Mouth-based non-manual coding schema used in the Auslan corpus: explanation, application and preliminary results

Trevor Johnston, Jane van Roekel
Macquarie University
Sydney, Australia
E-mail: trevor.johnston@mq.edu.au, jane.vanroekel@mq.edu.au

Abstract

We describe a corpus-based study of one type of non-manual in signed languages (SLs) — mouth actions. Our ultimate aim is to examine the distribution and characteristics of mouth actions in Auslan (Australian Sign Language) to gauge the degree of language-specific conventionalization of these forms. We divide mouth gestures into categories broadly based on Crasborn et al. (2008), but modified to accommodate our experiences with the Auslan data. All signs and all mouth actions are examined and the state of the mouth in each sign is assigned to one of three broad categories: (i) mouthings, (ii) mouth gestures, and (iii) no mouth action. Mouth actions that invariably occur while communicating in SLs have posed a number of questions for linguists: which are ‘merely borrowings’ from the relevant ambient spoken language (SpL)? which are gestural and shared with all of the members of the wider community in which signers find themselves? and which are conventionalized aspects of the grammar of some or all SLs? We believe these schema captures all the relevant information about mouth forms and their use and meaning in context to enable us to describe their function and degree of conventionality.

Keywords: sign language, corpus, ELAN, non-manuals, Auslan, Australian Sign Language

1. Introduction

The mouth is prominent site of non-manual activity and movements of the mouth are an obvious accompaniment to manual signing. The linguistic status of mouth actions in SLs, like other non-manuals, is a question of debate. There are two major types of mouth actions: those that are transparently complete or partial silent articulations of the spoken words of the ambient SpL (mouthings), and those that are not (mouth gestures). An early issue of interest was the amount of mouth actions, especially mouthings, that various SLs typically manifested. An area of debate concerned the status of mouthings: were they an integral part of a SL or were they marginal, stemming from language contact or borrowing? Addressing both these questions involved describing what mouthing and mouth gestures did and thus categorizing them into types.

2. Previous research

Research has shown that mouthings frequently accompany manual signs in many SLs and there is some evidence—though the datasets have never been very large or varied—that the rate varies according to text-type. They occur very frequently with fingerspellings and some signers appear to always mouth when fingerspelling (e.g., Sutton-Spence & Day 2001). Mouthings have been shown to occur more with nouns and plain verbs than with morphologically complex signs such as indicating verbs (also known as agreement and spatial verbs) or with depicting signs (also known as classifier signs). Mouthing has been shown to add meaning to some signs by indicating a more specific reading of a sign, e.g., the Auslan sign spouse with the English mouthings ‘wife’ or ‘husband’ (Johnston & Schembri 2007) which may or may not be considered a specification of the form of the sign (e.g., Schermer 2001). A mouthing can even add independent semantic inform
the conventional spoken words of a SpL (Pizzuto 2003). Fontana (2008), drawing on the work of Kendon (2004, 2008) and McNeill (2000), makes the radical suggestion that all mouth actions can be analysed this way. A similar but more conservative observation is made in Dachkovsky and Sandler (2009) and Sandler (2009). They identify a category of gestural iconic mouth actions to distinguish them from syllabic E-type (or ‘lexical’) mouth components and from the conventional adverbial and adjectival A-type modifiers already identified by other SL researchers which they also accept (graphically represented in Figure 1).

Figure 1 Three potential categorizations of mouth actions

3. This study

3.1 Methodology

Fifty video texts were selected from the Auslan corpus for analysis. All signs and all mouth actions were examined and all mouthings and mouth gestures were identified, categorized and annotated.

The data in this study has been drawn from the Auslan corpus of native or near-native signers (for further details see Johnston & Schembri 2006). For this study, 50 video clips were selected from the corpus, representing 38 individuals, 3 text types (monologue, dialogue, and elicited) during 5 hours and 58 minutes of the corpus, representing 16,920 manual sign tokens. The signed texts ranged from 1:32 to 38:30 minutes in duration. The 50 video clips consisted of 25 monologues (narratives of which there were 25 retellings of two Aesop’s fables); 10 dialogic texts (free conversation or responses to a series of interview questions); and 15 sessions of 40 elicited picture descriptions.

3.2 Annotation schema

The 50 texts were chosen from a subset of the 459 texts that had previously been given at minimum a basic annotation (i.e., they had been glossed and translated) using ELAN multi-media annotation software according to guidelines detailed in the Auslan Corpus Annotation Guidelines1 and Johnston (2010). All signs and all mouth actions were examined and the state of the mouth in each sign was assigned to one of three broad categories: (i) mouthings, (ii) mouth gestures (both of which we have already briefly characterized), and (iii) no mouth action. Mouth gestures were divided into types that were based on Crasborn et al. (2008) but additional sub-groupings were (temporarily) created to accommodate finer distinctions we felt salient in the Auslan data (Figure 2). We are prepared to further adapt or even abandon these categories if needs be after considering the first annotation implementation, aggregation of data, and analysis.

These new sub-categories are: prosodic = a tensed posture of the mouth that is held for a period of time, even if relatively briefly, without changing dynamically rather than any specific mouth posture as such; spontaneous = involuntary or spontaneous expressions (indexes almost) of the state of the mind of the signer (e.g., amused, confused, concerned); editorial = expressions as meta-comments about what the signer is signing that do not intentionally modify the manual signs; constructed actions = full enactments that involve all of the face; congruent = a default expression that match the semantics of the lexical sign, such as smiling while signing happy; adverbial expressive = clearly intend to modify and add meaning to the manual sign(s) but they are not limited to the mouth and they are also strongly enacting (Figure 2).

1 Downloadable from www.auslan.org.au
Annotations were added to two ID-glossing and two grammatical class tiers (one for each hand of the signer), and four tiers for information on mouth actions. Annotations for mouth gestures were made on the ‘mouth gesture form’ tier (called MouthGestF), and on the ‘mouth gesture meaning’ tier (called MouthGestM). The annotations for mouthings were made on the ‘mouthing form’ tier (called Mouthing) (Figure 3).

The relevant tiers [file: SSNc2a00:00:29.000]

<table>
<thead>
<tr>
<th>Mouth action</th>
<th>MouthGestF begins with</th>
<th>MouthGestM tier contains</th>
</tr>
</thead>
<tbody>
<tr>
<td>No mouth action</td>
<td>no annotation</td>
<td>no annotation</td>
</tr>
<tr>
<td>M-type (mouthing)</td>
<td>English word (Table 2)</td>
<td>no annotation</td>
</tr>
<tr>
<td>E-type (echo or empty)</td>
<td>syllable:gloss</td>
<td>various meanings as needed</td>
</tr>
<tr>
<td>A-type (modifying)</td>
<td>intonational gloss (Table 3)</td>
<td>meaning: e.g., activity, emphasis</td>
</tr>
<tr>
<td></td>
<td>adverbia gloss (Table 3)</td>
<td>meaning: e.g., large-amount, careless, unpleasant</td>
</tr>
<tr>
<td>4-type (mouth for mouth)</td>
<td>CongruentMouth Only</td>
<td>enactment</td>
</tr>
<tr>
<td>W-type (whole-of-face)</td>
<td>spontaneous no annotation</td>
<td>no annotation</td>
</tr>
<tr>
<td></td>
<td>editorial comment</td>
<td>no annotation or various meanings as needed</td>
</tr>
<tr>
<td></td>
<td>constructed action</td>
<td>no annotation or various meanings or descriptions as needed</td>
</tr>
<tr>
<td></td>
<td>ConstructedAction</td>
<td>the gloss for an A-type mouth gesture</td>
</tr>
<tr>
<td></td>
<td>ConstructedAction:gloss (Table 3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>congruent CongruentWholeFace</td>
<td>expression, enactment, emphasis</td>
</tr>
<tr>
<td></td>
<td>adverbial expressive ConstructedAction:adv</td>
<td>expression</td>
</tr>
</tbody>
</table>

Table 1 The annotation schema for mouth actions

The annotation schema is summarized in Table 1. It is important to note that each type of mouth gesture has a unique MouthGestF annotation or a unique combination of values on both the MouthGestF and MouthGestM tiers. This enables various constraints to be applied with an ELAN search routine to identify and quantify mouth gestures of a certain type or sub-type only. For example, one may perform an ELAN multi-tier conditional search to extract statistics on, say, intonational mouth gestures because they will be the only ones that have the annotations “activity” or “emphasis” on the MouthGestM tier associated with any of the form glosses listed in Table 2 on the MouthGestF tier. For example, if “emphasis” occurs on a MouthGestM tier which occurs with a “Con- gruentWholeFace” annotation on the MouthGestF tier, it is an instance of a congruent W-type mouth gesture, not
an intonational one. Each combination of values in Table 1 is unique to a sub-category of mouth gestures.

<table>
<thead>
<tr>
<th>Degree of articulation</th>
<th>Representation</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete articulation</td>
<td>complete</td>
<td>race, rabbit, village, far</td>
</tr>
<tr>
<td>Initial segment</td>
<td>(initial)</td>
<td>village, sa(me), diff(eren), sh(eep)</td>
</tr>
<tr>
<td>Medial segment</td>
<td>(me)di(al)</td>
<td>(no)th(ing), (re)mem(ber), (b)e(st)</td>
</tr>
<tr>
<td>Final segment</td>
<td>(fi)nal</td>
<td>(success)ful, (fin)ish, (im)prove, (to)day</td>
</tr>
<tr>
<td>Initial &amp; final segment only</td>
<td>in(i)tial</td>
<td>f(in)ish, d(ea)f, s(uc)cesful</td>
</tr>
<tr>
<td>Suppressed</td>
<td>Suppressed</td>
<td>(lady), (have)</td>
</tr>
<tr>
<td>Unreadable</td>
<td>Unreadable</td>
<td></td>
</tr>
</tbody>
</table>

Table 2 Examples of glosses for mouth gesture forms

Annotating the alignment of mouth actions with manual signs Where a mouth action clearly spreads across two or more signs, it is marked separately on each one and the subsequent annotations are suffixed with -prog (progressive) or --regress (regressive). Only apparently significant spreading is annotated in this dataset—the exact onset and offset time of mouth gestures is not a focus of this study though it has been in other studies (Sandler 1999; Crasborn et al. 2008)

Annotating mouthings Mouthings are often incomplete or partial. We found that it was important to annotate which part or parts of the associated English word were mouthed (Table 3). This was partly based on the realization that competent but non-fluent or non-native signers can easily mistaken a partial mouthing for a mouth gesture when annotating the data thus resulting in an overall inflation of mouth gesture counts.

3.3 Preliminary results

The results presented here are from the first iteration of the implementation of this annotation schema. They are not definitive. They are merely indicative of the type of information that can be easily extracted from the corpus given these annotations. The annotations were partly motivated by the type of functions available in ELAN for searching, filtering and exporting annotations within that program.

The key functions used in ELAN included: multi-file multi-tier searches using explicit values or regular expressions; automatically generated statistical profiles across multiple annotation files (domains); multi-file (domain) processing, in particular, EXPORT MULTIPLE FILES AS > ANNOTATION OVERLAPS INFORMATION. The latter were exported as tab delimited files into Excel for further processing.
Figure 4 Overall distribution of mouth actions

Figure 5 Mouth actions by sign type
Lexical frequency  The rates of mouthing for lexical frequency suggest that lexical frequency, as such, has marginal impact on mouthing rates. However, it did emerge that the highest ranking lexical signs have a smaller the range of English words mouthed with each sign (and this is unsurprisingly often the same word that has been adopted for the ID-gloss in the corpus).

3.3.1 Characteristics of types of mouth actions
E-type (semantically empty mouth gestures) Syllabic mouth gestures were very rare in the data. Only 66 signs
and their mouth actions fell into this category. Several signs occurred with different mouth gestures, often with the same effect (Table 4).

The first observation to make is that at 11 potential forms, the number of different syllabic mouth patterns in this dataset is actually quite small. A second important observation is that syllabic mouth gestures in Auslan do have flexible but consistent meanings across a range of signs, they are not exactly semantically empty as such as suggested in the SL literature.

<table>
<thead>
<tr>
<th>Mouth gesture form</th>
<th>Mouth gesture meaning</th>
<th>Tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAH</td>
<td>SUDDEN</td>
<td>31</td>
</tr>
<tr>
<td>AP</td>
<td>EMPHASIS</td>
<td>12</td>
</tr>
<tr>
<td>(L/B)AM</td>
<td>DISAPPEAR</td>
<td>8</td>
</tr>
<tr>
<td>WOOF</td>
<td>EMPHASIS</td>
<td>4</td>
</tr>
<tr>
<td>PAH-PAH</td>
<td>EMPHASIS</td>
<td>2</td>
</tr>
<tr>
<td>POOH</td>
<td>REMOVE</td>
<td>2</td>
</tr>
<tr>
<td>(L/B)AM</td>
<td>EMPHASIS</td>
<td>2</td>
</tr>
<tr>
<td>POW</td>
<td>EMPHASIS</td>
<td>2</td>
</tr>
<tr>
<td>BOOM</td>
<td>EMPHASIS</td>
<td>1</td>
</tr>
<tr>
<td>AM</td>
<td>EMPHASIS</td>
<td>1</td>
</tr>
<tr>
<td>ALARM</td>
<td>EMPHASIS</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>66</td>
</tr>
</tbody>
</table>

Table 4 E-types by meaning and token count

<table>
<thead>
<tr>
<th>MG form (tokens)</th>
<th>MG meaning gloss</th>
<th>Meanings in more detail</th>
</tr>
</thead>
<tbody>
<tr>
<td>TONGUE (111)</td>
<td>CARELESS</td>
<td>carelessly, easily, without regard, petulantly, with deliberate careless enjoyment, reckless, slipshod, insouciant</td>
</tr>
<tr>
<td>LIPS-OUT (16)</td>
<td>EASE</td>
<td>easily, without regard, petulantly, with enjoyment</td>
</tr>
<tr>
<td>TRILL (13)</td>
<td>LARGE AMOUNT</td>
<td>large amount, a lot of, unimpeded, energetic, powerful, engine/machine-powered</td>
</tr>
<tr>
<td>BLOW (10)</td>
<td>SMOOTH</td>
<td>smooth, unimpeded, quickly, ongoing</td>
</tr>
<tr>
<td>TONGUE (7)</td>
<td>UNPLEASANT</td>
<td>unpleasant, distasteful, bad</td>
</tr>
<tr>
<td>TRILL (3)</td>
<td>EASE</td>
<td>easily, unimpeded, with enjoyment</td>
</tr>
<tr>
<td>LIPS-PRESSED (2)</td>
<td>EASE</td>
<td>easily but deliberately, enjoyable</td>
</tr>
<tr>
<td>BOT-TOM-LIP-OUT (2)</td>
<td>CARELESS</td>
<td>carelessly, easily, without regard, petulantly, with deliberate careless enjoyment, reckless, slipshod, insouciant</td>
</tr>
<tr>
<td>Total (165)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5 Form/meaning pairings for A-type

A-type (adverbials) the majority are actually intonational in character, according to our definitions. In effect, only a very small number of the total number of signs in the dataset represent putative dedicated conventional A-type adverbial mouth gestures. Only a very small set of recurring semantic descriptors needed to capture the apparent contribution of these mouth gestures. The data suggests that the semantic component of the mouth gesture is quite broad (Table 5).

Only two broad meaning labels appeared necessary to capture the effect of intonational mouth gestures: emphasis and activity. Emphasis was the broader default reading (71%), and wide was the most preferred mouth gesture with this force (21%); activity accounted for the force of the remaining 29%, of which 48% were achieved with wide. Overall, wide accounted for almost a third of all intonational mouth gestures.

W-type (whole of face) 82% are of the constructed action sub-type. There are very few of the other types. Indeed, constructed actions represent 45% of all mouth gesture types (i.e., mouth actions excluding mouthings)

4-type (mouth for mouth) Token count was extremely low (N = 68). The token frequencies are unremarkably linked to the narratives chosen for the re-tells or the elicitation materials, e.g., GRAZE, YELL, CAPTURE, EAT, AMERICAN, SPEECH, LAUGH, SHOUT, CHEW, ANGRY, etc.

Time alignment of mouth actions with manual signs (spreading and ‘holding’) Only mouthings have been thus far annotated and processed for spreading activity. There are approximately 305 spreading mouthings (245 progressive and 60 regressive). Importantly, they are strongly associated with pointing signs (PT) with approximately 50% spread to PTs (both progressively and regressively). There are >100 cases in which mouthing articulation proper spans only one sign but the mouth shape is held progressively (for the duration of the following sign). Once again, approximately 50% involves a following PT sign.

M-type (mouthings) Most, but not all, mouthings were completely and clearly articulated (over 95%). A small number of mouthing tokens (approximately >30) were not accompanied by any manual sign yet the were clearly not redundant. They provided essential disambiguating or logico-cohesive information to the utterance. They were conjunctions, prepositions or adverbial like but, or, for, just, maybe; sentence modifiers like I-don’t-know, I-think; or other interactives like no, yes, not-true.

<table>
<thead>
<tr>
<th>Degree of articulation</th>
<th>Tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete articulation</td>
<td>8911</td>
</tr>
<tr>
<td>Initial segment</td>
<td>262</td>
</tr>
<tr>
<td>Medial segment</td>
<td>13</td>
</tr>
<tr>
<td>Final segment</td>
<td>26</td>
</tr>
<tr>
<td>Initial &amp; final segment only</td>
<td>23</td>
</tr>
<tr>
<td>Suppressed articulation</td>
<td>6</td>
</tr>
<tr>
<td>Unreadable</td>
<td>64</td>
</tr>
</tbody>
</table>

Table 6 Type of mouthing
3.3.2. Variation

Individual variation in mouth action rates, text type variation, and sociolinguistic variation are all also very important for understanding the role of mouth actions in SLs. However, these results are not reported here as these deal with wider questions on the function and interpretation of the role of mouth actions in SLs rather than the annotation and classification of mouth actions and other non-manuals.

4. Acknowledgments

Australian Research Council (ARC) grant #DP1094572. Research assistants and annotators who have contributed to the Auslan corpus in approximate chronological order since 2005 (most recent first): Ben Hatchard, Michael Gray, Gabrielle Hodge, Lindsay Ferrara, Julia Allen, Gerry Shearim, Karin Banna, Louise de Beuzeville, Dani Fried, Della Goswell, Adam Schembri.

5. References


Woll, B. (2001). The sign that dares to speak its name: Echo phonology in British Sign Language (BSL). In P. Boyes-Braem & R. Sutton-Spence (Eds.), The hands are the head of the mouth: The mouth as articulator in sign languages. Hamburg: Signum Press.
LREC 2014, Ninth International Conference on Language Resources and Evaluation

May 26-31, 2014
Harpa Concert Hall and Conference Center
Reykjavik, Iceland

Editors:
Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, Stelios Piperidis.

Assistant Editors:
Hélène Mazo, Sara Goggi

Copyright by the European Language Resources Association
ISBN 978-2-9517408-8-4
EAN 9782951740884

The LREC 2014 Proceedings are licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

Organization:

European Language Resources Association

Evaluation and Language resources Distribution Agency

Istituto di Linguistica Computazionale

LREC 2014 Committees
LREC 2014 Programme at a Glance

ARCHIVE CONTENT:

This archive contains the 742 papers accepted for presentation at LREC 2014. All papers are in PDF format.

Please read instructions below to see how you can read the files.
Workshops on the Representation and Processing of Sign Languages

LREC

scope and aims

LREC2004 Lisbon
LREC2006 Genoa
LREC2008 Marrakech
LREC2010 Malta
LREC2012 Istanbul
LREC2014 Reykjavik
LREC2016 Portorož

imprint

SCOPE AND AIMS

Workshops on sign language representation and processing at the last day of LREC are now becoming a tradition: At LREC 2004 in Lisbon, the series started with a workshop dedicated to sign notation and sign writing as well as image processing. In 2006, at LREC in Genova, the focus was on information technology in learning sign language or learning with sign language. At the third incarnation of this workshop in Marrakech in 2008, the focus already was on corpora. At LREC 2010 in Malta, we had another workshop with more than 100 participants focussing on corpora. At LREC 2012 in Istanbul, we narrowed the focus a bit down to one of the hot topics in discussions on sign language corpora: The interactions between corpus and lexicon. At LREC 2014 in Reykjavik, we had another workshop focussing on specific aspects of sign language corpora: How to deal with anything except the manual channel. At LREC 2016 in Portorož, the focus was on Corpus Mining.

Proceedings of the workshops are all available online:

of Sign Languages. 8th International Conference on Language Resources and Evaluation, LREC 2012, Istanbul, Turkey.] Paris. ELRA.


https://www.sign-lang.uni-hamburg.de/rec/workshops.html
Workshop Proceedings

6th Workshop on the Representation and Processing of Sign Languages:
Beyond the Manual Channel

Language Resources and Evaluation Conference (LREC)
Reykjavik, Iceland, 31 May 2014
Editors and Workshop Organizers

Onno Crasborn Radboud University, Nijmegen NL
Eleni Efthimiou Institute for Language and Speech Processing, Athens GR
Evita Fotinea Institute for Language and Speech Processing, Athens GR
Thomas Hanke Institute of German Sign Language, University of Hamburg, Hamburg DE
Julie Hochgesang Gallaudet University, Washington US
Jette Kristoffersen Centre for Sign Language, University College Capital, Copenhagen DK
Johanna Mesch Stockholm University, Stockholm SE

Workshop Programme Committee

Richard Bowden University of Surrey, Guildford GB
Penny Boyes Braem Center for Sign Language Research, Basel CH
Annelies Braffort LIMSI/CNRS, Orsay FR
Christophe Collet IRIT, University of Toulouse, Toulouse FR
Kearsy Cormier Deafness Cognition and Language Research Centre, London GB
Onno Crasborn Radboud University, Nijmegen NL
Svetlana Dachkovsky University of Haifa, Haifa IL
Eleni Efthimiou Institute for Language and Speech Processing, Athens GR
Stavroula-Evita Fotinea Institute for Language and Speech Processing, Athens GR
John Glauert University of East Anglia, Norwich GB
Thomas Hanke Institute of German Sign Language, University of Hamburg, Hamburg DE
Alexis Heloir German Research Centre for Artificial Intelligence, Saarbrücken DE
Jens Heßmann University of Applied Sciences Magdeburg-Stendal, Magdeburg DE
Julie Hochgesang Gallaudet University, Washington US
Trevor Johnston Macquarie University, Sydney AU
Reiner Konrad Institute of German Sign Language, University of Hamburg, Hamburg DE
Jette Kristoffersen Centre for Sign Language, University College Capital, Copenhagen DK
Lorraine Leeson Trinity College, Dublin IE
Petros Maragos National Technical University of Athens, Athens GR
John McDonald DePaul University, Chicago US
Johanna Mesch Stockholm University, Stockholm SE
Carol Neidle Boston University, Boston US
Christian Rathmann Institute of German Sign Language, University of Hamburg, Hamburg DE
Adam Schembri National Institute for Deaf Studies and Sign Language, La Trobe University, Melbourne AU
Rosalee Wolfe DePaul University, Chicago US
# Table of contents

Mouth features as non-manual cues for the categorization of lexical and productive signs in French Sign Language (LSF)  
*Antonio Balvet and Marie-Anne Sallandre*  
1

Segmenting the Swedish Sign Language corpus: On the possibilities of using visual cues as a basis for syntactic segmentation  
*Carl Börstell, Johanna Mesch and Lars Wallin*  
7

Synthesizing facial expressions for sign language avatars  
*Yosra Bouzid, Oussama El Ghoul and Mohamed Jemni*  
11

Eye gaze annotation practices: Description vs. interpretation  
*Annelies Braffort*  
19

An annotation scheme for mouth actions in sign languages  
*Onno Crasborn and Richard Bank*  
23

Implementation of an automatic sign language lexical annotation framework based on Propositional Dynamic Logic  
*Arturo Curiel and Christophe Collet*  
29

Creation of a multipurpose sign language lexical resource: The GSL lexicon database  
*Athanasia-Lida Dimou, Theodore Goulas, Eleni Efthimiou, Stavroula-Evita Fotinea, Panagiotis Karioris, Michalis Pissaris, Dimitis Korakakis and Kiki Vasilaki*  
37

A hybrid formalism to parse sign languages  
*Rémi Dubot and Christophe Collet*  
43

Non-manual features: The right to indifference  
*Michael Filhol, Mohamed Nassime Hadjadj and Annick Choisier*  
49

When nonmanuals meet semantics and syntax: Towards a practical guide for the segmentation of sign language discourse  
*Sílvia Gabarró-López and Laurence Meurant*  
55

Last train to “Rebaudengo Fossano”: The case of some names in avatar translation  
*Carlo Geraci and Alessandro Mazzei*  
63

Annotation of mouth activities with iLex  
*Thomas Hanke*  
67

Release of experimental stimuli and questions for evaluating facial expressions in animations of American Sign Language  
*Matt Huenerfauth and Hernisa Kacorri*  
71

How to use depth sensors in sign language corpus recordings  
*Rekha Jayaprakash and Thomas Hanke*  
77

**Mouth-based non-manual coding schema used in the Auslan corpus: Explanation, application and preliminary results**  
*Trevor Johnston and Jane van Roeke*  
81

**Weakly supervised automatic transcription of mouthings for gloss-based sign language corpora**  
*Oscar Koller, Hermann Ney and Richard Bowden*  
89
First Call for papers

Share this page!  

FIRST CALL FOR PAPERS

ELRA is glad to announce the 9th edition of LREC, organised with the support of a wide range of international organisations.

CONFERENCE AIMS
LREC is the major event on Language Resources (LRs) and Evaluation for Human Language Technologies (HLT). LREC aims to provide an overview of the state-of-
the-art, explore new R&D directions and emerging trends, exchange information regarding LRs and their applications, evaluation methodologies and tools, on-going and planned activities, industrial uses and needs, requirements coming from e-science and e-society, with respect both to policy issues and to scientific/technological and organisational ones.

LREC provides a unique forum for researchers, industrials and funding agencies from across a wide spectrum of areas to discuss problems and opportunities, find new synergies and promote initiatives for international cooperation, in support of investigations in language sciences, progress in language technologies (LT) and development of corresponding products, services and applications, and standards.

**CONFERENCE TOPICS**

**Issues in the design, construction and use of LRs: text, speech, multimodality**  
* Guidelines, standards, best practices and models for LRs interoperability  
* Methodologies and tools for LRs construction and annotation  
* Methodologies and tools for extraction and acquisition of knowledge  
* Ontologies, terminology and knowledge representation  
* LRs and Semantic Web  
* LRs and Crowdsourcing  
* Metadata for LRs and semantic/content mark-up

**Exploitation of LRs in systems and applications**  
* Sign language, multimedia information and multimodal communication  
* LRs in systems and applications such as: information extraction, information retrieval, audio-visual and multimedia search, speech dictation, meeting transcription, Computer Aided Language Learning, training and education, mobile communication, machine translation, speech translation, summarisation, web services, semantic search, text mining, inferencing, reasoning, etc.  
* Interfaces: (speech-based) dialogue systems, natural language and multimodal/multisensorial interactions, voice-activated services, etc.  
* Use of (multilingual) LRs in various fields of application like e-government, e-culture, e-health, e-participation, mobile applications, digital humanities, etc.  
* Industrial LRs requirements, user needs

**Issues in LT evaluation**  
* LT evaluation methodologies, protocols and measures  
* Validation and quality assurance of LRs  
* Benchmarking of systems and products  
* Usability evaluation of HLT-based user interfaces and dialogue systems  
* User satisfaction evaluation

**General issues regarding LRs & Evaluation**  
* International and national activities, projects and collaboration  
* Priorities, perspectives, strategies in national and international policies for LRs  
* Multilingual issues, language coverage and diversity, less-resourced languages  
* Open, linked and shared data and tools, open and collaborative architectures  
* Organisational, economical, ethical and legal issues.

**LREC 2014 HOT TOPICS**

**Big Data, Linked Open Data, LRs and HLT**

The ever-increasing quantities of large and complex digital datasets, structured or unstructured, multilingual, multimodal or multimedia, pose new challenges but at the same time open up new opportunities for HLT and related fields. Ubiquitous data and information capturing devices, social media and networks, the web at large with its big data / knowledge bases and other information capturing / aggregating / publishing platforms are providing useful information and/or knowledge for a wide range of LT applications. LREC 2014 puts a strong emphasis on the synergies of the big Linked Open Data and LRs/LT communities and their complementarity in cracking LT problems and developing useful applications and services.

**LRs in the Collaborative Age**

The amount of collaboratively generated and used language data is constantly increasing and it is therefore time to open a wide discussion on such LRs at LREC. There is a need to discuss the types of LRs that can be collaboratively generated and used.
Are lexicons, dictionaries, corpora, ontologies (of language data), grammars, tagsets, data categories, all possible fields in which a collaborative approach can be applied? Can collaboratively generated LRs be standardised/harmonised? And how can quality control be applied to collaboratively generated LRs? How can a collaborative approach ensure that less-resourced languages receive the same digital dignity as mainstream languages? There is also a need to discuss legal aspects related to collaboratively generated LRs. And last but not least: are there different types of collaborative approaches, or is the Wikimedian style the best approach to collaborative generation and use of LRs?

**LREC 2014 SPECIAL HIGHLIGHT**

**Share your LRs!**

In addition to describing your LRs in the LRE Map – now a normal step in the submission procedure of many conferences – LREC 2014 recognises that the time is ripe to launch another important initiative, the LREC Repository of shared LRs! When submitting a paper, you will be offered the possibility to share your LRs (data, tools, web-services, etc.), uploading them in a special LREC META-SHARE repository set up by ELRA. Your LRs will be made available to all LREC participants before the conference, to be re-used, compared, analysed, .... This effort of sharing LRs, linked to the LRE Map for their description, may become a new "regular" feature for conferences in our field, thus contributing to creating a common repository where everyone can deposit and share data.

**PROGRAMME**

The Scientific Programme will include invited talks, oral presentations, poster and demo presentations, and panels, in addition to a keynote address by the winner of the Antonio Zampolli Prize.

**SUBMISSIONS AND DATES**

Submission of proposals for oral and poster (or poster+demo) papers: 15 October 2013

Abstracts should consist of about 1500-2000 words, will be submitted through START and will be peer-reviewed.

Submission of proposals for panels, workshops and tutorials: 15 October 2013

Proposals should be submitted via an online form on the LREC website and will be reviewed by the Programme Committee.

**PROCEEDINGS**

The Proceedings will include both oral and poster papers, in the same format.

There is no difference in quality between oral and poster presentations. Only the appropriateness of the type of communication (more or less interactive) to the content of the paper will be considered.

In addition a Book of Abstracts will be printed.

**CONFERENCE PROGRAMME COMMITTEE**

Nicoletta Calzolari – CNR, Istituto di Linguistica Computazionale “Antonio Zampolli”, Pisa - Italy (Conference chair)

Khalid Choukri – ELRA, Paris - France

Thierry Declerck – DFKI GmbH, Saarbrücken - Germany

Hrafn Loftsson – School of Computer Science, Reykjavík University - Iceland

Bente Maegaard – CST, University of Copenhagen - Denmark

Joseph Mariani – LIMSI-CNRS & IMMI, Orsay - France

Asuncion Moreno – Universitat Politècnica de Catalunya, Barcelona - Spain

Jan Odič – UIL-OTS, Utrecht - The Netherlands

Stelios Piperidis – Athena Research Center/ILSP, Athens - Greece

---

**Important Dates**

**Links**