Facing Nadine's Speech. Multimodal Annotation of Emotion in Later Life

Catherine Bolly
CNRS, UMR 7023 SFL and
Université catholique de Louvain
catherine.bolly@uclouvain.be

Anaïs Thomas
University of Paris Ouest Nanterre
anais2.thomas@gmail.com

Abstract

This pilot study aims at reconstructing the empathic profile of conversation participants from their interaction in real-world settings. It addresses the question of how verbal and nonverbal modes converge in conveying information about the emotional and attitudinal behavior in everyday communication. In particular, the empathic ability of older people is explored studying physiological patterning from nonverbal resources, in relation to emotions expressed through the face. In addition, the IRI psychometric test of empathy provides the participant's overall empathic profile. The data is taken from the CorpAGEst multimodal corpus, and focuses on the language of four healthy very old women who obtained a normal score on the MoCA cognitive test. Preliminary results indicate that, despite the highly idiosyncratic use of nonverbal resources, some inter- and intra-individual tendencies seem to emerge.

1 Introduction

The main objective of this paper is to investigate the extent to which we can (or cannot) reconstruct the emotional and attitudinal profile of conversation participants from their interaction in real-world settings. At the core of the study, the following question is addressed: "To what extent do the verbal and nonverbal modes converge in the information they convey about the emotional and attitudinal behavior of people in their everyday communication?" In particular, the empathic ability of very old healthy people is explored by analyzing their physiological patterning from nonverbal resources, in relation to emotions expressed through the face. In addition to the corpus-based approach, the recourse to the IRI psychometric test of empathy provides a more precise picture of the participant's overall empathic profile. Taking for granted the multidimensionality of empathy and the multimodality of emotions (Martin et al., 2006), the present study is a tentative effort to access the complexity of human beings' communication through the lens of complementary approaches to language in interaction.

2 Background

2.1 Pragmatic competence in later life

To date, only very little attention has been paid to the study of *pragmatic competence* – that is, the ability to use language resources in a contextually appropriate manner (Kasper and Rose, 2002) – of healthy older people from the angle of language production in a natural environment. Yet, the existence of pragmatic features specific to communication mode in the older people is recognized, which shows change in the interlocutors' behavior and increased (off-target) verbosity. On the one hand, it has been observed that speakers often adjust their way of speaking and gesturing to accommodate to the older people speech, and switch from their common way of speaking to a so-called "elderspeak" (Harwood, 2007). On the other hand, the Pragmatic Change Hypothesis (James et al., 1998) argues that the decrease in coherence – which goes together with an increase in amount of speech (*viz.* verbosity) in the older people – would result from a strategy to adapt their speech style according to communicative goals and social context.

K. Jokinen and M. Vels. 2015. Proceedings of The 2nd European and the 5th Nordic Symposium on Multimodal Communication. This work is licensed under a Creative Commons Attribution 4.0 International License:
http://creativecommons.org/licenses/by/4.0/

2.2 Embodied emotion and empathic ability in aging

Emotions are grounded in the here-and-now experience of conversation participants. As such, they form part of our pragmatic competence: "Emotions are defined as short-term, biologically based **patterns** of perception, subjective experience, physiology, and action (or action tendencies) that constitute responses to specific physical and social problems posed by the environment" (Niedenthal et al., 2005: 22). It is worth stressing that the recognition of these emotional patterns is a complex process depending on the situational context, on the affective state, social and cultural identity of the participants (Russell et al., 2003: 334). What research in interpersonal pragmatics should therefore include in its scope goes far beyond the level of discourse, and must also address the embodied dimension of emotion as being part of the fuller context of interaction (Niedenthal, 2007). Facial expressions are particularly recognized as a major conveyance of both affective and cognitive stance, that is, of intersubjective evaluation, positioning, and alignment of language users in a situation of collaborative interaction (Englebretson, 2007). They may also have an emotion-regulating function (in the communicating person) and provoke empathic inferences (in the interlocutor).

Empathy is generally defined as the cognitive and affective ability to understand others' emotions and points of view, as well as to be in-tune with their emotional states (Eisenberg et al., 2014). We will distinguish here between *empathy*, seen as the result of both affective and cognitive processes that are self- and other-oriented, and *sympathy*, defined as the "other-oriented desire for the other person to feel better" (Eisenberg and Fabes, 1990: 132). In the domain of aging and neuropsychology, results indicate that the healthy subjects' advancing age may be accompanied by a loss of empathic ability (Bailey and Henry, 2008), liable to affect their ability to successfully engage in social interaction.

2.3 The CorpAGEst project: pragmatics, aging and language in use

The present paper is part of the CorpAGEst project ("A corpus-based multimodal approach to the pragmatic competence of the elderly"), which aims to establish the gestural and verbal profile of very old people in aging, looking at their pragmatic competence from a naturalistic perspective. The CorpAGEst assumption is that multimodal (inter)subjective markers of stance are highly relevant cues for the measurement of empathic ability of the older people: "Since evaluation is tied to affect, stance taking in the here-and-now of interaction serves to link affect to aspects of ideological systems and their expressions, including language, gesture, body practices, rhetoric, socialization, prior text, arts, aesthetic artifacts, and more" (Du Bois and Kärkkäinen, 2012: 437).

3 Objectives and research questions

This paper investigates the extent to which we can (or cannot) reconstruct the emotional and attitudinal profile of healthy very old people in their everyday interactions. The focus is on the synchronic and individual aspect of language competence in later life, without any longitudinal perspective or comparison between age groups at that stage. Hence, this pilot study has to be considered as a preliminary step for further investigation in a longer-term perspective within the framework of the project CorpAGEst: (i) the ongoing annotation of hand moves and body gestures, together with the functional analysis of pragmatic markers and gestures in the corpus, will strengthen the multimodal scope of the study; (ii) the longitudinal approach will allow for the detection of any individual change in the old persons' way of speaking and gesturing with advancing age; (iii) the widening of the sample of study subjects will make the results more likely to be generalized, at least to some reasonable extent.

For the purpose of the present study, the following questions are addressed: "To what extent do the verbal and nonverbal modes converge in the information they convey about the emotional and attitudinal behavior of people in their everyday communication?" And, more precisely: "What can verbal and nonverbal emotional or attitudinal markers reveal about the empathic ability of the very old person?" Such markers may consist in verbal pragmatic markers of stance (e.g., *enfin* 'well'; *tu sais* 'you know') or in nonverbal resources that have an expressive or interactive function (e.g. a wide opening of the eyes to indicate surprise; a gaze towards the interlocutor to maintain his or her attention).

4 Data

4.1 Study subjects and tasks

The CorpAGEst corpus (Bolly, 2013) is comprised of semi-directed, face-to-face conversations between an adult and a very old subject (75 y. old and more) living at home or in a residential home, which have been audio and video recorded. The corpus is two-fold: the cross-tasks corpus currently comprises 18 interviews (9 subjects; mean age: 85; duration: 16.8 hrs.); the longitudinal corpus (still in progress) will comprise interviews based on a shortened protocol from reminiscence tasks (see http://corpagest.org). Contextual independent variables are part of the corpus design, namely environment (private *vs.* residential home), the social tie between the participants (familiar *vs.* unknown interviewer) and the task type (focusing on past events *vs.* present-day life) (see Table 1). Metadata provide information about the interaction situation (e.g., date, place, quality of the recordings) and the participants (e.g., sex, education, profession, mother tongue, geographic origin, etc.).

Task Type	*Interview N°1 (with a familiar person)*	*Interview N°2 (with an unknown person)*
Task A: Focus on past events	Task 1A: Milestones in aging	Task 2A: Milestones in progress
Task B: Focus on present-day life	Task 1B: Self-perception of aging	Task 2B: Self-perception of every-day environment

Table 1. Tasks for the transversal corpus data collection

4.2 Clinical evaluation

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). The IRI test takes the form of a questionnaire that takes into account four components of empathy, the first two being part of the cognitive dimension of empathy, and the last two being part of its affective dimension: (i) *Fantasy* is defined as "the tendency to imaginatively transpose oneself into fictional situations"; (ii) *Perspective-Taking* relates to "the tendency to spontaneously adopt the psychological view of others in everyday life"; (iii) *Empathic Concern* corresponds to "the tendency to experience feelings of sympathy or compassion for unfortunate others"; (iv) *Personal Distress* concerns "the tendency to experience distress or discomfort in response to extreme distress in others" (Davis, 1994: 55-57). Among the nine old people from the CorpAGEst corpus, only the four who obtained a normal score at the cognitive test (equal or more than 26/30) were selected for the present study (n: 4; sex: F; mean age: 80 – see Table 2).

Recordings	*hh:mm:ss*	*Speaker ID Code*	*Pseudo*	*Age*	*Birth*	*Sex*	*Education (*n *years)*	*Cognition (MoCA)*	*Empathy (F-IRI)*
ageBN1r-1 ageBN1r-2	1:01:14 0:49:02	ageBN1	Nadine	75	1938	F	12	29/30	64 %
ageLL1r-1 ageLL1r-2	1:13:41 1:14:25	ageLL1	Louise	79	1933	F	12	26/30	66 %
ageBM1r-1 ageBM1r-2	0:59:02 0:50:36	ageBM1	Anne-Marie	82	1932	F	12	28/30	61 %
ageDA1r-1 ageDA1r-2	0:59:07 0:52:41	ageDA1	Albertine	84	1929	F	14	29/30	61 %

Table 2. Main characteristics of the study subjects by chronological age (transversal corpus)

5 Method

Taking for granted the multidimensionality of empathy and the multimodality of emotions, the perspective adopted combines notions and methods inherited from various disciplines. About 1 hour of video data was fully annotated on the basis of facial physiological parameters (section 5.2) and emotional states (section 5.3). In addition, the data were partly analyzed in terms of multimodal relationship with speech (section 5.4). All annotations were done by one investigator and partly crosschecked by the other one, mainly during the learning phase, in order to develop, improve and stabilize the cod-

ing scheme. The second investigator also served as control for uncertain and ambiguous cases.

5.1 Multimodal approach

This pilot study corresponds to the very first step of the annotation procedure within the framework of the CorpAGEst project, which aims *in fine* at a better understanding of the way in which the verbal and gestural dimensions interact to make sense in real-world settings. Starting with mono-modal analyses (gesture *vs.* speech) and focusing on one group of articulators at a time within each modality (*viz.* face, gaze, head, shoulders, torso, hands, legs, and feet), the annotation procedure next moves to multimodal analyses. Consequently, the present study mainly concentrates on facial displays, gaze, and emotions perceived from the face. A first insight into the interaction of physiological and emotional parameters with contextual and discursive cues is given at the end of the paper (see section 6.4).

The text, sound and video data were aligned using the *ELAN* software (Wittenburg et al., 2006). The multi-level annotation of the audio and video samples (3*5 min. per interview) was performed as follows: (i) annotation of the physiological parameters for the face; (ii) annotation of emotions expressed through the face (no recourse to the sound signal); (iii) annotation of the relation between the tagged emotion and the contextual information (taking into account gestures and linguistic information).

5.2 Facial expressions and gaze

In line with form-based approaches to gesture (Müller et al., 2013) and mainly inspired from the *MUMIN* project (Allwood et al., 2007), the ELAN annotation scheme dedicated to the physiological description of facial expressions is comprised of 7 parameters (see Table 3 below).

Articulator	*Variable*	*Values / Labels*
Eyebrows	Form	Frowning, Raising, Other
Eyes	Form	Exaggerated Opening, Closing-Both, Closing-One, Closing-Repeated, Other
Gaze	Direction	Forward-Front, Forward-Right, Forward-Left, Up-Front, Up-Right, Up-Left, Down-Front, Down-Right, Down-Left, Other
	Target	Addressee, Other participant, Vague, Object, Body part, Camera, Other
Mouth	Openness	Open
	Lips' corners	Up, Down, Other
	Lips' shape	Protruded, Retracted, Other

Table 3. Articulators and physiological parameters for facial expressions

Facial displays (including gaze) were identified according to their location in the face (eyebrows, eyes, gaze, mouth) and then annotated in terms of physiological features (e.g., *closed-both* for the eyes, *corners up* or *retracted* for the lips). The annotation was made independently of the sound signal to avoid any interpretive bias in the semiotics of gesture at this stage in the analysis. Movements were identified according to the following principle: the left boundary of each annotation – that is, the beginning of the move – has been assigned to the first frame that corresponds to a visible change in the face, mostly on a blurred image (e.g., when the eyes begin to close, not when they are completely closed); in the same manner, the right boundary of facial expressions – that is, the end of the move – has been put on the frame corresponding to the absence of any visible change, mostly a fixed image (e.g., when the eyes are fully open again). It is of great importance here to stress some methodological issues. First, although the beginnings of facial moves were quite easy to detect, many of them were disappearing with a fading effect. In those cases, the right boundary has been put on the frame corresponding to the recovered neutral position. Other physiological features (e.g., wrinkling the forehead while eyebrow raising) were also used as support to detect the very end of such fading moves. Secondly, openness of the mouth and moves from the lips were only taken into account when not accompanying speech production. For these exceptions only, the sound signal was activated to distinguish between the two possibilities (with or without speech). Thirdly, we chose to annotate gaze all along the samples (with the obvious exception of closing eyes) rather than delimiting so-called "gaze-units", in order to highlight transitions in gaze direction and nature of the target.

5.3 Attribution of emotions

In the present study, emotions were annotated by looking at facial expressions. We followed the Plutchik's multidimensional model (1980) based on eight primary emotional dimensions, which are organized in polarity dyads (e.g., *ecstasy* as opposed to *grief*), declined into several combinations (e.g., *optimism* resulting from the combination of *anticipation* and *joy*), and nuanced according to their degree of intensity in tryads (e.g., *acceptance – trust – admiration*, from weakest to strongest) (Figure 1).

Anchored in biological and neurobiological grounds, the model includes 32 emotion labels that are said to be both discrete and gradual, insofar as intensity and polarity are considered to be central criteria for distinguishing between the emotions at stake. Three more recurrent emotions (*nervousness*, *disappointment* and *nostalgia*) emerged from the video data analysis and were therefore *a posteriori* added to the model. The closed list gives the advantage of providing a rich set of labels, which seems to be more accurate for the study of naturalistic data than other models strictly based on the 6 Ekman's basic emotions (Ekman, 1992). Moreover, all emotions in the Plutchik's model can be reduced to intrinsic positive and negative values. This is in line with the view that the broad bipolar dimensions of emotions (positive *vs.* negative valence) are the best (if not the only), most efficient way to distinguish between emotions from the face (Russell et al., 2003: 334).

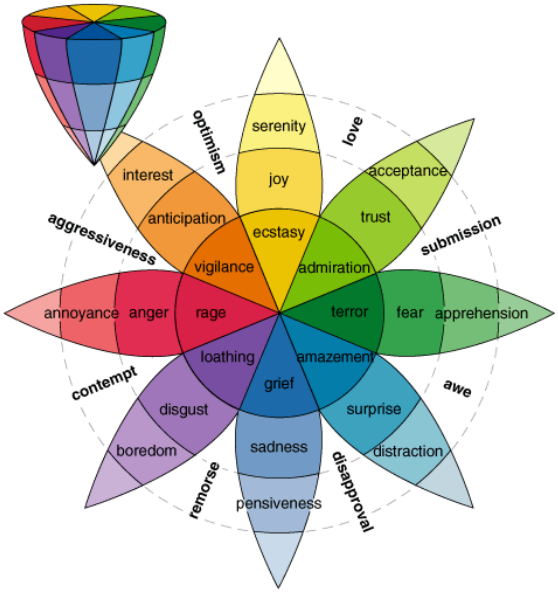


Figure 1. Plutchik's circumplex and wheel of emotions (reproduced from www.6seconds.org)

It is worth noting that the annotation of emotions was done without any access to the previously annotated physiological parameters. The boundaries of the emotion tags were determined on the basis of a holistic perception of the emotion expressed through the face, independently of the existing segmentation at the physiological level. Using emotion tags as a filter in the next step allowed for a bottom-up, relatively objective approach to the data. As a consequence, the boundaries of emotion tags do not correspond to physiological tags and emotions are mostly comprised of several physiological tags (e.g., one single emotion may include successive eye-closing moves and changes in gaze direction).

5.4 Contextual disambiguation of emotions

Emotional and attitudinal expression can be transmitted through multiple modes of communication (e.g., face, voice, words, and gestures) and may therefore result in complementary, redundant or even conflicting information (Gendron et al., 2012). According to that view, the semantic relation between emotions perceived from the face and their context of appearance (including the whole body, the linguistic and extralinguistic context) has been recognized to be either: (i) redundant: a similar emotion (even though not necessarily synchronous) is expressed from the face and from the linguistic context; (ii) complementary: the facial emotion is compatible with and adds some value to the linguistic information conveyed (e.g., modalization, emphasis, hedge, specification, elaboration, etc.); (iii) contradictory: the facial emotion is not compatible with the linguistic information conveyed; (iv) independent: there is no relation between the two modes, which fulfill their proper function in the language interac-

tion; (v) accordant: facial emotions are in accordance with information transmitted by the extralinguistic context at large (e.g., as reaction to external stimuli such as noises). This classification is mainly inspired by Colletta et al. (2009), following the pioneering classification in the field by Poggi and Magno Caldognetto (1996). These relationships were attributed to emotion tags in the data of one speaker only, namely Nadine (3 samples; 16 min. 24 sec.), with an additional focus on the relation between emotions and discourse markers in one of the three samples.

Discourse markers can briefly be defined as short linguistic items, which have no or little referential meaning and are not syntactically connected to their host clause. They serve pragmatic purposes by guiding the addressee in the decoding of the information conveyed: they "can connect to the speaker or addressee, provide information about the attitude of the communicator, introduce assumptions, or provide information about the context of interpretation" (Brinton, 2008: 5). Once transcribed and aligned to the sound signal, discourse markers were semi-automatically retrieved from speech and aligned to the video signal in the ELAN files.

6 Results

Preliminary results from the study indicate that, despite the highly idiosyncratic use of nonverbal resources, some inter- and intra-individual tendencies emerge.

6.1 Empathic ability

Results from the empathy test (F-IRI, see above) suggest that the healthy subjects obtain a relatively homogeneous global score of empathy (from 61% to 66%). This seems to posit that their empathic ability is relatively well preserved. Yet, a highly significant variability has been observed in the individual profiles with respect to the four subscales of empathy ($X^2 = 30.94$; $df = 9$; $p < 0.001$) (Table 4).

Speaker ID Code	*Pseudo*	*Fantasy (F)*	*Perspective-Taking (PT)*	*Empathic Concern (EC)*	*Personal Distress (PD)*	*F-IRI Score (%)*
ageBN1	Nadine	60 **[+1.93]**	57 **[-2.05]**	86 [-0.08]	51 [+0.77]	64
ageLL1	Louise	51 [+0.31]	80 [+0.23]	91 [+0.05]	43 [-0.69]	66
ageBM1	Anne-Marie	29 **[-2.4]**	94 **[+2.59]**	91 [+0.8]	31 **[-1.98]**	61
ageDA1	Albertine	46 [+0.1]	66 [-0.75]	77 [-0.77]	57 **[+1.9]**	61

Table 4. Subscales of empathy in percent [with standardized residuals]

These results partly confirm Gilet et al.'s findings (2013), who stressed *Fantasy* as being the most age-sensitive subscale. As a matter of fact, our data range from 60% in the youngest (Nadine, 75 y. old) to 46% in the oldest (Albertine, 85 y. old) for *Fantasy*, with the lowest score in Anne-Marie (29%). Moreover, *Empathic Concern* – which has been evidenced in several works to be genre-specific – shows a very high score in every participant (from 77% to 91%). Looking at intra-individual differences, even more striking results were found in Anne-Marie's profile (82 y. old) who seems to be more likely to experiment feelings of sympathy and compassion (EC: 91%), as well as to adopt the point of view of others (PT: 94%), than to transpose herself into fictional characters (F: 29%) or to feel concerned by stressful situations (PD: 31%). In addition, participants highly differ in their ability to cognitively adopt the point of view of somebody else (PT), ranging from the lowest ability in Nadine (57%) to the highest in Anne-Marie (94%).

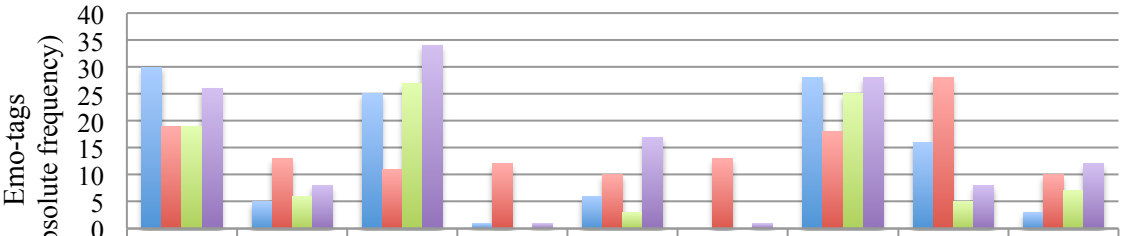
6.2 Emotional variety and richness

From the 581 emotions identified in the corpus data (including 8 undetermined emotions labeled as "Other"), 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). Only 9 categories of emotion for a total of 108 annotated emotions were counted in Albertine's speech samples [TTR = 0.08], while a wider emotional panel of facial expressions was observed in Louise's speech (20 types for a total of 161 emotions tagged in the samples [TTR = 0.124]). Anne-Marie and Nadine obtained intermediate scores, with respectively 14 types for 143 emotions tagged [TTR = 0.097] and 19 types for 169 emotions tagged [TTR = 0.112]. Even though these results were not statistically significant, we would like to highlight the fact that only 23 types among the 35 emotion tags available in the Template were identified as such, from a wider variety in Louise and Nadine (with more than 50%

of the tags used), to much less diversity in Anne-Marie (40% of the available tags used) and Albertine (only 26% of the available tags used). Interestingly, some emotions were quite infrequent in the data (e.g., only 1 to 3 cases of *amazement*, *boredom*, *contempt*, *ecstasy*, and *nervousness*), while others seem to be specific to one single participant. For instance, *fear* and *nostalgia* were mostly recognized from Nadine's face (with 12 and 13 out of 14 cases, respectively), while *attention* is mainly attributed to Albertine (with 9 out of 11 cases) and *disgust* to Anne-Marie (7 out of 9 cases). The most frequent emotions will be analyzed in the next section, by crossing emotional tags and physiological features.

6.3 Physiological patterning

Looking at physiological patterning from face and gaze expressions with regard to frequent emotions in the corpus, no clear physiological pattern could be considered specific to one emotion or another, neither to one speaker or another. However, some regularity was noticed from a closer investigation of the nine most frequent emotions (equal to or more than 10 occ. in the speech of at least one participant): *pensiveness* (99 occ.), *disapproval* (97 occ.), *annoyance* (94 occ.), *surprise* (57 occ.), *joy* (36 occ.), *trust* (32 occ.), *disappointment* (32 occ.), *fear* (14 occ.) and *nostalgia* (14 occ.) (see Figure 2).



	Annoyance	Disappointment	Disapproval	Fear	Joy	Nostalgia	Pensiveness	Surprise	Trust
ageBM1	30	5	25	1	6	0	28	16	3
ageBN1	19	13	11	12	10	13	18	28	10
ageDA1	19	6	27	0	3	0	25	5	7
ageLL1	26	8	34	1	17	1	28	8	12

Figure 2. Distribution of the most frequent emotions across participants

For instance, results for the *annoyance* emotion showed some variance between individuals when comparing cases of frowning and raising eyebrows. Notably, the recognition of this emotion in Anne-Marie' face [ageBM1] is mainly correlated with eyebrow frowning (55% of her eyebrow moves, with a positive standard residual of 3.21 for 15 cases), whereas the other three participants preferably raise their eyebrows (with 94% of eyebrows' raising moves in Nadine [ageBN1], 88% in Albertine [ageDA1], and 60% in Louise [ageLL1]). When looking at eyes' moves, a difference is also observed between participants, showing (i) much complex and repeated closing of the eyes in Louise (in one out of two cases with a positive standard residual of 1.51), (ii) a lower proportion of eyes' moves linked to *annoyance* in Anne-Marie's face, noticeable through the absence of any eyes' move in 44% of the cases (with a standard residual of 2.68), and (iii) a specific use of exaggerated opening of the eyes in Nadine (with a standard residual of 2.28), by comparison with the other participants and the other types of eyes' moves. In spite of these individual differences, overall results showed a well-balanced use of single and double closing of the eyes corresponding, respectively, to 20% and 16% of the 91 *annoyance* tags. To sum up, we can say that the expression of *annoyance* tends to be more idiosyncratic in Anne-Marie's face, as she mostly frowns without any other characteristics in closing or opening the eyes, by contrast to the other three participants who mainly raise their eyebrows either with many more eye-closings (cf. Louise) or with exaggerated opening of the eyes (cf. Nadine).

Again, we could reasonably expect a strong correlation between exaggerated openings of the eyes and eyebrow raisings, as a means to express *surprise*. But, even though this combination is relatively frequent to express *surprise* (23% of the cases), it is above all true for *fear* (57% of the cases). In a much less remarkable degree, it also applies to *disappointment* (9% of the cases), *annoyance* (7% of the cases), *nostalgia* (7% of the cases), *joy* (3% of the cases), *disapproval* (2% of the cases), and *pensiveness* (2% of the cases). Going a step further, it appeared that this combination of physiological pa-

rameters was specific to one single participant: among the 37 cases of "exaggerated-opening/eyebrow-raising" pattern, 32 were identified in Nadine's face, as conveying one of the above-mentioned emotional states. More in-depth and exhaustive analyses, which would embrace all the physiological parameters (including the whole body) and examine the way they combine in every participants, would undoubtedly help distinguish between individual and shared uses of gestural patterns with regard to the emotional and attitudinal states of people interacting in real-world settings.

6.4 Multimodality and the speech-gesture interface

A closer look at Nadine's speech, with a focus on the first 5 min. of interaction, allows for a better understanding of the relationship between (frequent and infrequent) perceived emotions, the facial expressions, and the linguistic context. Results give a first insight on the role of discourse markers in the multimodal expression of emotional states. From a context-sensitive angle, results showed that emotions usually appeared to be congruent with the contextual and linguistic information. Adding some compatible semantic or pragmatic value to the meaning conveyed in language use, facial emotions were mostly identified as being *complementary* (42 out of 74 emotions, see the red bars in Figure 3). Yet, facial emotions sometimes contradict the information conveyed by the context (in 14 out of the 74 cases). For instance, most of the time, the annotation of *joy* does not mirror the information expressed by Nadine and should be disambiguated thanks to the linguistic context.

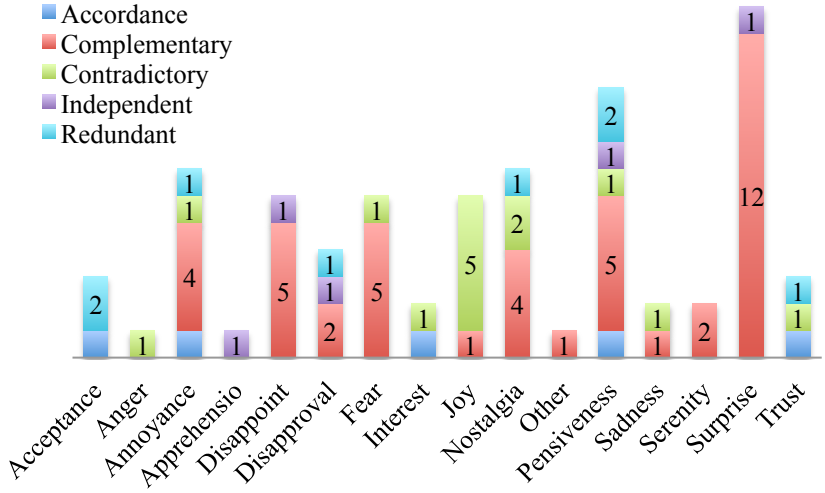


Figure 3. Semantic relationships between emotions perceived from the face and the contextual meaning (sample n°1 of Nadine's video data)

As the screen capture from the ELAN annotation file shows (Figure 4), the emotion perceived – independently of any contextual cues at the first step of analysis – is *joy*, but what the speaker is actually saying concerns a painful episode in her childhood (*c'était* ***un peu*** *je,une* ***quoi hein*** *j'ai été* ***un peu*** *malheureuse* ***là*** 'I was **a little bit** too young **well you know** I have been **quite** unhappy **there**').

The further question is to explore how far the global emotional and attitudinal state can be inferred from speech and from nonverbal resources. Our hypothesis is that Nadine is smiling here to mitigate the pain she is remembering (another interpretation would be that she is smiling because of being embarrassed to talk about an intimate and painful experience). Facial displays would then be redundant with the modal marker *un peu* 'a bit', notably repeated once. Concerning the function of discourse markers in the synchronous co-text of the emotion tag (*viz. quoi* 'well', *hein* 'he', and *là* 'there'), their intersubjective function (Kärkkäinen, 2006) could be seen as stressing the need to share the speaker's painful experience with the interlocutor or as reassuring that full attention is paid to what she says. As Russell et al. emphasized (2003: 242), smiles can be spontaneous "reliable signs of positive feelings toward a specific receiver" (expressive function), but they can also be produced in a controlled manner as "volitional smiles" which seek appeasement or help in the addressee (interactive function).

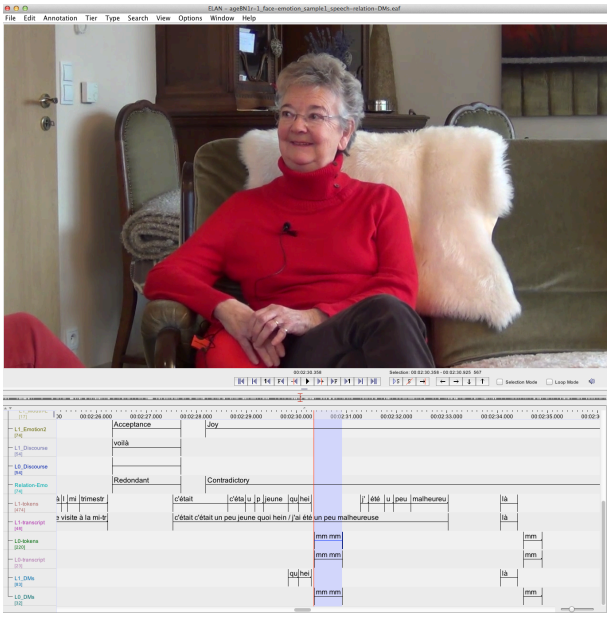


Figure 4. Example of contradiction in the emotion recognized from the face and the linguistic context.

7 Conclusion

Nonverbal language resources are recognized as a major channel of emotional expressivity and interactivity in the communicating person. But, due to their ambiguous and complex structure, emotional states are extremely challengeable to detect, even more in the natural context of language production (Douglas-Cowie et al., 2003: 36-38). Starting from the annotation of facial expressions and emotion recognition in authentic video data, it has been evidenced (however quite unsurprisingly) that the visual mode, if taken alone, was not sufficient to understand what kind of information the speaker is actually transmitting to the interlocutor. Rather, as the linguistic level of communication often needs to be contextualized, the nonverbal level of communication also needs more "words" in order to be interpreted in accordance with the speaker's intention.

This study represents the very fist step of the CorpAGEst research project, which aims at developing a multimodal model for the annotation of pragmatic functions in speech and gesture, as a means to detect any change with advancing age in the pragmatic competence of very old people. The study has given a first insight into what we can infer from emotional and attitudinal expressions of very old healthy people by means of "naturalistic" corpus data. Even though providing only part of the big picture, the approach, we assume, allowed for a better understanding of the way older people show and express their emotions in real language use. It seems obvious that the pragmatic part of language communication is not of little interest in the field of aging research, and would need further investigation moving from experiments in the laboratory towards empirical studies "into the wild".

Acknowledgment

The research leading to these results has received funding from the European Union Seventh Framework Programme ([FP7/2007-2013]) under grant agreement n° [PIEF-GA-2012-328282]. The authors are grateful to the two anonymous reviewers and L. Crible for their useful review of the paper. C. T. Bolly would like to warmly thank all the participants in the CorpAGEst project, who will certainly recognize themselves. She also thanks A. Sáfár, A. Thomas and D. Belin, for their help in the annotation procedure, and D. Boutet, A. Lacheret and L. Degand, for numerous inspiring discussions and their confidence in the project.

References

Allwood, J., Cerrato, L., Jokinen, K., Navarretta, C. and Paggio, P. 2007. The MUMIN coding scheme for the annotation of feedback, turn management and sequencing phenomena. *Language Resources and Evaluation*, 41(3-4), 273-287.

Bailey, Phoebe E. and Julie D. Henry. 2008. Growing less empathic with age: Disinhibition of the self-perspective. *Journal of Gerontology: Psychological Sciences*, 63B(4): 219–226.

Bolly, Catherine T. 2013. *CorpAGEst. Multimodal corpus for the elderly's language*. Louvain-la-Neuve, Paris: Université catholique de Louvain (Valibel - Discours et Variation) and CNRS (UMR 7023 Structures Formelles du Langage), <http://corpagest.org>.

Brinton, Laurel J. 2008. *The comment clause in English: Syntactic origins and pragmatic development*. Cambridge : Cambrige University Press.

Colletta, Jean-Marc, Kunene, Ramona N., Venouil, Aurélie, Kaufmann, Virginie and Simon, Jean-Pascal. 2009. Multi-track annotation of child language and gestures. In Michael Kipp, Jean-Claude Martin, Patrizia Paggio and Dirk Heylen (eds.), *Multimodal corpora* (*Lecture Notes in Computer Science* 5509). Berlin, Heidelberg: Springer, 54–72.

Davis, Mark H. 1994. *Empathy: A Social Psychological Approach*. Boulder: Westview Press.

Douglas-Cowie, Ellen, Campbell, Nick, Cowie, Roddy, and Peter Roach. 2003. Emotional speech: Towards a new generation of databases. *Speech Communication*, 40: 33–60.

Du Bois, John W. and Elise Kärkkäinen 2012. Taking a stance on emotion: Affect, sequence, and intersubjectivity in dialogic interaction. *Text & Talk*, 32(4): 433–451.

Eisenberg, Nancy and Richard A. Fabes. 1990. Empathy: Conceptualization, measurement, and relation to prosocial behavior. *Motivation and Emotion*, 14(2): 131–149.

Eisenberg, Nancy, Shea, Cindy L., Carlo, Gustavo, and George P. Knight 2014. Empathy-related responding and cognition: A "chicken and the egg" dilemma. *Handbook of moral behavior and development: Research*, 2: 63–88.

Ekman, Paul. 1992. An argument for basic emotions. *Cognition and Emotion*, 6(3/4): 169–200.

Englebretson, Robert (ed.). 2007. *Stancetaking in discourse: Subjectivity, evaluation, interaction* (Pragmatics & Beyond New Series 164). Amsterdam, Philadelphia: John Benjamins Publishing.

Gendron, Maria, Lindquist, Kristen, Barsalou, Lawrence, and Lisa F. Barrett. 2012. Emotion words shape emotion percepts. *Emotion*, 12(2): 314–325.

Gilet, Anne-Laure, Mella, Nathalie, Studer, Joseph, Grühn, Daniel and Gisela Labouvie-Vief. 2013. Assessing dispositional empathy in adults: A French validation of the Interpersonal Reactivity Index (IRI). *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement*, 45(1): 42–48.

Harwood, Jake. 2007. *Understanding communication and aging: Developing knowledge and awareness*. Thousand Oaks: Sage.

James, Lori E., Burke, Deborah M., Austin, Ayda and Erika Hulme. 1998. Production and perception of "verbosity" in younger and older adults. *Psychology and Aging*, 13(3): 353–367.

Kärkkäinen, Elise. 2006. Stance taking in conversation: From subjectivity to intersubjectivity. *Text & Talk*, 26(6): 699–731.

Kasper, Gabriele and Kenneth R. Rose. 2002. *Pragmatic development in a second language*. Mahwah: Blackwell (Also *Language Learning : Supplement 1*, 52).

Martin, Jean-Claude C., Niewiadomski, Radoslaw, Devillers, Laurence, Buisine, Stéphanie and Catherine Pelachaud. 2006. Multimodal complex emotions: Gesture expressivity and blended facial expressions. *International Journal of Humanoid Robotics*, 3(3): 269–291.

Müller, Cornelia, Jana Bressem and Silva H. Ladewig. 2013. Towards a grammar of gestures: A form-based view. In Cornelia Müller, Alan Cienki, Ellen Fricke, Silva H. Ladewig, David McNeill and Sedinha Teßendorf (eds.), *Body – language – communication: An international handbook on multimodality in human interaction. Vol. 1*. Berlin, Boston: Mouton de Gruyter, 707–733.

Nasreddine, Ziad S., Phillips, Natalie, Bédirian, Valérie, Charbonneau, Simon, Whitehead, Victor, Collin, Isabelle, Cummings, Jeffrey L. and Howard Chertkow. 2005. The Montreal Cognitive Assessment (MoCA): A brief screening tool for Mild Cognitive Impairment. *Journal of the American Geriatrics Society*, 53: 695–699, <http://www.mocatest.org/default.asp>.

Niedenthal, Paula M. 2007. Embodying emotion. *Science*, 316(5827): 1002–1005.

Niedenthal, Paula M., Barsalou, Lawrence W., Ric, François and Silvia Krauth-Gruber. 2005. Embodiment in the acquisition and use of emotion knowledge. In Lisa Feldman Barrett, Paula M. Niedenthal and Piotr Winkielman (eds.), *Emotion and consciousness*. New York: The Guilford Press, 21–50.

Plutchik, Robert. 1980. *Theories of Emotion. Vol. 1: Emotion: Theory, Research, and Experience*. New York: Academic.

Poggi, Isabella and Magno Caldognetto, Emanuela. 1996. A score for the analysis of gestures in multimodal communication. In L. Messing (ed.), *Proceedings of the Workshop on the Integration of Gesture and Language in Speech, Applied Science and Engineering Laboratories*, Newark and Wilmington, Delaware: 235–244.

Russell, James A., Jo-Anne Bachorowski and José-Miguel Fernández-Dols. 2003. Facial and vocal expressions of emotion. *Annual review of psychology*, 54(1): 329–349.

Wittenburg, Peter, Brugman, Hennie, Russel, Albert, Klassmann, Alex and Hans Sloetjes. 2006. Elan: A professional framework for multimodality research. In *Proceedings of LREC 2006*, Fifth International Conference on Language Resources and Evaluation. Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands, <http://tla.mpi.nl/tools/tla-tools/elan/>.