Visualizing the Importance of Medical Recommendations with Conversational Agents

Gersende Georg[1,2], Marc Cavazza[3], and Catherine Pelachaud[4,5]

[1] Centre des Cordeliers UMRS 872 Eq. 20, Paris, France
[2] French National Authority for Health (HAS), Saint-Denis La Plaine, France
[3] School of Computing, University of Teesside, Middlesbrough, United Kingdom
[4] University of Paris 8
[5] INRIA Rocquencourt, France
gersende.georg@spim.jussieu.fr, m.o.cavazza@tees.ac.uk,
catherine.pelachaud@inria.fr

Abstract. Embodied Conversational Agents (ECA) have the potential to bring to life many kinds of information, and in particular textual contents. In this paper, we present a prototype that helps visualizing the relative importance of sentences extracted from medical texts (clinical guidelines aimed at physicians). We propose to map rhetorical structures automatically recognized in the documents to a set of communicative acts controlling the expression of the ECA. As a consequence, the ECA will dramatize a sentence to reflect its perceived importance and degree of recommendation (advice, requirement, open proposal, etc). This prototype is constituted of three sub-systems: i) a text analysis module, ii) an ECA and iii) a mapping module which converts rhetorical structures produced by the text analysis module into nonverbal behaviors driving the ECA animation. This system could help authors of medical texts to reflect on the potential impact of the writing style they have adopted. The use of ECA reintroduces an affective element which won't be captured by other methods for analyzing document style.

Keywords: Embodied Conversational Agents, Emotional Natural Language Processing, Medical Informatics.

1 Introduction and Rationale

ECAs have been demonstrated to bring added value to many applications for which a more human-like presentation [1,2] is beneficial, including assistance, help and guidance [3,4]. In this paper, we investigate the use of ECAs to visualize the importance of specific sentences within medical documents. The document we studied are Clinical Guidelines, which are normative texts, produced by various Health authorities, promoting best practice in Medicine based on the concept of *evidence-based medicine* [5]. They are written by expert physicians and aimed at physicians, for instance General Practitioners. Clinical guidelines are based on the notion of recommendation as their elementary unit. These can be characterized linguistically through specific

syntactic constructs which manifest a rhetorical intention. For instance, "*In case of extension to pedicle lymph nodes, if surgical accessibility falls into Class I, surgery cannot be contraindicated, but this decision should nevertheless be part of a multi-disciplinary consultation*". Clinical guidelines are produced by expert committees through a complex process of consensus building, as committee members assess the style of initial formulations. Guidelines' authors need to be able to assess the potential perception by their readers of the strength of recommendations they have been writing, to anticipate the impact of the specific recommendations they contain as a function of the style used. Expressive communication is one mode of visualizing a recommendation's strength in which various dimensions could be combined seamlessly through multimodal channels, e.g. facial expressions and/or gestures synchronized to the utterance. This importance manifests itself, within certain constraints to be discussed below, through the choice of syntax and vocabulary, which can be mapped to certain rhetorical structures, from advice to orders.

2   System Overview and Architecture

This work is based on experiments carried out with our prototype, which presents itself as an ECA interface "reading aloud" specific recommendations selected from a clinical guideline. It is actually constituted of three sub-systems: i) G-DEE, a document engineering environment [6] which performs an automatic identification of recommendations in a Guideline, ii) Greta, an ECA system [7] and iii) a mapping module which converts rhetorical structures produced by G-DEE into the communicative act format used by Greta (a mark-up language known as APML [8]). The system operates as follows. Firstly, G-DEE is run offline to analyze the clinical guidelines as a whole. It produces a document in which all recommendations are identified through a set of specific mark-ups for their operators and the contents they apply to (referred to as the scopes of the operator).

A marked-up recommendation appears as highlighted text in the system interface (Fig. 1). This text fragment can be selected interactively, which triggers the generation of an APML file animating Greta on that sentence (the generation uses an XSLT conversion module). In this process, tags on communicative acts linked to recommendation strength have been added to the text automatically. Finally, Greta processes this APML file and utters the corresponding recommendation, displaying appropriate nonverbal behavior, which reflects the importance of the recommendation and places emphasis on relevant scopes. In this way, the actual strength of the recommendation and its potential impact can be visualized.

2.1   The Document Engineering Environment

Clinical guidelines belong to the generic category of normative texts, to which much research has been dedicated. These texts are naturally structured through the occurrence of specific linguistic expressions, known as "deontic operators" [9], which characterize the linguistic expression of recommendations. These operators manifest themselves (in French) through such verbs as "*pouvoir*" ("to be allowed to or may"), "*devoir*" ("should or ought to"), "*interdire*" ("to forbid").
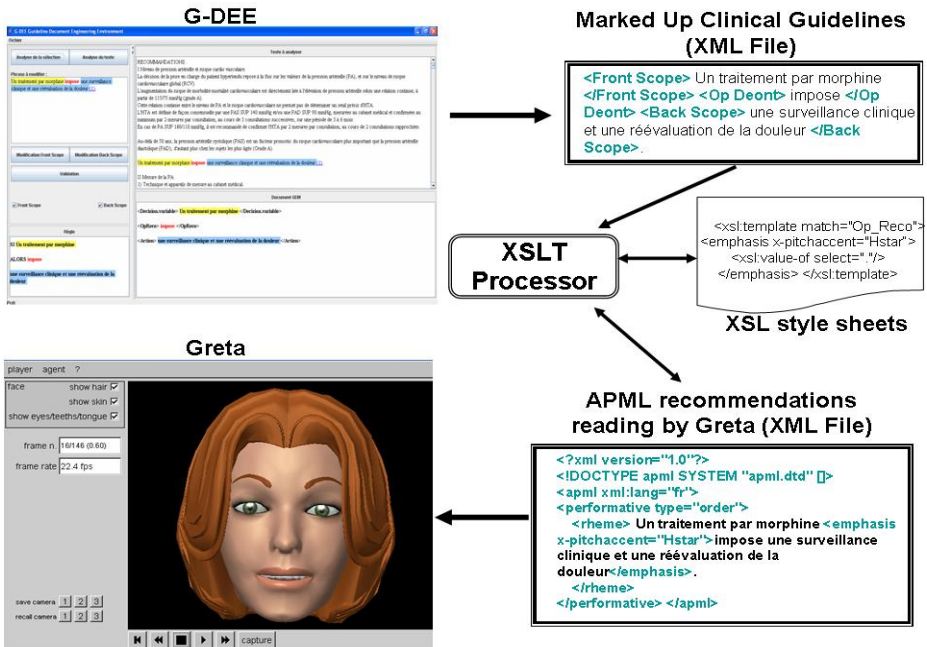
Fig. 1. System architecture

G-DEE [6] is a document analysis environment dedicated to the study of clinical guidelines[1]. It automatically detects recommendations using shallow natural language processing (NLP) techniques which recognize deontic operators in medical texts such as "authorize", "forbid", "ought to" [9]. Using specific grammars embedded in Finite-State Automata (FSA), G-DEE parses the whole document, identifying deontic operators and the text segments they apply to. The final output of this process is to structure the document around recommendations as shown below.

```
<Front-Scope> A treatment with morphine </Front-Scope> <Op_Reco>
imposes </Op_Reco> <Back-Scope> a clinical supervision and a reas-
sessment of pain </Back-Scope>²
```

Fig. 2. Example of marked-up recommendation

Let us now consider the different aspects that determine the strength and emphasis of a recommendation. Firstly, deontic operators fall within the broad categories of permission, obligation or interdiction and can be classified according to their

---

[1] G-DEE is actually deployed at the French National Authority for Health (HAS) to assist the early stages of Guideline development and has been used in the analysis of over 30 Guidelines over the past 14 months.

[2] This is an English translation of the original French Guideline (the translation does not affect the recommendation structure).

"strength" within these categories (e.g. strength of interdiction). Strength is not just an issue of vocabulary, but relates also to syntactic constructs (which have been uncovered in the process of deontic operator extraction). In other words, that a specific drug "should not be used" is stronger than it being "not recommended". It can also be noted that this concept bears some similarity with the illocutionary strength of communicative acts (which was one of our initial inspirations for this project). The reason why deontic operators play such a dominant role in the expression of the recommendations strength is to be found in the authoring rules which are enforced in several governmental agencies (including the French National Authority for Health (HAS)). In that sense, other linguistic phenomena are prevented from playing a role in the expression of recommendations' strengths: implicit nuances are discouraged, few adverbs are actually employed, and there is little use, if any, of affective categories[3].

2.2 The Greta Platform

The Greta agent [7] used in these experiments is a platform developed for the purpose of research in non-verbal behavior. It includes an animation system with facial parameters supporting detailed expressive animations synchronized to a Text-To-Speech (TTS) system. Greta's animations are controlled using instructions in the APML language [8], which is a markup language embedding communicative functions as well as the text fragments corresponding to the utterance to be passed to the TTS system. The baseline system incorporates a set of pre-defined communicative acts for which all facial animation parameters have been preset. The mapping between communicative acts and nonverbal behaviors is based on studies that have reported the communicative value of given signals (e.g. frowning is linked to goal obtrusiveness [10, 11]) and on video corpus annotation [12]. Communicative acts are grouped into classes depending on the information they convey [12]. These categories cover communicative acts related to the agent's beliefs (e.g. confidence in what is being said), to the goal of the agent (emphasizing the focus of its utterance), and finally its emotional state. APML also contains markers linked to the intonational structure of the sentences uttered by the agent. The traditional theme/*rheme* distinction can be thought of as a distinction between new and old information [13]. The pitch accents and boundary tones follow the ToBI annotation [14].

In particular, a previous study [15] has investigated the relations between performatives (communicative acts), such as: order, suggest, propose, warn, refuse, etc., and facial expressions. Three main classes of performatives were considered: *request*, *inform* and *question*. Within a given class, performatives (either from the request, inform or question class) share a common goal: respectively to elicit an action, to give information or to ask for an information.

Let us look more in detail performatives of the class request. They have been characterized along three dimensions: whom the action is requested from, how certain one is of the information provided and the power relationship [15]. The first dimension allows us to differentiate performatives of *advice* from those of *command*. We advise someone to perform an action that we believe could be beneficial to herself but we

---

[3] This may not be true of all types of Medical documents, yet it is a characteristic of clinical guidelines (easily understandable if one considers such an authoring rule as the one discouraging the use of adverbs).

command a person to carry out an action for our own benefit. Along the second dimension, we may suggest something when we are unsure. Power relations relate to our potential to coerce the person from whom an action is requested. As a result, we order or suggest depending on those power relations.

Based on the representation of the performative along these three dimensions, we have proposed a mapping between each of these dimensions and facial expressions [12]. That is, the facial expression associated to a given performative is obtained by combining the expressions arising from each dimension. Being certain or uncertain can be shown on the eyebrow region: one frowns when being very much certain of what one says, but raises one's eyebrow if uncertain [10, 11]. Head orientations (such as head kept straight up or tilted aside) can be a sign of a power relation: submissiveness is often shown by displaying the neck [16] while dominance is characterized by a straight up head). Performatives contain an intrinsic emotional factor [12]. When giving an order, one can potentially show anger if the requested action is not performed. On the other hand, when imploring one can display sadness.

Thus facial expressions of a performative may encompass information related to the potential emotional state of the agent, the dominance/submissive relationship with the interlocutor, as well as the degree of certainty of the information conveyed. These dimensions can be mapped to the dimensions characterizing recommendations. We introduce this mapping later in the paper (see section 5).

Let us now explain how our system computes the agent's animation from its intended utterance. Having established the dimensions along which performatives can be characterized and having specified the link between these dimensions and facial expressions, we have elaborated the mapping for several performatives (in particular of the class Request and of the class Inform) into facial expressions. These definitions are stored in a lexicon.

The system takes as input the text to be uttered, augmented with APML tags specifying the communicative intentions of the agent. As a first step, the selected text is sent to the TTS system, which provides the list of phonemes allowing the computation of the lip movement, as well as giving timing information supporting the synchronization of verbal and nonverbal streams. The next step consists of converting communicative intentions into a set of nonverbal behaviors by looking up in the lexicon mentioned above. Finally, the set of nonverbal behaviors are matched to the verbal stream to ensure synchrony across modalities. The last step is to compute the values of the facial animation parameters over time and to play the animation.

3   Related Work

Our work focuses on the use of conversational agents for visualizing rhetorical structures extracted from medical texts. The majority of the work on ECAs has focused on speech acts and emotional communication. There has not been much emphasis on rhetorical structures, to the exception of the T2D system [17], which used textual input decomposed into segments linked by rhetorical discourse relations to generate dialogues between agents.

In the context of medical applications, the emphasis has been on facilitating the dialog between doctors and patients, the former having often been criticized for their

lack of empathy and ability to explain patients the course of their disease [18]. ECAs can thus play an important role in medical applications, as shown by Marsella et al. [19] with their system "Carmen's Bright IDEAS". MagiCster is another system whose applications have included an advice-giving dialog in a medical domain [20]. Here, the agent plays the role of a doctor giving the patient information about her disease. De Rosis et al. [21] have described a conversational agent advising on eating disorders. The aim of the dialog between the ECA and the user is to persuade them to improve their eating habits [22]. The use of ECAs for patients education has been shown to result in higher satisfaction rates [23]. ECAs have thus been demonstrated to bring added value to many applications for which a more human-like presentation [2] is beneficial, including assistance, help and guidance [24].

However, no work to date has investigated the use of ECAs to explore medical text perception by physicians themselves. Here, the equivalent to emotional content in Guidelines' recommendations would correspond to authority and responsibility, and the rhetorical expression of recommendations corresponds to communicative acts.

4   Identifying the Rhetorical Strength of Recommendations

The first step consisted in devising a scale to rate the strength of clinical recommendations. We carried out a study involving 14 medical experts from Inserm (French National Institute for Health) and HAS, who have been involved in the elaboration of clinical guidelines, to determine the level of consensus between experts about the strength of a given recommendation[4]. These experts rated the strength of 37 prototypical recommendations extracted from several hundreds recommendations occurring in recent clinical guidelines published by the HAS. They ranked the strength of each recommendation according to a predefined 6-point scale defined as follows:

```
CAT1- well-identified best practice, which is compulsory
CAT2- practice well adapted to the clinical situation that
presents demonstrable benefits
CAT3- accepted practice which can be advised, or to be considered
CAT4- practice left to the discretion of the physician
CAT5- statement explaining or justifying a course of action
CAT6- a useful information item
```

Fig. 3. Categories for evaluating the strength of recommendations

For each deontic verb, used in recommendations, we were subsequently able to associate a numerical score quantifying its rhetoric strength depending on the previous analysis of Guidelines corpora[5] and on explicit rules on guidelines vocabulary and terminology (including verbs for recommendations) mentioned in HAS internal documents on Guidelines' authoring methodology. This will serve as a starting point

---

[4] This represents a very significant sample, considering the total number of such experts within any given European country.

[5] Guidelines may contain explicit gradings on the level of certainty and strength for their recommendations. These make possible to analyse statistically the occurrence of deontic verbs as a function of these gradings.

to map the rhetorical strength of deontic expressions onto the emotional categories of Greta. It is an empirical solution, based on existing corpora, to the problem of relating grades of recommendations (as described by the experts) to the linguistic formulation of the strength of a recommendation. It also supports the extension of the set of performatives used by Greta as described in the next section.

5   Mapping Rhetorical Structures onto Multimodal Communicative Acts

The process by which the rhetorical strength of textual recommendations will be visualized rests on a mapping from deontic operators onto multimodal communicative acts. These can be described as the dynamic expression of traditional communicative acts (order, advice, propose, etc.), using communicative parameters and dynamic animation of non-verbal behavior, in particular facial expressions. The rationale for such a mapping derives from the pre-existing commonality between certain deontic operators used in the description of recommendations and the set of primitive communicative acts originally embedded in the APML control language (which contains communicative acts such as *advice*). This mapping attempts to generalize these commonalities by relating deontic operators to communicative acts but also their perceived strength to the *rheme* part of APML expressions. The *Rheme* corresponds to novel information that is brought into the conversation, and most of the nonverbal behavior occurs within the *rheme* [25].

5.1   Definition of Specific Expressions

As introduced previously (Section 2.2), performatives are described along three dimensions. We have elaborated the mapping between the six categories of the recommendations' strength scale and the performatives by looking at the common values for these 3 dimensions[6]. In particular, communicative acts are defined as a pair whose first element corresponds to the meaning of the communicative act. It is specified by APML tags, while the second element represents the signals that convey this meaning. Let us see the mapping for each category:

- CAT1: the practice is compulsory. It corresponds to an order, a request of carrying out an action. In APML, the performative "order" is described by a frown (sign of anger), head up and look down. Recommendations may not encapsulate social relationship thus any behavior linked to this aspect have been eliminated (in this case, head up and look down). To highlight the importance of the recommendation, the emphasis tag is added. It is shown through head nods.
- CAT2: the practice is a strong recommendation but not as much as CAT1. As for CAT1, no notion regarding social relationship is needed here. Thus,

---

[6] In the current version of our system, the mapping is done with the performatives that are already defined within the APML language. However, we aim to extend the list of performative to improve coverage.

CAT1 is represented by a less intense frown. The APML tag is "order" but with lesser intensity expression. No emphasis tag is added.

- CAT3: This type of recommendation can be considered as an advice. It is displayed using the eyebrow shape to convey the performative 'advice': slight rising of the eyebrows.
- CAT4: The recommendation mentions a possible course of action as a suggestion. Suggestion is characterized by raised eyebrows and tilted head.
- CAT5: the recommendation is used to inform physicians but with a certain emphasis. It is translated by looking at one's addressee and performing a head nod on the emphasized word. Two tags are used, one performative "inform" and the emphasis tag.
- CAT6: the physician is simply informed: that is displayed through gaze behavior, namely looking at the addressee.

The APML tags are added to the text automatically using the following rules:

- Performative tags: they mark the whole sentence.
- Emphasis tags: emphasis is linked to the intonational structure of the sentence. As new information should receive the most attention, emphasis is marked. We use the *rheme*/theme structure. New information is part of the *rheme*. The emphasis tag is set around the deontic verb while the *rheme* tag goes around the sentence (as the performative).
- Certainty tags: to mark negation contained in deontic verb (e.g. French *interdire* (to forbid) of CAT1; *ne pas prescrire* (not to prescribe) of CAT2), we use the tag "certainly-not". It spans the same text as the performative. Certainly-not is shown by a frown.

Let us see in the next two sections how this automatic transformation between G-DEE and APML happens.

5.2   XSL-Based Transformations and APML Generation

Because Greta's input format for multimodal communicative acts (APML) is XML-based, it is a natural choice, from a technical standpoint, to use XSLT transformations to generate APML from the deontic operator structure based itself on XML. The XSLT processor already integrated in G-DEE has been extended to support the generation of APML formulas. XSL style sheets need to define the mapping between various categories, in particular the communicative acts defined in the APML DTD. This includes the transfer of sentence fragments belonging to the recommendation from the marked-up guideline to Greta's TTS system. Firstly, the XML file generated by the text analysis module contains marked-up recommendations which are the starting point for the further XSL transformations. An example of such a marked-up recommendation is presented in Figure 2. We thus defined XSL style sheets to transform automatically this file into an APML file controlling Greta. In this example, the presence of a deontic operator of the type CAT1 justifies the presence of an <emphasis> mark-up in APML. In turn, the type of deontic operator determines the type of communicative act. This transformation is based on the conceptual mapping between the

```
<apml>
    <performative type="order"> <rheme>
    A treatment with morphine
    <emphasis> imposes
    </emphasis>
    a clinical supervision and a reassessment of pain levels
    </rheme> </performative>
</apml>
```

Fig. 4. APML expression resulting from the XSL transformation to map recommendations to speech acts

deontic operators recognized by G-DEE and the set of previously predefined communicative acts in Greta (see Fig. 4 for an example of transformation).

This conceptual mapping essentially establishes a correspondence between the expressivity of the written text and that of the recommendation pronounced by Greta. XSL style sheets are specific to each category of the recommendations strength. G-DEE characterizes which kind of deontic verb the recommendation contains using the mapping described in the section above. In some cases there exists a direct, one-to-one mapping such as with the "advice" operator, or, to some extent the "propose" one (as they exist in both mark-up systems). The deontic operators which are meant to "suggest" (e.g. French *être laissé à* (to be left to) / *pourrait* (may) of CAT4) can be mapped to the *suggest* communicative act. Strong negative recommendations (such as the French "*déconseiller*" (Advice not to, discourage)) have been mapped to the *refuse* communicative act in APML, while for mild negative recommendations we use the disagree communicative act.

To intensify negative deontic verbs the communicative function Certainly-not is added to the text. Certainly-not is part of the cluster 'certainty' and is marked by a frown. Frown is attached to a negative signal as it is often linked to goal obstruction [10]. Finally, for CAT1 and CAT5 cases, the emphasis tag is added around the deontic verb.

Table 1. Excerpt of the mapping table between deontic verbs and APML performative types

Deontic verb	APML
CAT1 – APML: order	
ordonner (to order) / *impose* (to impose) / *devra associer (*will have to associate)	Performative "order"+emphasis/*rheme*
interdire (to forbid)	Performative "order" + certainty "certainly_not" +emphasis/*rheme*
CAT2 – APML: recommend	
recommander (to recommend) / *prescrire* (prescribe) / *contre-indiquer* (to counterindicate)	Performative "recommend"
déconseiller (to advise not to) / *ne pas recommander* (not to recommend) / *ne pas prescrire* (not to prescribe)	Performative "recommend" + certainty "certainly_not"

Overall, the existing APML performative set of the communicative acts can support a consistent mapping: the only limitation lies in the lack of explicit nuances between some forms of positive recommendations, which should be the object of further work can be compensated in part using redundancy in behaviors, that is where meanings are conveyed over different modalities (e.g. raised eyebrow and head nod). Next section illustrates this mapping through an example.

6   Example Results

The XSLT transformations generate the most appropriate communicative act for the deontic verb considered in the recommendation. The following screenshots of Greta enable to visualize the differences between expressions according to performative type which are mainly focus on the eyebrow and the head nod (although differences can only be really seen on the dynamic animation).

The following examples correspond to two of the six categories defined for the strength scale. For the category 2, the dedicated style sheet enables to transform a marked-up recommendation to an APML format (Fig. 5) that supports the mapping of the French "*il est recommandé*" ("it is recommended") deontic verb to the *recommend* performative type, defined in section 5.1.

```
<apml>
      <performative type="recommend"> <rheme>
      <emphasis> It is recommended
       </emphasis> to perform a venous Doppler examination as
       part of the management of all patients with ulcers of the
       lower limbs. </rheme> </performative>
</apml>
```

Fig. 5. The resulting APML file corresponds to a recommend performative type

Fig. 6. The resulting expression of Greta corresponding to a *recommend* performative type

Fig. 7. The resulting APML file corresponding to a *suggest* performative type

The corresponding expression of Greta consists of a recommendation with an emphasis on the deontic verb "*il est recommandé*" (it is recommended) and a frown with a head nod (Fig. 6), while the suggest conversational act (Fig. 7) is associated to a slight raising of the eyebrows and a head tilt.

7   An Application to Consensus Judgments of Recommendations' Strength

We conducted an evaluation of the system with six medical experts drawn from the group of the fourteen experts that have participated in the definition of recommendations' strengths. For this evaluation, we devised a test suite of nine prototypical recommendations, representative of the whole spectrum of recommendations' strength. The main objective of this evaluation consists of determining whether Greta improves the perception of recommendations strengths, for instance by generating a stronger consensus or helping to disambiguate between neighboring categories. Since standard deviation is a simple and well described measure of consensus [26], we analyzed its value throughout our experiments. Each of the six medical experts rated the recommendations strength first from reading them and secondly from seeing them presented by Greta. The average strength score as well as the standard deviation were calculated for each recommendation, without and with Greta (Fig. 8).

The effect of Greta appears to vary greatly depending on the category in which a recommendation has been indexed. However, one of the main problem in guideline authoring is the level of consensus. In that sense, figures obtained from isolated users do not reflect the actual dynamics of a working group. A lack of consensus (measured e.g. through a high value for standard deviation) can have significant implications during a face-to-face consensus meeting and this is why improving consensus is a major objective in the process of Guidelines' elaboration.
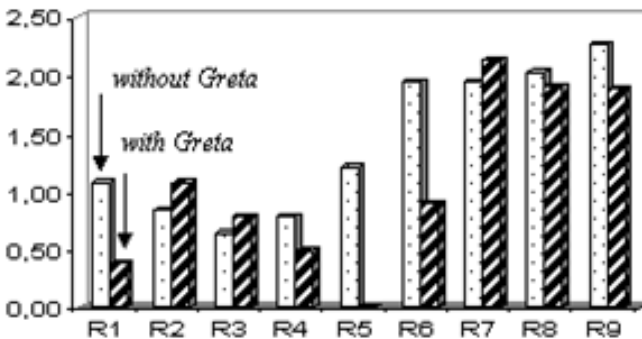
Fig. 8. Impact of Greta on the standard deviation of experts' judgments of recommendations' strength: this impact is stronger for "borderline" categories (R4, R5 and R6), where consensus is most difficult to reach

Most importantly, we observed a very significant effect of Greta on the standard deviation of recommendations' strength, and that effect is more pronounced, and highly significant, for intermediate categories, such as CAT3 (R4), CAT4 (R5) and CAT5 (R6), which are known to be the object of significant debate in working groups. This effect is also remarkable for the strongest type of recommendations CAT1 (R1), for which difficulties in reaching consensus have been often reported. The decrease in standard deviation can be interpreted as a better consensus between experts: this suggests that using Greta would potentially improve the efficiency of a working group. To a large extent, the system presented here can restore the link between the wording of a recommendation and its intended impact on the reader. As a tool to assist the authoring of guidelines, it should help selecting the appropriate level of emphasis required as well as balancing the importance of recommendations across the document as a whole.

8   Conclusions

ECAs have been mostly described in dialogue and interface applications, with little work on their use in the visualization of textual properties. The most natural applications in that area would be to dramatize the affective aspects of the underlying text. Yet, we suggest that dramatization, as provided by the non-verbal behavior of ECAs, can also be of use to visualize the rhetorical content of texts, and that this can have practical applications as well. The principle behind this approach is that communicative acts which are used to define ECA non-verbal behavior naturally overlap with some of the rhetorical intentions embedded in texts. The mapping between these two aspects may not be trivial, and in these first experiments we had to provide an empirical solution based on domain expertise.

These first results are very encouraging and future work will extend this approach using more sophisticated expressive mechanisms such as gestures, possibly also relating non-verbal behavior to further contents of the deontic operators, such as the recommended course of action.

References

1. Hoorn, J., Konijn, E.: Personification: Crossover between Metaphor and Fictional Character in Computer Mediated Communication. In: The annual meeting of the International Communication Association, San Diego, CA (2003)
2. Nass, C., Steuer, J., Tauber, E.: Computers are Social Actors. In: Proceedings of the SIGCHI conference on Human factors in computing systems: celebrating interdependence, Boston, Massachusetts, United States, pp. 72–78 (1994)

3. Abbattista, F., Lops, P., Semeraro, G., Andersen, V., Andersen, H.: Evaluating virtual agents for e-commerce. In: Falcone, R., Barber, S., Korba, L., Singh, M.P. (eds.) AAMAS 2002. LNCS (LNAI), vol. 2631. Springer, Heidelberg (2003)
4. Allbeck, J., Badler, N.: Toward Representing Agent Behaviors Modified by Personality and Emotion. In: Falcone, R., Barber, S., Korba, L., Singh, M.P. (eds.) AAMAS 2002. LNCS (LNAI), vol. 2631. Springer, Heidelberg (2003)
5. Sackett, D., Rosenberg, W., Gray, J., Haynes, R., Richardson, W.: Evidence-based medicine: what it is and what it isn't. BMJ 312(7023), 71–72
6. Georg, G., Jaulent, M.-C.: A Document Engineering Environment for Clinical Guidelines. In: Proceedings of the 2007 ACM Symposium on Document Engineering, Winnipeg, Manitoba, Canada, pp. 69–78. ACM Press, New York (2007)
7. Pelachaud, C.: Multimodal expressive embodied conversational agent. In: ACM Multimedia, Brave New Topics session, Singapore, pp. 683–689 (2005)
8. De Carolis, B., Pelachaud, C., Poggi, I., Steedman, M.: APML, a Markup Language for Believable Behavior Generation. In: Prendinger, H., Ishizuka, M. (eds.) Life-like Characters. Tools, Affective Functions and Applications, pp. 65–86. Springer, Heidelberg (2003)
9. Moulin, B., Rousseau, D.: Knowledge acquisition from prescriptive texts. In: Proceedings of the 3rd international conference on Industrial and engineering applications of artificial intelligence and expert systems, Charleston, South Carolina, United States, pp. 1112–1121 (1990)
10. Ekman, P.: About brows: Emotional and conversational signals. In: von Cranach, M., Foppa, K., Lepenies, W., Ploog, D. (eds.) Human ethology: Claims and limits of a new discipline: contributions to the Colloquium, pp. 169–248. Cambridge University Press, Cambridge (1979)
11. Chovil, N.: Discourse-oriented facial displays in conversation. Research on Language and Social Interaction 25, 163–194 (1991)
12. Poggi, I.: Mind Markers. In: Rector, M., Poggi, I., Trigo, N. (eds.) Gestures, Meaning and use, pp. 203–207. University Fernando Pessoa Press, Oporto (2003)
13. Bolinger, D.: Intonation and its Part. Stanford University Press (1996)
14. Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., Hirschberg, J.: ToBI: A Standard for Labeling English Prosody. In: Proceedings of the International Conference on Spoken Language Processing, Banff, Alberta, pp. 867–870 (1992)
15. Poggi, I., Pelachaud, C.: Performative faces. Speech Communication 26, 5–21 (1998)
16. Darwin, C.R.: The expression of emotions in man and animals. Murray, London (1872)
17. Piwek, P., Hernault, H., Prendinger, H., Ishizuka, M.: T2D: Generating Dialogues Between Vir-tual Agents Automatically from Text. In: Intelligent Virtual Agents 2007, Paris, pp. 161–174 (2007)
18. Charon, R.: Narrative Medicine - A Model for Empathy, Reflection, Profession, and Trust. The Journal of the American Medical Association 286(15), 1897–1902 (2001)
19. Marsella, S., Gratch, J., Rickel, J.: Expressive behaviors for virtual worlds. In: Prendinger, H., Ishizuka, M. (eds.) Life-like Characters. Tools, Affective Functions and Applications, pp. 317–376. Springer, Heidelberg (2003)
20. De Carolis, B., De Rosis, F., Carofiglio, V., Pelachaud, C., Poggi, I.: Interactive Information Presentation by Embodied Animated Agent. In: International Workshop on IPNMD, Verona, Italy (2001)
21. De Rosis, F., De Carolis, B., Carofiglio, V., Pizzutilo, S.: Shallow and Inner Forms of Emotional Intelligence in Advisory Dialog Simulation. In: Prendinger, H., Ishizuka, M. (eds.) Life-like Characters. Tools, Affective Functions and Applications, pp. 271–294. Springer, Heidelberg (2003)

22. De Rosis, F., Novielli, N., Carofiglio, V., Cavalluzzi, A., De Carolis, B.: User modeling and adaptation in health promotion dialogs with an animated character. Journal of Biomedical Informatics 39(5), 514–531 (2006)
23. Bickmore, T., Caruso, L., Clough-Gorr, K.: Acceptance and usability of a relational agent interface by urban older adults. In: CHI 2005 extended abstracts on Human factors in computing systems, pp. 1212–1215 (2005)
24. Bickmore, T., Pfeifer, L., Paasche-Orlow, M.: Health Document Explanation by Virtual Agents. In: Intelligent Virtual Agents 2007, Paris, pp. 183–196 (2007)
25. Cassell, J., Torres, O., Prevost, S.: Turn Taking vs Discourse Strcuture: How Best to Model Multimodal Conversation. In: Wilks, Y. (ed.) Machine Conversations. Kluwer, Dordrecht (1999)
26. Burke, M., Dunlap, W.: Estimating interrater agreement with the average deviation index: a user's guide. Organizational Research Methods 5(2), 159–172 (2002)