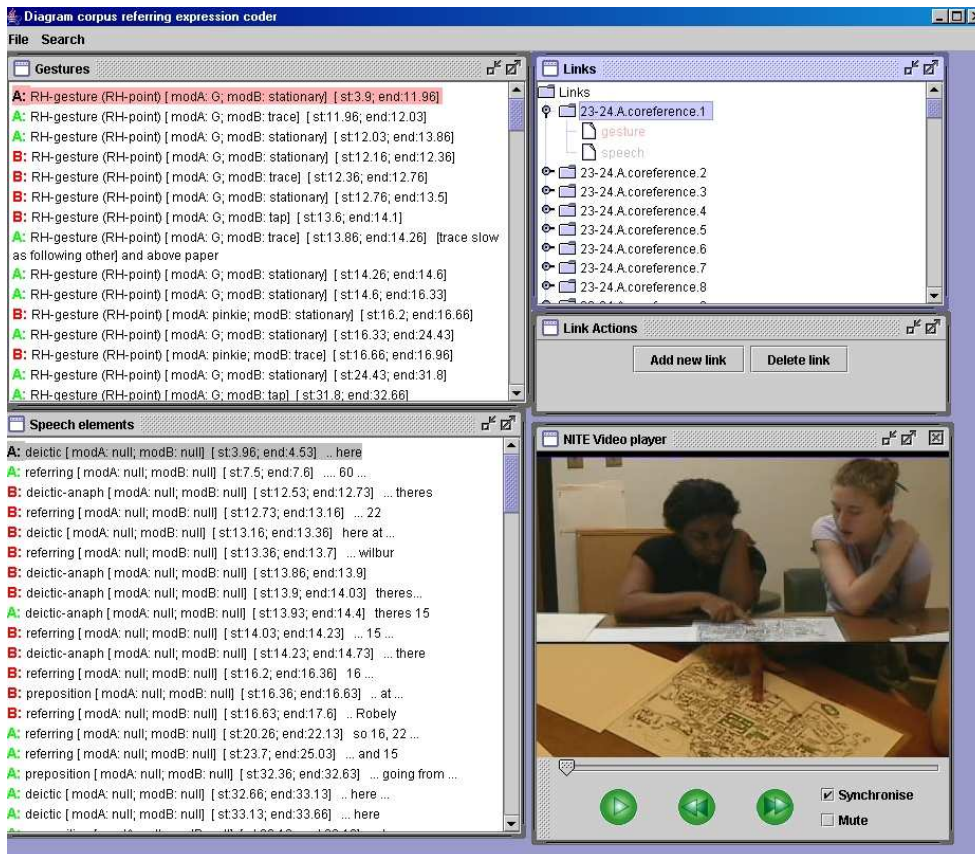


Figure 2: DIAGRAMS corpus: linking gesture and dialogue acts.

### Example 2: Multimodal reference

This example is a small project that is looking at the relationship between referring expressions and the hand gestures used to point at a map. Although the transcription, referring expression, and gesture annotations were done in other tools and then up-translated, NXT gave the best support for linking referring expressions with gestures and analysing the results. Figure 2 shows the linking tool. One interesting aspect of this project was that the analysis was performed by a postgraduate psychologist. Analysts with no computational experience find it more difficult to learn how to use the query language, but several have done so. With this kind of data set, simply the ability to play the signals and annotations together and highlight query results provides insights into behaviours that are difficult to reach otherwise.

### Example 3: Syntax in Genesis

This example is an annotation of Genesis in classical Hebrew that shows its structural division into books, chapters, verses, and half-verses. The data itself, which is purely textual, was originally

stored in an MS Access relational database, but overlapping hierarchies in the structure made it difficult to query in this format. After finding NXT on the web and consulting us about the best way to represent the data using the NXT data model, the user successfully up-translated his data, searched it using NQL, and exported counts to SPSS to create corpus statistics.

### Example 4: The Switchboard Corpus

The Switchboard Dialogue Corpus (Godfrey et al., 1992) has been popular for computational discourse research. (Carletta et al., 2004) describes an effort which up-translated its Penn Treebank syntactic analysis to NXT format, added annotations of “markables” for animacy, information structure, and coreference, and used this information all together. This project made heavy use of NXT’s query language, including the ability to index query results in the data storage format itself for easy access. The work is now being extended to align an improved version of the transcriptions that includes word timestamps derived by forced alignment with the transcriptions used for the syntactic and discourse

annotation, and to add annotations for phonology and syllable structure, all within the same corpus structure.

## 4 Discussion

It should not be supposed from our list of examples that NXT has been used only for these applications. Particularly novel NXT uses include simultaneous display of annotation with a re-enactment of a human-computer tutorial dialogue driven by the dialogue system itself; hand-annotation of head orientation from video using a flock-of-birds motion sensor mounted on a coffee mug; and annotation of the critiques expressed in conversations about movies. However, most NXT users are applying some kind of discourse annotation. They choose NXT because they need to combine signal labellings with annotations that give structure over the top of orthography, because they want to combine annotations from different sources and so find the stand-off format attractive, or because they need the GUI library in order to develop novel interfaces. Academic software is often inadequately documented, and therefore only usable with the help of its developers. It is inevitable given the size of the target user community that most of them are at least “friends of friends”. Enough users have worked independently of the developers that we are confident that all but the newest parts of NXT are understandable from the documentation.

Although NXT is mature enough for use, several projects are investing in further development. The largest current efforts are to create an annotation and analysis tool with a time-aligned “tiered” data display and a query processor with better performance (Mayo et al., 2006). Another priority is better packaging, particularly of the configurable interfaces and of the existing translations from the formats used in transcription and other annotation tools. Finally, contributing projects plan work that will improve interoperability between NXT and other tools, including eyetrackers, NLP applications such as part-of-speech taggers, and machine learning software.

## Acknowledgments

NXT prototyping and development has been supported primarily by the European Commission (MATE, LE4-8370; NITE, IST-2000-2609; and AMI, FP6-506811) and Scottish Enterprise via the

Edinburgh-Stanford Link. The examples shown are by kind permission of Scottish Enterprise, the AMI Project, and Dr. Matthew Anstey of Charles Sturt University.

## References

- [Carletta and Kilgour2005] J.C. Carletta and J. Kilgour. 2005. The NITE XML Toolkit meets the ICSI Meeting Corpus: Import, annotation, and browsing. In S. Bengio and H. Bourlard, editors, *Machine Learning for Multimodal Interaction: First International Workshop, MLMI 2004, Martigny, Switzerland, June 21-23, 2004, Revised Selected Papers*, Lecture Notes in Computer Science 3361, pages 111–121. Springer-Verlag, Berlin.
- [Carletta et al.2004] Jean Carletta, Shipra Dingare, Malvina Nissim, and Tatiana Nikitina. 2004. Using the NITE XML Toolkit on the switchboard corpus to study syntactic choice: a case study. In *Fourth Language Resources and Evaluation Conference*, Lisbon, Portugal.
- [Carletta et al.2005] J. Carletta, Simone Ashby, Sebastien Bourban, Mike Flynn, Mael Guillemot, Thomas Hain, Jaroslav Kadlec, Vasilis Karaiskos, Wessel Kraaij, Melissa Kronenthal, Guillaume Lathoud, Mike Lincoln, Agnes Lisowska, Iain McCowan, Wilfried M. Post, Dennis Reidsma, and Pierre Wellner. 2005. The AMI Meeting Corpus: A pre-announcement. In S. Renals and S. Bengio, editors, *2nd Joint Workshop on Multimodal Interaction and Related Machine Learning Algorithms*, Edinburgh.
- [Godfrey et al.1992] J. Godfrey, E. Holliman, and J. McDaniel. 1992. Switchboard: Telephone speech corpus for research development. In *International Conference on Acoustics, Speech, and Signal Processing*, pages 517–520, San Francisco, CA, USA.
- [ICSInd] ICSI. n.d. Extensions to Transcriber for meeting recorder transcription. <http://www.icsi.berkeley.edu/Speech/mr/channeltrans.html>; accessed 14 Oct 05.
- [Janin et al.2003] A. Janin, D. Baron, J. Edwards, D. Ellis, D. Gelbart, N. Morgan, B. Peskin, T. Pfau, E. Shriberg, A. Stolcke, and C. Wooters. 2003. The ICSI Meeting Corpus. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, Hong Kong.
- [Mayo et al.2006] Neil Mayo, Jonathan Kilgour, and Jean Carletta. 2006. Towards an alternative implementation of NXT’s query language via XQuery. In *EACL Workshop on Multi-dimensional Markup in Natural Language Processing (NLPXML)*, Trento, Italy.
- [Sumecond] Stanislav Sumec. n.d. Event Editor. <http://www.fit.vutbr.cz/research/grants/m4/editor/>; accessed 14 Oct 05.