



**Marie Curie Intra-European Fellowship for career  
development  
7<sup>th</sup> Framework Programme**

# CorpAGEst

## Mid-term report

### 1. General information

<b>Grant Agreement Number:</b>	PIEF-GA-2012-328282
<b>Project Acronym:</b>	CorpAGEst
<b>Project Title:</b>	A corpus-based multimodal approach to the pragmatic competence of the elderly
<b>Funding Scheme:</b>	Marie Curie Intra-European Fellowship for career development (IEF)
<b>Duration of the Project:</b>	24 months (starting date : 01-10-2013; end : 30-09-2015)
<b>Periodic Report:</b>	Mid-term report
<b>Period Covered:</b>	from 1 <sup>st</sup> October 2013 to 30 September 2014

#### Name, title and organization of the scientific representative

<b>of the project's coordinator:</b>	Mr. Philippe CAVELIER, Délégué régional CNRS Ile-de-France Ouest et Nord Me Laurence CORVELLEC, Adjointe au délégué régional CNRS Ile-de-de-France Ouest et Nord
<b>Tel:</b>	00 33 1 45 07 57 53
<b>Fax:</b>	00 33 1 45 07 58 99
<b>E-mail:</b>	<a href="mailto:spv@dr5.cnrs.fr">spv@dr5.cnrs.fr</a>
<b>Project website address:</b>	<a href="http://corpigest.org">http://corpigest.org</a>

<b>Marie Curie Fellow:</b>	Catherine T. Bolly
<b>Scientist in Charge:</b>	Dominique Boutet
<b>Beneficiary Institution:</b>	Centre National de la Recherche Scientifique
<b>Host Laboratory:</b>	UMR7023 Structures Formelles du Langage (Dir.: Sophie Wauquier)

**October 2014**

## 2. Summary description of project context, objectives and expected results

### 2.1. Project summary [cf. Annex 1 to the Grant Agreement]

This research project aims to establish the gestural and verbal profile of very old people in normal aging, looking at their pragmatic competence from a naturalistic and ecological perspective. To reach this goal, a multimodal corpus (integrating text, sound and gesture) will serve to carry out corpus-based analyses with the aim of testing several hypotheses: (i) a change in the use of (inter)subjective verbal markers (e.g., *tu sais/vous savez* 'you know') could be an indicator of a change in the empathic ability of the elderly; (ii) we expect a change in the use of gesture (incl. hand gesture, facial expression, and eye contact) associated with increasing age, which could be an indicator of an adaptive strategy used to reduce the cost of gestural production during interaction; (iii) we expect a change in the concurrent use of verbal and non-verbal pragmatic units considered within temporality (synchronization or heterochrony between modes), which could be an indicator of a change in the pragmatic competence of the elderly's language. The dual approach, combining parametric and functional analysis of both verbal and non-verbal modes, will allow quantitative and statistical processing of language data, while guaranteeing maximum reproducibility of the analysis results. In response to European socio-economic concerns about the aging population, the project's contribution is multiple: (i) enrichment of the discussion of the concrete strategies to be implemented to promote successful aging, in particular by improving care of the elderly; (ii) enrichment of multimodal interaction annotation systems, which is a key issue for linguistic description and more generally for understanding language mechanisms; (iii) provision of a reference multimodal corpus of the language in healthy very old people that may serve as a control corpus for further comparative studies (e.g., as an aid to early detection of dementia).

### 2.2. Stated research objectives [cf. Part B of the proposal, Section B1.1 – Annexed to the Grant Agreement]

The objectives of the project are as follows [summarized here]:

- 1) The profiling of the verbal and non-verbal pragmatic competence of healthy very old people in a situated context of language communication (i.e., in their natural environment), especially in terms of their empathic and gestural ability (theoretic objective).
- 2) The establishment and annotation of a corpus that is both representative of the target population (here, healthy very old people) and the object of the study (here, the verbal and non-verbal pragmatic language markers); to understand the act of language communication in its global enunciation context, the approach will be multimodal (incorporating text, sound and gesture) (methodology objective).

### 2.3. Expected results [cf. Part B of the proposal, Section B1.1 – Annexed to the Grant Agreement]

To achieve these objectives, analyses of the study of verbal (in this instance, discourse markers) and non-verbal (in this instance, gestural markers) pragmatic language units will be performed. These studies are based on the premise that expressive (i.e., subjective) and interactional (i.e., intersubjective) pragmatic markers (be they verbal or gestural) can be seen as key indicators of pragmatic competence and thus, by extension, of empathic ability in the elderly's everyday language. The quasi-longitudinal perspective will make it possible to verify whether any observed change is able to reveal a developmental trend linked to advancing age (middle-aged > old > very old), or if the individual variation takes precedence over the developmental perspective. The expected results are set out below in the form of three main hypotheses [see Section B1.1, summarized below]:

- 1) (Inter)subjective discourse markers as indicators of a change in the pragmatic competence of very old people: (i) an overuse of subjective markers in an interactional context may indicate a strengthening of the self-centered perspective in the aging subject; (ii) a significant change in the frequency of use of intersubjective markers may indicate a difficulty in adopting the point of view of others, in other words, a loss of empathic ability, or conversely, a tendency to take interactivity more into account.
- 2) Coverbal gestures as indicators of a change in the pragmatic competence of very old people: (i) a certain gestural economy is to be expected, which would result in a reduction in the speed and amplitude of gestures; (ii) this reduction may in exchange lead to reinforcing facial mimics or a change in gaze; (iii) greater use is expected of non-representational gestures (or with a weakened representational function) co-appearing with discourse markers in very old people; (iv) a change is also expected regarding the functions fulfilled by beats, taking into account the nature of the task, the age group concerned, and the type of social tie between the interviewer and interviewee.
- 3) Interaction of verbal and non-verbal language modes in very old people: given that older people may preferentially have recourse to one or the other language modes (verbal or gestural) to achieve their communicative goal, depending on the best preserved language mode, we may therefore observe in certain subjects a dissymmetry between the use of discourse markers and the use of gestures to fulfill the same communicative function.

### 2.4. Training objectives [cf. Part B of the proposal, Section B2.1]

- 1) Introduction to a new area of research (theoretical objective): acquisition of theoretical knowledge in the area of multimodality and gesture studies to constitute a solid basis for the elaboration and verification of hypotheses relative to verbal and non-verbal language pragmatic competence of the elderly.
- 2) Expertise in multimodality analysis tools (technical objective): selection of the most appropriate tools to respond as efficiently and as best possible to the research questions raised, involving the exploration of interrogability, existing analysis tools, especially for verbal alignment, annotation and archiving multimodal corpus (meta)data.

- 3) Development of an interoperable annotation model (methodological objective): development of an annotation scheme for pragmatic competence, at the interface between the verbal and the gestural modes, to assist maximum interoperability and reproducibility, while complying with international standards.

## **2.5. Additional research training and transferable skills** [cf. Part B of the proposal, Section B2.2]

The targeted expertise and competences are [summarized below]:

- 1) Expertise in a complementary discipline (horizontal competence): (i) extension of the annotation model at the international level, applicable to different target populations (e.g., pathological elderly subjects) and in various areas of research; (ii) integration of CorpAGEst in the Valibel database (Université catholique de Louvain) to enrich the database with very old people's language and video data.
- 2) Transition towards an independent professional status (activity competence): springboard to access a university academic position in the area of research [...] to steer future research projects in global language competence of the elderly [and acquiring] the necessary expertise to be able to lead projects in the area of multimodality.
- 3) Research management, international exposure and leadership (managerial competence): strengthening of leadership, management and coordination abilities; participation in networks of researchers and in research projects will enhance the project's visibility at the European level.

### 3. Activity report

#### 3.1. Work progress and main achievements during the period

Results obtained during the reporting period are compared to the original work programme [cf. the [Work Plan in Section B4.3 of the proposal](#)]. During the first year of the project, the focus has been put on the data collection and design (Task 1), on the annotation of the nonverbal mode (Task 5) in parallel with corpus studies (Task 6), and on the ELAN software training (Task 7). The ongoing speech data treatment (Tasks 2-3) and annotation (Task 4) will be at the core of the second year of the project.

#### Task 1. Data collection

- Protocol: As planned at the very first step of the corpus building, an informed consent, a (meta)data cards and interview guides were developed (available online at <http://corpagem.org/working-papers/>). An local Ethical Committee validated the procedure for the data collection in December 2012.

As shown in table 1, each interview was replicated twice and subdivided into two subtasks. The subtasks 1A and 2A are focused on past events, whereas the subtasks 1B and 2B are centered on the current life of the old people.

Task Type	Interview N°1 (with a familiar person)	Interview N°2 (with an unknown person)
<b>Task A:</b> Descriptive task with a focus on past events	<b>Task 1A:</b> Milestones in aging	<b>Task 2A:</b> Milestones in progress
<b>Task B:</b> Explicative task with a focus on present-day life	<b>Task 1B:</b> Self-perception of aging	<b>Task 2B:</b> Self-perception of every-day environment

Table 1. Tasks for the transversal corpus data collection  
(see the *Interview Guide* – Bolly CorpAGEst © 2013)

The longitudinal part of the corpus is still in progress. It will comprise interviews from reminiscence tasks. Every interview consists in a shortened version of the interviews conducted within the framework of the CorpAGEst transversal data collection. Note that the first interview is optional and will be conducted only if there is no pre-existing transversal corpus (see for instance the G. Duboisindien Ph.D Thesis in France). This is the guarantor for the comparability of results between subcorpora.

Task Type	(Interview N°1)	Interview N°2	Interview N°3	Interview N°4
<b>Task A:</b> Focus on past events	<b>Task 1A:</b> Milestones in aging	<b>Task 2A:</b> Visual reminiscence from a picture	<b>Task 3A:</b> Olfactory reminiscence	<b>Task 4A:</b> Tactile reminiscence
<b>Task B:</b> Focus on present-day life	<b>Task 1B:</b> Society's perception of aging	<b>Task 2B:</b> Self-perception of everyday environment	<b>Task 3B:</b> Family and social relationships	<b>Task 4B:</b> Self-perception of aging

Table 2. Tasks for the longitudinal corpus data collection  
(see the *Interview Guide* – Bolly CorpAGEst © 2014)

- Subjects for the transversal corpus: It was planned to record 8 very old subjects (>75 years), healthy (without any major injury or cognitive impairment), living at home or in a residential home. To date, the corpus comprises 18 semi-directed, face-to-face interviews (see Table 3) with a very old subject (9 subjects: 8 F, 1 M; 3 in residential home, 6 at home; 16.8 hrs; approx. 250,000 words).

Recordings	hh:mm:ss	Speaker	Pseudo	Age	Birth	Sex	Education ( <i>n</i> years)	Cognition (MoCA)	Empathy (F-IRI %)
ageBN1r-1	1:01:14	ageBN1	Nadine	75	1938	F	12	29	63,57
ageBN1r-2	0:49:02								
ageLL1r-1	1:13:41	ageLL1	Louise	79	1933	F	12	26	66,43
ageLL1r-2	1:14:25								
ageBM1r-1	0:59:02	ageBM1	Anne- Marie	82	1932	F	12	28	61,43
ageBM1r-2	0:50:36								
ageDA1r-1	0:59:07	ageDA1	Albertine	84	1929	F	14	29	61,43
ageDA1r-2	0:52:41								
ageMM1r-1	1:20:40	ageMM1	Marie- Louise	84	1928	F	12	23	71,43
ageMM1r-2	0:57:06								
ageAE1r-1	0:41:35	ageAE1	Emile	86	1927	M	15	30	55,00
ageAE1r-2	0:47:00								
ageSM1r-1	0:51:14	ageSM1	Marie- Thérèse	89	1924	F	9	23	57,86
ageSM1r-2	0:58:38								
ageTL1r-1	0:49:56	ageTL1	Lucie †	92	1920	F	6	n.a.	n.a.
ageTL1r-2	0:12:47								
ageDI1r-1	1:25:34	ageDI1	Irène †	94	1919	F	8	13	75,71
ageDI1r-2	0:51:14			95					

Table 3. Main characteristics of the study subjects by chronological age (transversal corpus)

- Selection of comparable video samples (transversal corpus): The annotation procedure of the nonverbal mode required selecting and sampling the primary video sources. This was done with respect to the following methodological principles:
  - Sample 1 (interview N°1): it consists in the selection of the first 5 min. of every first interview, with the aim of exploring the way older people actually manage their language competence in a new (stressful) communication situation;
  - Samples 2 and 3 (interview N°1): they consist in one excerpt of 5 min. each occurring respectively in the middle of the first part (Task 1A: focus on the past) and second part (Task 1B: focus on the present-day time) of the interview; the aim was to build comparable samples taking sub-tasks as dependent variables;
  - Sample 4 (interview N°2): it consists in one excerpt of approx. 5 min. taken from the second part (Task 2B) of the second interview, whose thematic content must be on the perception of places to live, with the aim to compare, on the one hand, the Samples 4 to each other (dependent variable: individuals) and, on the other, to their corresponding Sample 3 (dependent variable: type of social tie between interlocutors).

To date, 14 samples have been created (see Table 4) by means of the video editor *Adobe Premiere Elements*:

Speaker	Pseudo	Recordings	Samples	hh:mm:ss.ms
ageBN1	Nadine	ageBN1r-1	Sample 1	00:05:05.00
			Sample 2	00:06:05.01
		ageBN1r-2	Sample 3	00:05:14.01
			Sample 4	00:07:59.02
ageLL1	Louise	ageLL1r-1	Sample 1	00:05:40.17
			Sample 2	00:06:38.02
			Sample 3	00:05:33.13
ageBM1	Anne-Marie	ageBM1r-1	Sample 1	00:05:34.14
			Sample 2	00:06:26.01
			Sample 3	00:05:01.11
ageDA1	Albertine	ageDA1r-1	Sample 1	00:05:10.04
			Sample 2	00:04:43.10
		ageDA1r-2	Sample 3	00:05:03.12
			Sample 4	00:05:59.22
Total duration :				01:20:11.10

Table 4. Audio-video samples for the first annotation phase (transversal corpus)

- Longitudinal corpus:** In the proposal, it was planned to carry out quasi-longitudinal studies to compare results between age groups (middle-aged people > old > very old). But, the following consideration led us to revise the protocol and to opt for a strict longitudinal approach: (i) the small size of the samples with control subjects (4 subjects planned per age group) won't allow any generalization in terms of developmental approach in aging; (ii) an additional obstacle to highlight developmental regularities is the growing individual variation that accompany the advancing age; (iii) adopting a multimodal (speech and gesture), integrative (incl. face, gaze, head, shoulders, arms, hands, legs, and feet) and form-based (fine-grained, objective and systematic) approach to language data, implies to devote a lot of time to the development of interoperable coding schemes and, therefore, doesn't allow for its application to a sufficiently wide sample of control subjects at that stage of the project. As a result, the choice was to build a coherent longitudinal corpus (see Bolly CorpAGEst © 2014). The objective is to meet and record four people among the study subjects every three months during one year, in such a way that any change in their pragmatic competence would be noticed (see the hypotheses above). The protocol has been developed in order to apply to several varieties of French (Belgium and France – see the PhD. Thesis of G. Duboisdindien in France). Recordings are planned to start in December 2014 for the longitudinal part of the project.
- Clinical tests for the cognitive and empathic abilities:** In addition to planned corpus-based methodology, clinical evaluation scales were used to serve as a basis for methodological comparison and validation: the *Montreal Cognitive Assessment* test (MoCA, Nasreddine *et al.* 2005); and the French version of the *Interpersonal Reactivity Index* (F-IRI, Gilet *et al.* 2013) for the assessment of empathy. The IRI test takes into account four cognitive and affective components of empathy: fantasy, perspective-taking, empathic concern, and personal distress. The overall scores are presented in table 3 (see above).

### Task 2 and Task 3. Transcription and alignment of the primary data

- Orthographic transcription of the sound signal: In accordance with the *Valibel* conventions (Dister *et al.*, 2009), 2 interviews (among the 18 interviews) have been fully transcribed, anonymized, standardized and partially aligned (via *EasyAlign*), with the help of one working student (Delphine Belin: Master 2 student at Paris Ouest Nanterre / 266 hrs of work). The transcription of audio data is still ongoing.

### Task 4. Pragmatic annotation of the discourse markers

- Parameter analysis: Within the framework of the *Model for Discourse Marker Annotation* (MDMA) research group at the Université catholique de Louvain, a parameter analysis of discourse markers in French have been carried out. The MDMA project (see Bolly *et al.*, Como 2014) aims at reaching feature-based criteria for the identification of discourse markers by disambiguating their uses through formal (e.g. position, co-occurrence) and semantic-pragmatic parameters (e.g. procedural vs. conceptual meaning). First results showed that, despite the great grammatical diversity of the discourse marker class, recurrent patterns of features could be revealed through multivariate analysis of several parameters annotated by four different coders.
- Functional analysis: In collaboration with L. Crible (Crible & Bolly submitted, ICLC-13 Newcastle; Bolly & Crible submitted, Ipra Antwerp), corpus-based studies aim at giving new insight in the functional annotation of discourse markers based on operational categories validated over languages (French and English) and modalities (speech and gesture). This part of the project is still ongoing. It benefits from the collaboration with the ARC project on *Fluency and disfluency marker* at the Université catholique de Louvain and from the participation in the fourth Working Group “Interoperable Annotation Guidelines” of the European COST Action *Textlink: Structuring Discourse in Multilingual Europe*.

### Task 5. Gestural pragmatic annotation

- Annotation scheme: The annotation scheme was developed in collaboration with the scientist in charge (D. Boutet) and resulted in the creation of a list of physiological parameters and tags for annotating gesture in the ELAN software. The scheme takes into account several physical articulators for the nonverbal mode (see table 5 below).

In order to make the ELAN annotation schemes (called “templates” in ELAN) easily usable from one coder to another and transposable from one recording to another, one template per group of articulators was created: (i) facial displays, (ii) hand gestures, and (iii) body gestures.



Modality: nonverbal/gesture	Articulators
1. Facial displays	
	Eyebrows
	Eyes
	Gaze
	Mouth
2. Hand gestures	
	Hands
3. Body gestures	
	Head
	Shoulders
	Torso
	Arms
	Legs
	Feet
Modality: verbal/speech	Levels of analysis
	Pragmatic markers
FUNCTION-BASED ANALYSIS	
- Multimodal annotation of emotions	
- Multimodal annotation of pragmatic functions	

Table 5. Modalities, articulators and levels of analysis in CorpAGEst (cf. *CorpAGEst Annotation Manual (I. Gestural Annotation Guidelines) (Version 1.3)* – Annexed to this report)

(i) *Facial displays*: The template is comprised of 9 annotation lines (called “Tiers” in ELAN) in relation to 4 physiological parameters (*viz.* eyebrow, eye, gaze, and mouth). In addition to the formal approach, emotions perceived from the face are annotated according to their emotion category and to their interaction with contextual and discursive cues (see Task 6 below). To date, 8 interviews have been annotated following this scheme with the help of one working student (Anaïs Thomas: Master 2 student at the University of Paris Ouest Nanterre, France / 266 hrs of work).

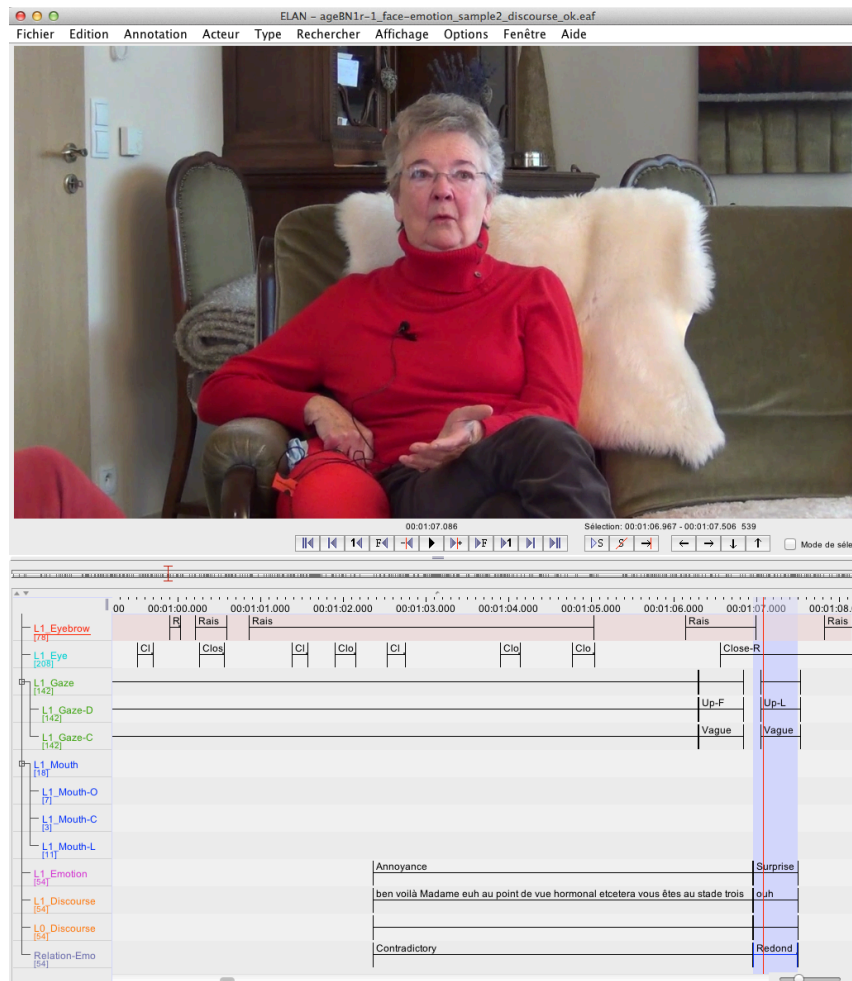


Figure 1. Annotation of facial displays (ELAN file: ageBN1r-1\_sample2)

To date, about 1 hour of video data has been fully annotated on the basis of the facial and emotional annotation scheme (4 speakers: Nadine, Albertine, Anne-Marie, Louise; samples 1, 2 and 3; duration: 66 min. 12 sec.). In addition, the study of Nadine's interactions includes an analysis of the type of relationship with speech. Those 12 annotated samples served as a basis for the study presented at the conference in Tartu (Bolly 2014, to appear).

(ii) *Hand gestures*: The template is comprised of 21 annotation lines, describing the hand moves according to their segmentation into phases, to their form (parameters: shape, orientation, position, and movement) and to the contact possibly involved (parameters: target, body/object, activity type). These parameters apply to the left and right hand, respectively. The last annotation line describes the type of symmetry for the hands, if any occurred (parameters: type of plane, parallel/alternate). This was done with the help of one temporary assistant (Anna Sáfár: Ph.D student at the Radboud University Nijmegen, The Netherlands / 157.5 hrs of work).

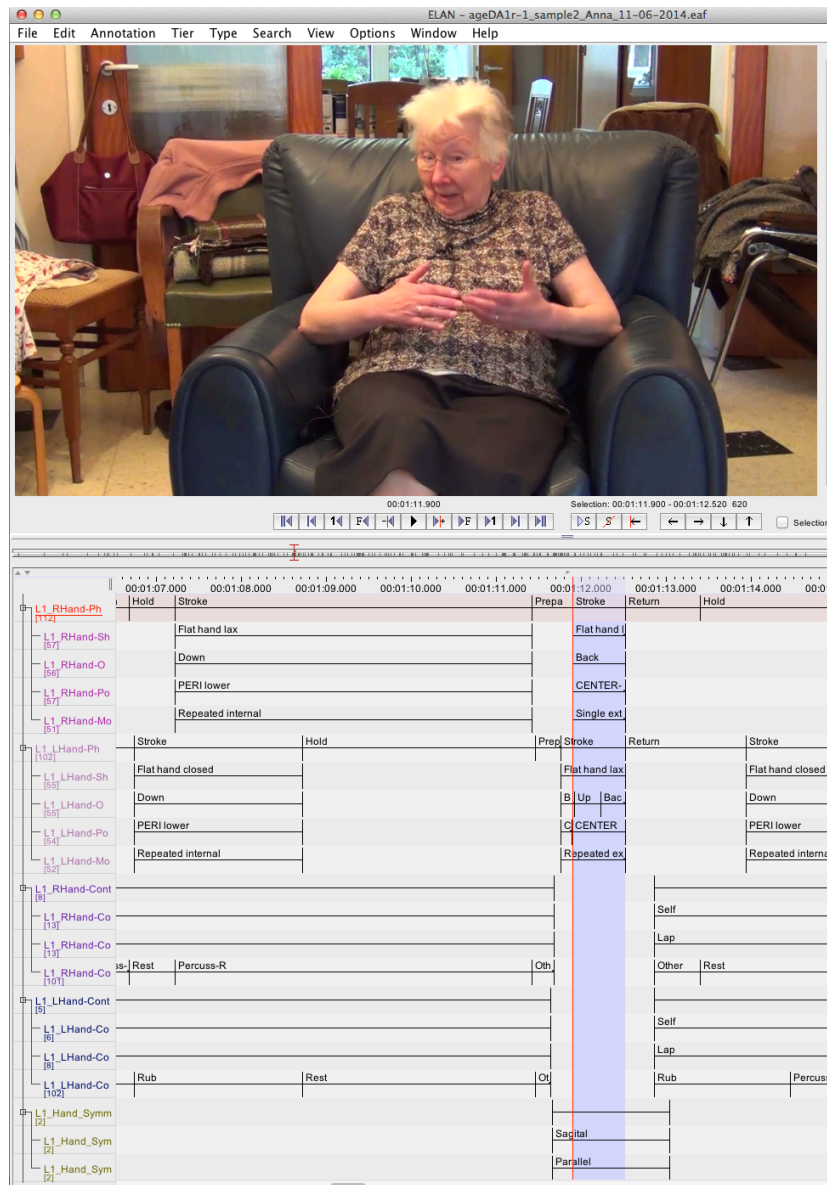


Figure 2. Annotation of hand gestures (ELAN file: ageDA1r-1\_sample2)

To date, about 40 min. of video data have been fully annotated on the basis of the annotation scheme for hand gestures (4 speakers: Nadine, Albertine, Anne-Marie, Louise; 7 samples; duration: 39 min. 07 sec.). The annotation is still ongoing for the other samples. These samples served as a basis for the study to be presented at the conference in Leuven (MaMuD 2014).

(iii) *Body gestures*: This part of the annotation scheme is still ongoing and describes visible bodily actions that concern the following articulators: head, shoulders, torso, arms, legs, and feet. It is currently comprised of 5 annotation lines dedicated to the head, shoulders and torso moves. Head moves are described according to the position and direction of the head in a three-dimensional space, with respect to three orthogonal body planes (*viz.* frontal, sagittal, and horizontal). Two working students have been of assistance in the development and stabilization of the ELAN annotation scheme for the head, shoulders and torso articulators (Alysson Lepeut, Julie Kairet: Master 2 students at the Université catholique de Louvain, Belgium / 156 hrs and 208 hrs of work).

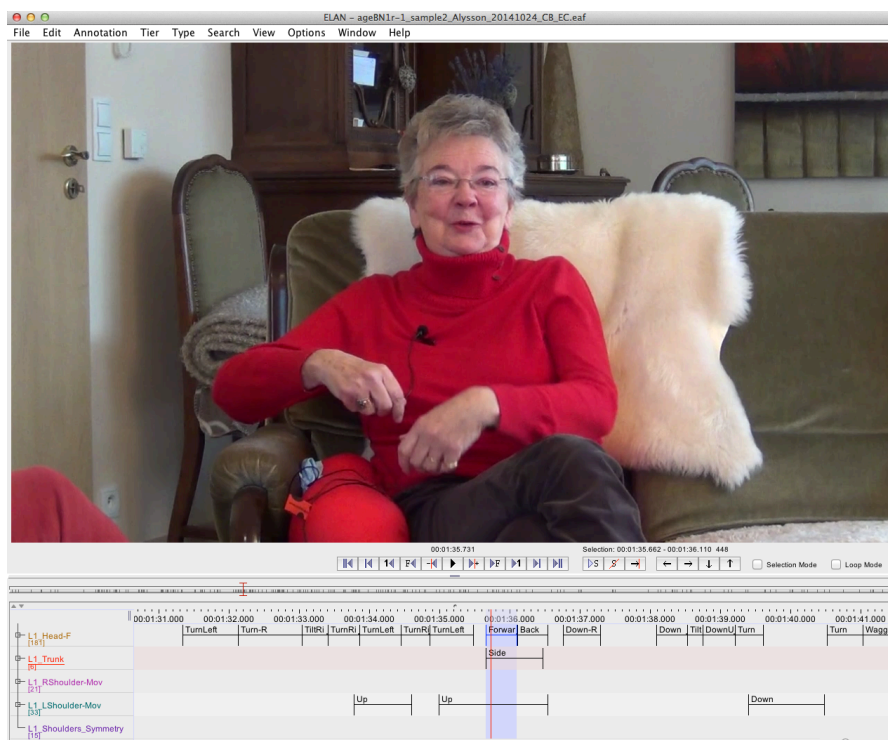


Figure 3. Annotation of body gestures (ELAN file: ageBN1r-1\_sample2)

- **Annotation guide** (for working students and external annotators): The development and trial of the annotation guide, with its necessary adjustments, are supervised by the main researcher (C. T. Bolly) with the assistance of the scientist in charge (D. Boutet) and of the temporary working students/assistants. In response to the need for maximum reproducibility, the manual is written in English. It has been designed to be further transposable to other languages (be they signed or vocal languages), to target-specific populations (e.g., younger people or people with dementia), and multiple modalities (e.g., functions of pragmatic markers in speech and gesture). The guide is currently at its 1.3 version and will be available on-line before the end of 2014. The last version of the guide includes the principles for the annotation of nonverbal resources (*viz.* face, gaze, emotions, hands, head, torso and shoulders). These principles are formulated in such a manner to be compatible with the ELAN software. The detailed description of the parameters goes hand in hand with concrete illustrations pointing to relevant excerpts from the audio-video data.
- **Pragmatic annotation of gesture:** The functional ongoing annotation scheme will be developed in such a way to be interoperable with respect to the functional annotation of speech (see above, Task 4). This task will be at the core of the second year of research.

### Task 6. Corpus-based analyses

The starting order of the corpus-based studies is the reverse of the one announced in the proposal, but it nevertheless follows the principle according to which we should go from 'mono-modal' analyses (here, from the nonverbal mode) to 'multi-modal' ones (including the level of discourse), while combining quantitative and qualitative

approaches to the data. The focus, during the first year of the project, has thus been on the gestural analysis, in relation to the emotional and attitudinal states of the very old healthy subjects. In particular, their empathic ability has been explored by means of analyses of the physiological patterning from nonverbal resources (*viz.* facial displays), in relation to emotions expressed through the face.

Preliminary results from a study of emotional and attitudinal states in four healthy elderly people (Nadine, Albertine, Anne-Marie and Louise) indicate that, despite the highly idiosyncratic use of nonverbal resources, some inter- and intra-individual tendencies emerge (see Tartu 2014; Bolly, 2014 to appear):

- (i) *Empathic ability*: Results from the empathy test (F-IRI, see above) suggest that, despite a highly significant variability in the individual profiles with respect to the four sub-categories of empathy, the healthy subjects obtain a relatively homogeneous global score of empathy and use a similar proportion of polarity emotions;
- (ii) *Emotional richness*: Whereas not in a statistically significant manner, it appears that the four subjects slightly differ with respect to their facial emotional richness, measured in terms of types of expressed emotions within the samples (Type/Token Ratio): for instance, only 9 categories of emotion for a total of 108 annotated emotions were counted in the Albertine's speech samples [TTR = 0.08], while a wider emotional panel of facial expressions were observed in the Nadine's speech (18 types for a total of 122 emotions tagged in the samples [TTR = 0.15]);
- (iii) *Physiological patterning*: Again, whereas no clear physiological pattern seems to be emotion-specific, some regularity is observed for the most frequent emotions (up to 5 occurrences per subject): for instance, in the case of Nadine, *surprise* is mainly expressed by means of eyebrow raising (19/22 cases), often combined with an exaggerated opening of the eyes (12/22 cases), but this combination of parameters is also true for *fear* (5/7 cases), *disappointment* (3/8 cases) and *annoyance* (3/16 cases);
- (iv) *Multimodality and the speech-gesture interface*: A closer look at the Nadine's speech has shown that the most frequent emotions appeared to be congruent (complementary or redundant) with the contextual and linguistic information. Yet, some facial emotions sometimes contradict the information conveyed by the context. For instance, most of the time, the annotation of *joy* does not mirror the information expressed by Nadine and must be disambiguated thanks to the linguistic context (e.g., *c'était un peu jeune quoi hein j'ai été un peu malheureuse là* 'I was a little bit too young well I have been quite unhappy there').

This pilot study thus provided a first insight into the interaction of the verbal and nonverbal language modes (see Study No. 3 in the proposal). Studies whose aim was to detect any change in the use of (inter)subjective discourse markers and gestures in the elderly people (see Studies No. 1 and No. 2 in the proposal) will be carried out during the last year of the project, on the basis of the longitudinal data.

Additional corpus-based studies are undergone, which aim at exploring more in-depth several aspects of the project:

- Exploring the link between multimodal pragmatics (including pragmatic gestures and discourse markers) and cognitive processes/representation? (cf. Bolly 2014,

MaMuD Leuven; Bolly submitted, ICLC-13 Newcastle; Crible & Bolly submitted, ICLC-13 Newcastle);

- Exploring the intersubjective functions of (non)verbal pragmatic markers in intergenerational interactions (cf. Bolly 2015, CRIS Neuchâtel; Master Thesis of A. Lepeut in Louvain);
- Exploring the role of non-representational gestures in the interaction (cf. Bolly 2014, MaMuD Leuven; Master Thesis of R. Martin in Paris);
- Exploring the interoperability of models dedicated to the annotation of discourse markers (cf. MDMA working group in Louvain) and of pragmatic gestures (cf. Bolly & Crible submitted, Ipra Antwerp; Crible & Bolly submitted, ICLC-13 Newcastle);
- Exploring comparable corpora comprised of face-to-face interviews with very old people, with a focus on specific phenomena such as repetition (cf. Gerstenberg & Bolly submitted, Ipra Antwerp);
- Exploring the impact of (non)verbal pragmatic markers on the communicative competence of very old people over time (cf. Ph.D. Thesis of G. Duboisdindien in Paris).

### **Task 7. ELAN training**

The progress in this software training is described in Section 3.2 (“Project management”).

### **Task 8. Dissemination and publication**

Several working documents directly linked to the project are available online (see at <corpagest.org>), making them publicly available for the scientific community. Additional scientific activities (such as collaboration and network, theses supervision, conferences, publication, etc.) are listed in Section 3.3 (“Dissemination and transfer of knowledge activities”).

### **Task 9. Coordination and logistics**

The research fellow has benefited during the first year of research in Paris from:

- Administrative support in the host research center UMR7023 (administrative staff: G. Morand and I. Biercewicz): furniture, logistics, secretariat, missions’ management, etc.;
- Human resources support in the host research center UMR7023: management of contracts for working students;
- A room equipped with four computers, headphones and Internet facilities, made available for the CorpAGEst meetings (working students, assistant, main researcher, etc.).

A dedicated project website was also created at the initiative of the research fellow (see <http://corpagest.org>).

### 3.2. Project management and progress of training activities

- 1) Introduction to a new area of research (theoretical objective): Thanks to her participation in numerous seminars, conferences and thematic school in France and abroad (see 3.3.), the fellow has strengthened her knowledge in the field of gesture studies and aging studies. She also benefited from the access to the local library of the host research center UMR7023 Structure Formelles du Langage and from regular fruitful discussion with the scientist in charge of the project. The acquisition of the most recent scientific publication in the field of (i) language and aging, (ii) discourse studies and pragmatics, and (iii) gesture studies, also consolidated the bases of the required multi-disciplinary theoretical background, while permitted the development of more fine-grained hypotheses in the project.
- 2) Expertise in multimodality analysis tools (technical objective): Multiple competences have been acquired during the first year of the project, including the following ones:
  - Edition of video-data, by means of the video editor *Adobe Premiere Elements*.
  - Edition of audio-data, by means of the audio editor *Audacity*.
  - Training in the Valibel – Discours et Variation center (Université catholique de Louvain, Belgium) to the verbal transcription (using the *Praat* program) and alignment (using the *EasyAlign* plugin for *Praat*).
  - As previously planned, the research fellow has benefited from personalized training to become familiar with the annotation of audio-video data through the *ELAN* software. This training took the form of:
    - Regular research meetings in the host laboratory at the UMR7023 Structures Formelles du Langage, under the supervision of the scientist in charge (D. Boutet), with all the members working on the project at the time concerned (temporary assistants, working students, etc.);
    - Attendance of courses dedicated to the annotation of audio-video data by means of the *ELAN* software, for the technical aspects of the annotation: IRCOM training school in France (Paris, November 2013) and international summer school in Estonia (Tartu, August 2014).
 Note that in order to limit the number of different tools to be used and, *in fine*, to gain time in the audio-video data treatment (at a preliminary stage of analysis), annotation was made only by means of the *ELAN* software (excluding thus the use of the *EXMaRALDA* software). Moreover, it appeared from a comparison between equivalent annotation tools that *ELAN* was one of the most interoperable ones to date, notably for the import/export function and queries.
- 3) Development of an interoperable annotation model (methodological objective): An annotation scheme for pragmatic competence, at the interface between the verbal and the gestural modes, is underway. It aims at assisting maximum interoperability and reproducibility, while complying with international standards. See above the description of the work in progress for the Task 5 and, in particular, the creation of the annotation guide.

### 3.3. Dissemination and transfer of knowledge activities

- 1) Permanent archival [Task 8.3 in the Work Programme]: Archiving multimodal corpus (meta)data: thanks to her coordination of the 6<sup>th</sup> IRCOM Group on archival of oral and multimodal data, C. T. Bolly has gained a comprehensive view of the existing resources in France with this respect. She is also actively involved in the creation of a Digital Humanities group at the University of Louvain (LODIH). Two possibilities are envisaged. First, the archival of the audio and video data in the Valibel database, together with their transcription, implies to develop the actual interface in order to welcome the deposal of video data. Secondly, the archival of the audio and video data, with their transcription and annotation, could be made on the CoCoON or Ortolang resource in France, but this implies a much longer procedure for the metadata transformation in XML language. The decision will be made before the end of 2014.
- 2) Dissemination within the scientific community [Task 8.2 in the Work Programme,]: The project's content has mainly been disseminated by the research fellow (C. T. Bolly) during the period concerned, through the creation of one international research group (CLARe), the participation in experts' networks and interdisciplinary consortia (e.g., IRCOM, Louvain4Ageing), the organization of and participation in several international conferences (see below). In addition, research results were/will be disseminated by means of (peer-reviewed) oral presentations at international scientific events (see below). Several scientific papers are planned to appear or to be submitted during the second year of the project. The research fellow is also currently involved in the supervision of PhD. Theses (2) and Master Theses (3), in relation to the CorpAGEst project, mainly in collaboration with the University of Louvain (Belgium) and of Paris Ovest Nanterre (France) (see below). All the scientific activities undergone during the reporting period (and the upcoming events) are listed in the following pages.



## List of scientific activities for the reporting period

### (i) Collaboration and Scientific Network

- April 2014 Co-founder of the CLARE research group (*Corpora for Language and Aging Research*) (with. A. Gerstenberg, Freie Universität Berlin) (Belgium and Germany)
- 2013- Member of the Pilot Committee of the Linguistic Consortium *IRCOM (Corpus Oraux et Multimodaux)* [Consortium for Spoken and Multimodal Corpora of Corpus-IR], TUL - FR 2559 (France)
- 2013- Member of the interdisciplinary Consortium Louvain4Ageing, Université catholique de Louvain (Belgium)
- 2013- Member of the European COST Action *Textlink: Structuring Discourse in Multilingual Europe* (Working Group “Interoperable Annotation Guidelines”) (ISCH IS1312)
- 2013- Collaboration in the *Orfeo Project (Outils et Recherches sur le Français Ecrit et Oral)* [Tools and Studies on Written and Spoken French] (Coord. J.-M. Debaisieux), ANR Corpus (call 2011), Université Paris 3 Sorbonne Nouvelle (France)
- 2013- 2014 Coordinator of the Working group 6 (with M. Toda): *Stockage sécurisé et mutualisation des corpus oraux et multimodaux* [Archival of Spoken and Multimodal Corpus Data] in the Linguistic Consortium *IRCOM (Corpus Oraux et Multimodaux)* [Consortium for Spoken and Multimodal Corpora of Corpus-IR], TUL - FR 2559 (France)
- 2012- Scientific coordinator of the Working Group *MDMA (Model for Discourse Marker Annotation)* (with F. Ciabbarri, L. Crible, L. Degand, D. Uygur-Distexhe), Centre Valibel – Discours et variation, Université catholique de Louvain, Louvain-la-Neuve (Belgium)
- 2012- Member of the Linguistic Consortium *IRCOM (Corpus Oraux et Multimodaux)* – Working group 4: *Multimodalité et modalité visuo-gestuelle* [Consortium for Spoken and Multimodal Corpora of Corpus-IR], TUL - FR 2559 (France)
- 2012- Collaboration in the A.R.C. Project *Fluency and disfluency markers. A multimodal contrastive perspective*, Université catholique de Louvain & Université de Namur, Louvain-la-Neuve (Belgium)

### (ii) Publication

#### *Journal Articles (peer-reviewed)*

- Bolly, C. T. (2014). Gradience and gradualness of parentheticals. Drawing a line in the sand between phraseology and grammaticalization. Mouton – De Gruyter (eds.), *Yearbook of Phraseology* 5, 25-56.
- Bolly, C. T. & Degand, L. (2013). Have you seen what I mean? From verbal constructions to discourse structuring markers. *Journal of Historical Pragmatics* 14(2), 210-235.

#### *Proceedings (peer-reviewed)*

- Bolly, C. T. (to appear). “Facing Nadine’s speech. Multimodal annotation of emotion in the elderly”, Proceedings of *The Second European and the 5th Nordic Symposium on Multimodal Communication*, 6-8 August 2014, University of Tartu, Tartu (Estonia).

#### *Working Papers (online publication)*

- Bolly, C. T. (in progress). *CorpAGEst Annotation Manual (Version 1.3)*.
- Bolly, C. T. (2014). *CorpAGEst Interview Guide and Corpus Design (II. Longitudinal Part)*. Working Paper available online at <http://corpagest.org> [Bolly CorpAGEst © 2014]
- Bolly, C. T. (2013). *CorpAGEst Interview Guide and Corpus Design (I. Transversal Part)*. Working Paper available online at <http://corpagest.org> [Bolly CorpAGEst © 2013]

### (iii) Conferences, Seminars

#### Conference Organization

Pilot Committee: International CLARe Workshop on *Communicating with the elderly people. What about their language in use?* (Org.: C. T. Bolly; CLARe research group, Louvain4Ageing consortium and Valibel center), 8 October 2014, Université catholique de Louvain, Louvain-la-Neuve (Belgium)

Organizing Committee: One-day Conference on *Multimodality: New questions, interdisciplinary perspectives, and new methodologies* (Org.: C. T. Bolly, B. Garcia, E. Soroli, V. Vapnarsky, C. Vincent; Working group 4 of the Linguistic Consortium *IRCOM*), 6 June 2014, INALCO, Paris (France)

#### Invited Talks

“Pragmatique et gestualité : pour une meilleure compréhension du langage des personnes âgées”. International CLARe Workshop on *Communicating with the elderly people. What about their language in use?* (Org.: C. T. Bolly; CLARe research group, Louvain4Ageing consortium and Valibel center), 8 October 2014, Université catholique de Louvain, Louvain-la-Neuve (Belgium)

“Valibel. La gestion des corpus oraux à l’ère du numérique” (with L. Degand, M. Francard, A. C. Simon). Conference *Corpus de français parlés et français des corpus* (Org.: M. Avanzi, M.-J. Béguelin & F. Diémoz), 8-9 May 2014, Université de Neuchâtel, Neuchâtel (Switzerland)

“What corpora for discourse and aging studies ? From words to gestures (and conversely) / Quels corpus pour nos aînés? Du verbe au geste et du geste au verbe” (Org.: A. Gerstenberg). 23 April 2014, Freie Universität of Berlin, Berlin (Germany)

“Quand la linguistique parle de, avec et pour les aînés”. Communication at the 2<sup>nd</sup> Scientific Meeting of the *Louvain4Ageing* network, 21 May 2014, Université catholique de Louvain, Louvain-la-Neuve (Belgium)

“Discourse and aging” (Hamilton, 2001) and “Language and (inter)subjectivity in normal aging” (Bolly, 2012). Presentation at the Journal Club de Gérontologie et Gériatrie & Aging Research Group (IRSS) (Org.: B. Boland, I. De Brauwer), 2 April 2014, Université catholique de Louvain, Woluwé-St-Lambert (Belgium)

“Des gestes et des mots pour dire la vieillesse”. Research seminar of the UMR 7023 Structures Formelles du Langage (Org.: P. Cabredo Hofherr, E. Soare), 24 March 2014, University of Paris 8, UMR7023 Structures Formelles du Langage (SFL), Paris (France)

#### Conference Talks (peer-reviewed)

“Going cognitive. Linguistic categories for pragmatic markers across languages and modalities” (submitted) (with L. Crible). Theme session: “The cognitive commitment 25 years on: are linguistic categories cognitively real(istic) (and do they need to be)?” (Org.: D. Divjak, J. Klavan, N. Levshina), *13th International Cognitive Linguistics Conference (ICLC-13)*, 20-25 July 2015, Northumbria University, Newcastle (United Kingdom)

“Towards pragmatic gestures: From repetition to construction in multimodal pragmatics” (submitted). Theme session: “Grammar, Speakers’ Gestures, and Conceptualization” (Org.: A. Cienki, G. Brône), *13th International Cognitive Linguistics Conference (ICLC-13)*, 20-25 July 2015, Northumbria University, Newcastle (United Kingdom)

“From context to functions and back again: Disambiguating pragmatic uses of discourse markers” (accepted) (with L. Crible). Panel session: “Anchoring utterances in co(n)text, argumentation, common ground” (Org.: K. Fischer, M. Alm), *14th International Pragmatics Conference (IPra)*, 26-31 July 2015, Antwerp (Belgium) “On the meaning potentials of pragmatic (micro-) gestures” (accepted), 2<sup>nd</sup> *MaMuD – Mapping Multimodal Dialogue* workshop, 21-22 November 2014, KU Leuven, Louvain (Belgium)

“Functions of repetition in the discourse of elderly speakers: The role of prosody and gesture” (accepted) (with A. Gerstenberg). Panel session: “Age and language use” (Org.: Chr. Englert), *14th International Pragmatics Conference (IPra)*, 26-31 July 2015, Antwerp (Belgium)

“Towards a Model for Discourse Marker Annotation in spoken French: From potential to feature-based discourse markers” (with L. Crible, L. Degand, D. Uygur-Distexhe), International Workshop *Pragmatic Markers, Discourse Markers and Modal Particles: What do we know and where do we go from here?*, 16-17 October 2014, Università dell’Insubria, Como (Italy)

“Facing Nadine’s speech. Multimodal annotation of emotions in the elderly”, *The Second European and the 5th Nordic Symposium on Multimodal Communication*, 6-8 August 2014, University of Tartu, Tartu (Estonia)

#### *Participation (without communication)*

One-day Conference: *Maladie d’Alzheimer: Des pratiques pour plus de dignité* (Org.: Le Bien Vieillir ASBL), 22 September 2014, Centre culturel Marcel Hicter, La Marlagne, Wépion (Belgium)

One-day Conference: *Construire le bien-être des personnes en perte d’autonomie* (Org.: Convergences pour l’Innovation Sociale), 27 March 2014, Louvain-la-Neuve (Belgium)

International Workshop: JET AFLiCo (Association Française de linguistique cognitive), *International Workshop on Stance and (Inter)Subjectivity* (Org.: C. Debras, G. Furmaniak, A. Morgenstern), 22 March 2014, Université Paris 3 Sorbonne Nouvelle, Paris (France)

One-day Conference: *Usages des sources numériques en histoire des sciences et des techniques III* (Org.: C. Blondel, S. Pouyllau, T. Charmasson; Pôle Histoire des Sciences et des Techniques en Ligne), 29 November 2013, Cité des sciences et de l’industrie, Paris (France)

#### **(iv) Training Schools**

Aug. 2014 Research Training Course *Pointing to Gestures* (Org.: K. Jokinen, E. Ahlsèn, J. Allwood, C. Navarretta, P. Paggio, S. Tenjes), 4-5 August 2014, University of Tartu (Estonia)

Nov. 2013 Training Course IRCOM on multimodal corpora *Notation, annotation et analyse de corpus multimodaux avec ELAN* (Org.: D. Boutet - IRCOM), 6-7 November 2013, Paris 8, CNRS Pouchet, Paris (France)

#### **(v) Direction, Supervision of Master and Ph.D Theses**

2014-2017 Co-director of the Guillaume Duboisindien Ph.D Thesis: *Analyse multimodale des marqueurs pragmatiques à fonction intersubjective au sein du vieillissement langagier, en tant qu’indices précoces de démence* [Multimodal approach to intersubjective pragmatic markers in aging as indicators of early-dementia], Modyco, Université Paris Ouest Nanterre, Paris (France)

2013-2017 Member of the Supervision Doctoral Committee of the Iulia Grosman Ph.D Thesis: *La (dis)fluence prosodique en français* [Prosodic (Dis)fluency Markers in French]. A.R.C. Project *Fluency and disfluency markers. A multimodal contrastive perspective*, Université catholique de Louvain & Université de Namur, Louvain-la-Neuve (Belgium)

2013-2015 Promoter (with F. Meunier): Master Thesis in Linguistics on the gestural expression of intersubjectivity in the elderly people (A. Lepeut), Université catholique de Louvain, Louvain-la-Neuve (Belgium)

2013-2014 Promoter (with A. Lacheret): Master Thesis in French Linguistics *Etude des expressions faciales, émotions et empathie chez les personnes âgées* [Facial expressions, emotions and empathy in the elderly people] (A. Thomas), Université Paris Ouest Nanterre, Paris (France)

2011-2013 Promoter (with L. Degand): Master Thesis in French Linguistics: *Analyse sur corpus des marqueurs de discours cumulés chez les personnes âgées* [Complex discourse markers in the elderly] (F. Sprumont), Université catholique de Louvain, Louvain-la-Neuve (Belgium)