

# Politeness and Alignment in Dialogues with a Virtual Guide

Markus de Jong, Mariët Theune, Dennis Hofs  
Human Media Interaction  
University of Twente  
The Netherlands

justmarkus@gmx.net, m.theune@ewi.utwente.nl, d.h.w.hofs@ewi.utwente.nl

## ABSTRACT

Language alignment is something that happens automatically in dialogues between human speakers. The ability to align is expected to increase the believability of virtual dialogue agents. In this paper we extend the notion of alignment to affective language use, describing a model for dynamically adapting the linguistic style of a virtual agent to the level of politeness and formality detected in the user's utterances. The model has been implemented in the Virtual Guide, an embodied conversational agent giving directions in a virtual environment. Evaluation shows that our formality model needs improvement, but that the politeness tactics used by the Guide are mostly interpreted as intended, and that the alignment to the user's language is noticeable.

## Categories and Subject Descriptors

H.5.2 [Information interfaces and presentation]: User interfaces—*natural language*

## General Terms

Human Factors, Languages

## Keywords

virtual agents, politeness, alignment

## 1. INTRODUCTION

Language alignment is something that happens automatically in dialogues between human speakers [10], not only at the syntactic level [5, 8] but also at the level of linguistic style in relation to affective factors [2, 6, 17]. Implementing alignment and the corresponding variations in linguistic style in virtual dialogue agents can help to give them a recognizable personality [5], make them appear more socially intelligent [1, 9] and increase their believability [17].

In this paper we describe a model for dynamically aligning the language use of a virtual agent to the level of politeness and formality displayed by the user's utterances. Our agent, the Virtual Guide, is an embodied conversational agent that gives directions in a virtual environment (a 3D theatre building). The Virtual Guide can answer questions related to the

**Cite as:** Politeness and Alignment in Dialogues with a Virtual Guide, Markus de Jong, Mariët Theune, and Dennis Hofs, *Proc. of 7th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, Padgham, Parkes, Müller and Parsons (eds.), May, 12-16., 2008, Estoril, Portugal, pp. XXX-XXX.

Copyright © 2008, International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

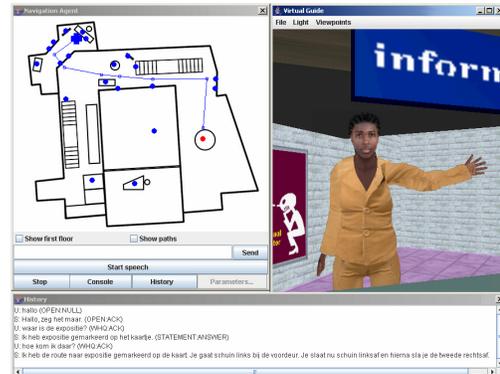


Figure 1: The Virtual Guide.

locations of objects in the virtual building, display routes and locations on a map, and give route directions using speech and gesture [13]. See Figure 1 for a screen shot.

Like most information-giving dialogue agents, the original version of the Virtual Guide always used the same, neutral language in all circumstances. Recently we have extended it with the possibility of linguistic alignment, which should result in a more human-like and varied style of interacting with the user. When in 'alignment mode', politely asking the Guide a question will result in a polite answer, while a rude question will result in a less polite answer, with the level of alignment determined by a parameter setting.

Our alignment model has three dimensions: *politeness* is reflected in the choice of sentence structures, *formality* relates to word choice, and *T-V distinction* (named after the Latin informal and formal personal pronouns "tu" and "vos") concerns the choice of personal pronouns. During interaction, our agent analyses each user utterance on these three dimensions, and adapts its own language accordingly.

Below, we first discuss related work (Section 2). We then describe the analysis of user input in our system (Section 3) and the generation of appropriate responses (Section 4). We present our first attempts at evaluation (Section 5), and end with a discussion and pointers to future work (Section 6).

## 2. RELATED WORK

### 2.1 Politeness and linguistic style

Brown and Levinson's Theory of Politeness [3] has formed an inspiration for most work on politeness and linguistic style, including ours. At its heart lies the idea that speakers

are polite in order to save the hearer’s face: a public self-image that every person wants to pursue. The concept of face is divided in *positive face*, the need for a person to be approved of by others, and *negative face*, the need for autonomy from others. Whenever a speech act goes against either of these needs, this is called a Face Threatening Act (FTA). Brown and Levinson discuss three main strategies that influence the linguistic choices the speaker makes: the On-record, the Off-record and the Don’t do FTA strategy.

When using the *On-record strategy*, the speaker’s intention is stated unambiguously. This strategy is split into two sub-strategies: bald and redressive. When a *bald* strategy is used, the FTA is phrased in direct terms without accounting for face threat, for example using an imperative. This strategy is used when there is no time or need to care about the hearer’s face, for example in emergency situations (“Help me!”) or when the speaker is (or perceives himself to be) superior to the hearer. It can also be safely used for speech acts that hardly pose a face threat, for example straightforward Informs such as “The car is in the parking lot”. The *redressive* substrategy involves unambiguous speech acts that are toned down in force by redressive language. Redressive language can be approval-oriented (positive face), for example by using first names or in-group names when addressing the hearer (“Hey pal, what’s the time?”) or by flattering the hearer. It can also be autonomy-oriented (negative face), for example by using *hedges*: words or phrases that diminish the face-threatening force of a speech act, such as the word “just” in “I just want to ask...”.

The *Off-record strategy* is an indirect way of executing the FTA by phrasing it in such a way that more than one interpretation is possible, and one of these interpretations does not pose a threat to the hearer’s face. For instance, when someone says “This weather always makes me thirsty” this is probably a hint that he would like a drink. However, the speaker can always ignore the indirect requesting interpretation and claim the literal (and non face-threatening) informing interpretation instead. Finally, when using the *Don’t do FTA strategy* the speaker omits to perform the FTA, thus completely avoiding any threat to the hearer.

Presumably the first attempt at implementing these politeness strategies in virtual agents was made by Walker et al. [17]. In their approach, the desired level of politeness of an utterance depends on the social distance between the dialogue participants, the power one has over the other, and the estimated face threat posed by the speech act to be expressed. A recent follow-up to their work is that of Gupta et al. [7], who present the POLLY system that models Brown and Levinson’s theory in task-oriented dialogue.

André et al. [1], Johnson et al. [9] and Porayska-Pomsta and Mellish [11] generate tutoring responses based on Brown and Levinson’s theory. Unlike Walker et al. [17], who used four basic politeness categories without making a more fine-grained distinction, André et al. distinguish different sub-strategies with varying levels of politeness. Johnson et al. [9] and Porayska-Pomsta and Mellish [11] use templates for sentence generation, which they annotate with separate numeric values for positive and negative face (where other approaches use only one numeric value to capture both).

## 2.2 Linguistic alignment

Based on evidence from psycholinguistics, Pickering and Garrod [10] have proposed an “interactive alignment ac-

count” of dialogue, claiming that the linguistic representations used by the conversation partners in a dialogue automatically become aligned at many levels. Isard et al. [8] focus on syntactic alignment in their implementation of the CRAG-2 system, which uses n-gram language models to achieve syntactic alignment between two dialogue participants (both virtual agents). The system can vary its degree of alignment to simulate different agent personalities, cf. [5].

Bateman and Paris [2] suggest extending Pickering and Garrod’s alignment model to affective alignment, so that the system matches the user’s affective style (or ‘register’), including the level of politeness. Bateman and Paris stress the importance of interpretation: in order to achieve alignment, the system must be able to recognize stylistic variations in the utterances of the user. An architecture capable of achieving this is described by Guinn and Hubal [6], who use semantic grammars for both analysis and generation of affective language, distinguishing different emotional and attitudinal factors including politeness.

## 3. INPUT ANALYSIS FOR ALIGNMENT

With the exception of [6], most work on linguistic style variation for dialogue agents seems to have focused on generation based on static input parameters. As future work, Walker et al. [17] mention exploring a ‘reciprocal feedback loop’ where the relevant parameters are not set in advance but change dynamically over the course of the dialogue, leading to interesting changes in the way the conversational partners interact with each other. Achieving such a system is exactly what we set out to do in the research described here.

The Virtual Guide is set up such in a way that the user must always initiate the dialogue. Thus it has the opportunity to scan the user’s input before deciding what linguistic style it should use. Our approach to analysing the level of politeness and formality in the user’s utterances is similar to that of Guinn and Hubal [6], who apply emotional or attitudinal tags to grammar rules to extract affective information from user utterances. For example, the use of words such as “please” increases the politeness of an utterance.

In the following sections we describe how we analyse the user utterances on the three dimensions distinguished in our model: politeness, formality and T-V distinction.

### 3.1 Politeness

Data on politeness in Dutch are available from the work of Roel Vismans [16], who asked Dutch speakers to rate the politeness of different linguistic variations of the same directive: “de deur dicht doen” (close the door). Vismans defined the context of the utterances as an informal setting with strangers, which more or less corresponds to the interaction context of the Virtual Guide. Therefore we use Vismans’ results as a basis for determining the level of politeness of the user input in our system, with politeness indicators being sentence structure and the use of modal particles. The total politeness value of a user utterance is calculated by adding the effect of modal particles (MP-Effect) to the basic sentence politeness (P):

$$\text{UserPoliteness} = P + \text{MP-Effect}$$

#### 3.1.1 Sentence structure

To investigate the influence of sentence structure on politeness, Vismans asked 24 subjects to rate 9 variations of a

**Table 1: Some sentence structures that can be recognized by the grammar. Politeness values (P) are based on the ratings from Vismans, converted to a scale from -5 (very impolite) to 5 (very polite).**

Form	Example sentences	P
IMP	Toon (me) de zaal. (Show me the hall.)	-3
DECL	Je moet me vertellen waar de zaal is. (You have to tell me where the hall is.)	-2
	Ik moet/ wil naar de zaal. (I have to / want to go to the hall.)	-1
	Ik zoek de zaal. (I am looking for the hall.)	0
	Je zou de zaal moeten tonen. (You should show me the hall.)	0
INT	Hoe moet ik bij de zaal komen? (How must I get to the hall?)	-1
	Waar is de zaal? (Where is the hall?)	0
	Waar / hoe vind ik de zaal? (Where/how do I find the hall?)	0
	Waar / hoe kan ik de zaal vinden? (Where/how can I find the hall?)	1
	Wil je de zaal tonen? (Do you want to show me the hall?)	2
	Zou je de zaal willen/kunnen tonen? (Would/could you show me the hall?)	3
	Kun je de zaal tonen? (Can you show me the hall?)	3
	Weet je waar de zaal is? (Do you know where the hall is?)	3

request to close the door. Imperative sentences (bald on-record) were rated as least polite, and interrogative sentences (redressive on-record) as most polite. Requests phrased as a declarative (“You should...”) were rated in between.

Based on Vismans’ results we associated the grammar rules used for input analysis in our system with tags indicating their level of politeness. Table 1 shows some variations of the request “Toon me de zaal” (Show me the hall) that can be analysed by our grammar.<sup>1</sup> For the sentence structures corresponding to those tested in Vismans’ experiments, we adopted the values found by Vismans. They were converted to the politeness scale used in our system, which ranges from -5 (least polite) to 5 (most polite). For the other sentence structures we determined a politeness value P based on the use of forceful verbs such as “moeten” (must) or mitigating verbs as “zouden” (could). Declarative sentences such as “Ik zoek de zaal” (I’m looking for the hall) can be seen as examples of Brown and Levinson’s *off-record* strategy. However, we felt that such utterances do create an expectation for the addressee to act that is hard to ignore, so we decided to assign them a neutral rather than a positive politeness value. Declaratives using the forceful phrases “Ik wil...” (I want...) and “Ik moet...” (I must...) were even ranked below neutral. The lowest-scoring imperative sentence structure found by Vismans, “Deur dicht” (lit. Door closed) is very specific and does not apply to the “show me the hall” example. As a result, -3 is the lowest value in Table 1. Other dialogue act types not illustrated in Table 1

<sup>1</sup>The English translations provided in this paper may differ from the Dutch originals in politeness and formality.

<code>\w</code>	VERB_MARK	
<code>\n</code>	markeren	; ‘indicate’
<code>\s</code>	1	
<code>\n</code>	aangeven	; ‘show’
<code>\s</code>	-1	
<code>\n</code>	aanstippen	; ‘mark’
<code>\s</code>	-3	

**Figure 2: A shaded lexicon entry**

include opening and closing acts (greetings/farewells) (+2) and thanking acts (+3).

### 3.1.2 Modal particles

One of the politeness tactics mentioned by Brown and Levinson [3] is *hedging*, which includes the use of modal particles such as “perhaps” or “possibly”, as in “Could I perhaps/possibly have the salt?” Based on his experiments [16], Vismans separates Dutch modal particles in *reinforcers* (dan, nou, ook, toch, eens) and *mitigators* (even, maar, misschien, soms), where reinforcers apply more pressure to the hearer of the speech act, while mitigators (i.e., hedges) do the opposite. However, the stronger the force of the FTA, the weaker the added effect of reinforcers or mitigators.

We use the following formula to calculate the effect of a modal particle on the overall politeness of an utterance:

$$MP\text{-Effect} = (5 - |P|)/5 * MP$$

*MP-Effect* is the total effect of the modal particle. *MP* is the politeness value of the mitigator or reinforcer, where mitigators have politeness +1, while reinforcers have politeness -1. *P* is the politeness value attributed to the sentence structure (see Section 3.1.1), which we take as representative of the force of the FTA. The closer this value is to either end of the scale, the stronger the force of the FTA – either in a positive or a negative direction. Thus, using this formula, the effect of mitigators and reinforcing particles is reduced on the extreme ends of the politeness scale.

## 3.2 Formality

In our model, formality values are attributed to words or phrases and their synonyms in a ‘shaded lexicon’. This part of the system was inspired by the work of Fleischman and Hovy [4], who use an emotionally shaded lexicon to achieve emotional variation in natural language generation. In our lexicon, the words and phrases are shaded with different levels of formality instead of emotion, and we use the lexicon not only for generation but also for analysis. The lexicon includes different greetings and salutations, ranging from the informal “Yo” to the formal “Goedemiddag” (Good afternoon), as well as numerous verb and noun synonyms that vary in formality as well. Compare the informal “plee” (loo) and the formal “toilet” (lavatory), informal “aangeven”(to show) and formal “markeren”(to indicate).

An example of a lexical entry in the shaded lexicon is shown in Figure 2. This is the entry for the concept MARK (indicated by `\w`). The different ways to express this concept are indicated by `\n` and their formality shades by `\s`. The shades in the lexicon range between -5 (very informal) to 5 (very formal). It must be noted that due to lack of proper data on the formality of Dutch words, the shades of the

**Table 2: T-V distinction in Dutch**

Formal	Informal	Translation
u	je/jij	you
uw	je/jouw	your
dank u	dank je	thank you
alstublieft	alsjeblijft	please

lexical entries were determined in a rather ad hoc fashion, partly based on information found in a dictionary (Van Dale On-line [14]) and partly on the authors’ intuition.

When analysing a user utterance to determine its formality (*UserFormality* in the formula below), each word is looked up in the lexicon and its formality shade, *wordFormality* is returned, if available. Then the individual values of the words in the utterance are summed up and the average formality score for the sentence is calculated by dividing the sum of the shades by the number (*k*) of shaded words.

$$\text{UserFormality}(i) = (\sum_1^k \text{wordFormality}(i))/k$$

### 3.3 T-V Distinction

In Dutch, the speaker can use either formal (“u”) or informal (“je”) personal pronouns to address the hearer. This distinction also affects other words and phrases that incorporate personal pronouns, as illustrated in Table 2.

In our model, T-V distinction is represented by a value of either 1 or 0. If formal versions of the phrases from Table 2 are detected in the user input, this value is set to 1, otherwise to 0. Our main reason to treat T-V distinction as separate from politeness and formality, even though it is clearly related to both, is a practical one (from the point of view of generation): if pronoun choice were to depend on the user’s current level of politeness and/or formality, this might cause our Virtual Guide to switch between the use of formal or informal pronouns without the user ever having made a change in T-V distinction at all.

So, for practical reasons we cannot let T-V distinction be influenced by politeness or formality. Currently, our model also does not allow for an influence in the other direction, even though the use of formal or informal pronouns likely affects the perceived politeness and formality of an utterance.

## 4. GENERATING ALIGNED UTTERANCES

After having analysed the user’s utterance for politeness, formality and T-V distinction, the system has to generate a response using the appropriate linguistic style. This section describes the generation process, which takes as input (1) a system dialogue act, selected by the system’s dialogue manager as described in [13], and (2) the current alignment state of the system, consisting of three values for politeness, formality and T-V distinction. The initial alignment state of the Guide can be set manually; default values are neutral (0) for politeness and formality, and informal for T-V distinction. As the dialogue advances, the Guide adapts its alignment state according to the following formula, where the degree of alignment is set by the variable  $\alpha$ :

$$\text{Politeness}(i+1) = \alpha * \text{Politeness}(i) + (1-\alpha) * \text{UserPoliteness}(i+1)$$

$$\text{Formality}(i+1) = \alpha * \text{Formality}(i) + (1-\alpha) * \text{UserFormality}(i+1)$$

Here *Politeness*(*i*) and *Formality*(*i*) are the current politeness and formality values in the Guide’s alignment state, and *UserPoliteness*(*i+1*) and *UserFormality*(*i+1*) are the politeness and formality values of the user’s most recent utterance. The value of  $\alpha$  varies between 0 and 1 and determines how changeable the alignment state is. The closer  $\alpha$  comes to 1, the slower the Guide will adapt its language to that of the user. Changing the value of  $\alpha$  allows us to experiment with different alignment settings, varying between no alignment at all when  $\alpha$  is 1 and full alignment when  $\alpha$  is 0.

In contrast to politeness and formality, the T-V distinction value is not computed but simply set to 0 or 1, depending on the user’s choice of addressing. The value is only changed if the user switches to a different form of addressing.

Based on the values in the alignment state, the system selects a surface realisation for the dialogue act to be carried out. First, an appropriate politeness tactic, in the form of a sentence template, is selected depending on the politeness value (Section 4.1). Then, certain slots in this template are filled with words that correspond with the formality value of the alignment state (Section 4.2). Finally, T-V distinction is applied by replacing personal pronouns and other relevant words (see Table 2) according to the current T-V value. As a basis for surface realisation we use Exemplars [18].

### 4.1 System Politeness

For the selection of appropriate politeness tactics we further extended the categories proposed by Walker et al. [17]. Based on the politeness theory of Brown and Levinson [3], they distinguish the following politeness strategies, in order of increasing politeness: DIRECT (bald on-record), APPROVAL (redressive on-record with positive politeness tactics), AUTONOMY (redressive on-record with negative politeness tactics), and INDIRECT (off-record). Walker et al. do not provide sub-rankings of the individual politeness tactics used to realise the strategies. However, clear differences in politeness exist between the individual tactics within each strategy, so we have split the strategies into multiple tactic clusters (cf. [1]). A list of strategies and tactics for the generation of requests, with example sentences and corresponding politeness ranges, is given in Table 3. This table shows tactical variations of a request to look at the map, and of a request to rephrase the last user utterance (issued after a speech recognition or text analysis failure). The tactics are based on the positive and negative tactics of Brown and Levinson [3] and Vismans’ data on the expression of politeness in Dutch [16], but also inspired by the work of Gupta et al. [7]. Their grouping into different clusters was partly based on Vismans’ data, and partly on intuition (for the majority of cases). Note that some templates at the higher politeness levels include mitigating modal particles, for example “Kun je eventjes op het kaartje kijken?” (Can you look at the map for a moment?).

Because many politeness tactics cannot be properly applied to Inform acts, which make up a large part of the system actions, we did not create tactic clusters for Informs. Also, because Informs do not need to be as polite as Requests, the *Indirect* tactic was omitted altogether. In order to increase variation, some Inform acts can also be formulated Requests. For example, the act of informing the user that the Guide has marked a location on the map, can be phrased as a request to look at the map (as in Table 3).

**Table 3: Politeness strategies and tactics for the generation of requests. The politeness ranges (P) use a scale from -5 (very impolite) to 5 (very polite).**

Strategy	Tactic	Example sentence	P
<b>DIRECT</b>	1) Imperative	Kijk op het kaartje waar ik de zaal heb aangegeven (Look at the map where I've marked the hall)	-4 to -2.5
	2) Declarative 1	Je moet het anders zeggen (You have to phrase it differently)	-2.5 to -2
<b>APPROVAL 1</b>	3) Ellipsis	Nog een keer proberen? (Try again?)	-2 to -0.5
	4) Inclusive	Zullen we het nog een keer proberen? (Shall we try again?)	
	5) Ability	Is het mogelijk dat je op het kaartje kijkt waar ik de zaal heb gemarkeerd? (Is it possible for you to look at the map where I've marked the hall?)	
	6) Ingroup name	Probeer het nog eens, vriend (Try again, mate)	
<b>APPROVAL 2</b>	7) Optimism	Je vindt het vast niet erg om het nog een keer te proberen (I'm sure you don't mind trying again)	-0.5 to 1
	8) Give reason	Als je het anders formuleert, lukt het vast wel (If you phrase it differently, it's bound to work)	
	9) Mind	Als je het niet erg vindt, kun je op het kaartje kijken waar ik de zaal heb gemarkeerd (If you don't mind, you can look at the map where I've marked the hall)	
	10) Declarative 2	Je zou het nog een keer moeten proberen (You ought to try again)	
<b>AUTONOMY 1</b>	11) Nominalize	De vraag is of je het anders wilt formuleren (The question is if you want to phrase it differently)	1 to 2.25
	12) Impersonalize	Is het mogelijk dat het nog een keer geprobeerd kan worden? (Would it be possible for it to be tried again?)	
	13) Distance in time	Ik vroeg me af of je op het kaartje wilt kijken waar ik de zaal heb gemarkeerd (I was wondering if you want to look at the map where I've marked the hall)	
	14) Conventionally indirect 1	Wil je het nog een keer proberen? (Do you want to try again?)	
<b>AUTONOMY 2</b>	15) Conventionally indirect 2	Kun je op het kaartje kijken waar ik de zaal heb gemarkeerd? (Can you look at the map where I've marked the hall?)	2.25 to 3.25
	16) Subjunctive pessimism 1	Zou je op het kaartje willen kijken waar ik de zaal heb gemarkeerd? (Would you like to look at the map where I've marked the hall?)	
	17) Subjunctive pessimism 2	Zou je het anders kunnen formuleren? (Could you phrase it differently?)	
<b>AUTONOMY 3</b>	18) Hedging	Kun je het misschien anders formuleren? (Can you perhaps phrase it differently?)	3.25 to 4
	19) Minimize imposition	Kun je eventjes op het kaartje kijken waar ik de zaal heb gemarkeerd? (Can you look at the map for a moment where I've marked the hall?)	
	20) Apologize	Sorry, kun je het nog eens proberen? (Sorry, can you try again?)	
<b>INDIRECT</b>	21) Indirect	Iemand zou het nog een keer moeten proberen (Someone should try again)	4 to 5

**Table 4: Evaluation results for the Politeness tactics from Table 3, ranked in order of increasing politeness on a scale from 1 (very polite) to 5 (very impolite). Results are average scores with standard deviation between brackets.**

Tactic	Result	Tactic	Result	Tactic	Result
2) Declarative 1	3.88 (0.60)	10) Declarative 2	2.92 (0.70)	13) Distance in time	2.24 (0.72)
3) Ellipsis	3.80 (0.82)	8) Give reason	2.88 (0.93)	5) Ability	2.20 (0.71)
6) Ingroup name	3.72 (0.79)	19) Minimize imposition	2.52 (0.77)	20) Apologize	2.15 (0.69)
1) Imperative	3.68 (0.75)	15) Conventionally indirect 2	2.36 (0.64)	16) Subj. pessimism 1	2.00 (0.41)
11) Nominalize	3.24 (1.00)	14) Conventionally indirect 1	2.28 (0.54)	12) Impersonalize	1.96 (0.45)
7) Optimism	3.08 (0.81)	4) Inclusive	2.24 (0.78)	18) Hedging	1.96 (0.54)
21) Indirect	3.08 (0.86)	9) Mind	2.24 (0.66)	17) Subj. pessimism 2	1.92 (0.57)

## 4.2 System Formality

The formality value determines the choice of isolated words and phrases within the sentence templates chosen based on politeness. Formality alignment in the Virtual Guide is accomplished by selecting formal or informal versions of words and phrases from a shaded lexicon (see Section 3.2) and inserting them into the template of the chosen politeness tactic to form a complete sentence. The words are selected from pools of synonyms that exist for different ranges of formality: 5 to 3 for very formal, 3 to 1 for formal, 1 to -1 for neutral, -1 to -3 for informal and -3 to -5 for very informal. These ranges are less precise than those used for Politeness, since we do not have exact data on formality.

So, the system determines for the current formality value in which range it falls, and then randomly selects a word within this range for each of the open slots in the sentence template. The following are examples of the insertion of different synonyms into a template. The inserted words or phrases are given in square brackets.

### Informal:

“Ik heb [de wc] [aangegeven] op [het kaartje]”  
(I have [marked] [the toilet] on [the little map])

### Formal:

“Ik heb [het toilet] [gemarkeerd] op [de plattegrond]”  
(I have [indicated] [the lavatory] on [the plan])

## 5. EVALUATING THE MODEL

To evaluate our model, we let 25 speakers of Dutch rate a number of pre-generated system utterances and dialogues on their level of politeness, formality and alignment. The evaluation procedures and results for each aspect are briefly discussed in the following sections. TV-distinction in the experiment was set to informal; this dimension was kept constant and was not evaluated.

### 5.1 Politeness evaluation

We asked the participants to rate the sentences from Table 3, embodying the different politeness tactics, on a scale ranging from 1 (very polite) to 5 (very impolite).<sup>2</sup> Participants were given the opportunity to comment on their ratings. Due to an error that was only discovered halfway through the experiment, the *Apologize* tactic (20) was rated only by 13 participants. Table 4 shows the average ratings for individual tactics, ordered by increasing politeness.

Table 5 shows how the average ratings of the politeness tactics fit within the range assigned by the model to their (sub)strategy. Tactics that fall outside the predicted range are given in italics. The size of the deviation is given between brackets, with + indicating that the tactic was rated as more polite than predicted, and - indicating it was rated as less polite. The largest deviation occurs in the INDIRECT strategy / tactic, which is rated as much less polite than predicted. Apparently, hearers do not appreciate being indirectly referred to as ‘somebody’ and this annuls the effect of indirectness. This result corresponds to that of Gupta et al. [7]. In a cross-cultural evaluation of their POLLY system they found that indirect tactics, which should have been the

<sup>2</sup>In the experiment, we used a five-point politeness scale rather than the more fine-grained scale employed in our model, to simplify the rating task for the participants.

**Table 5: Strategies with politeness ranges converted to a 5-point scale. Emphasized tactics received average scores outside the predicted range. The deviation is given between brackets, with + indicating that the tactic was rated as more polite than predicted, and - indicating it was rated as less polite.**

Strategy (range)	Tactics (deviation)
DIRECT 1 (4.6–4.0)	1) <i>Imperative (+0.32)</i>
DIRECT 2 (4.0–3.8)	2) Declarative 1
APPROVAL 1 (3.8–3.2)	3) Ellipsis
	4) <i>Inclusive (+0.96)</i>
	5) <i>Ability (+1.00)</i>
	6) Ingroup name
APPROVAL 2 (3.2–2.6)	7) Optimism
	8) Give reason
	9) <i>Mind (+0.36)</i>
	10) Declarative 2
AUTONOMY 1 (2.6–2.1)	11) <i>Nominalize (-0.64)</i>
	12) <i>Impersonalize (+0.14)</i>
	13) Distance in time
	14) Conv. indirect 1
AUTONOMY 2 (2.1–1.7)	15) <i>Conv. indirect 2 (-0.26)</i>
	16) Subj. pessimism 1
	17) Subj. pessimism 2
AUTONOMY 3 (1.7–1.4)	18) <i>Hedging (-0.26)</i>
	19) <i>Min. imposition (-0.82)</i>
	20) <i>Apologize (-0.45)</i>
INDIRECT (1.4–1.0)	21) <i>Indirect (-1.68)</i>

most polite, were rated most impolite of all.

Large deviations from the model are also found for the *Inclusive* (4) and the *Ability* (5) tactics. Although the *Inclusive* tactic was found patronizing (and thus not very polite) by some participants, the average ratings of both tactics were much more polite than predicted; the difference with the ratings of the other tactics in the APPROVAL 1 cluster is highly significant ( $p < 0.001$ ). In APPROVAL 2, the *Mind* tactic (9) was rated as significantly more polite than the other tactics ( $p < 0.01$ ). A frequent comment on this tactic was that subjects found the phrase “Als je het niet erg vindt” (If you don’t mind) out of place in the context of the request to look at the map, stating: “Why would I mind?”, indicating the absence of any threat to autonomy. In AUTONOMY 1, the *Nominalize* tactic (11) was rated as significantly less polite than the other tactics ( $p < 0.001$ ), whereas the *Impersonalize* tactic (12) was rated as slightly more polite (only significant w.r.t. the *Conventionally indirect 1* tactic,  $p < 0.05$ ). Both tactics were sometimes described as “unnatural”. In AUTONOMY 2, the *Conventionally indirect 2* tactic was rated as significantly less polite than the other tactics ( $p < 0.05$ ). All the tactics in AUTONOMY 3 were rated as less polite than predicted. In particular the *Minimalizing imposition* tactic (19) was rated as relatively impolite; apparently we overestimated the mitigating effect of the modal particle “eventjes” (a moment).

Finally, the *Imperative* (1) tactic deviated from its predicted range, but its average rating did not significantly differ from the other direct tactic *Declarative 1* (2), suggesting that the two could be assigned the same range in the model. However, for the moment we have adopted another solution, which is to shorten the *Imperative* template by removing the helpful clause “waar ik de zaal heb aangegeven” (where I

marked the hall), which may have mitigated the effect of the imperative. Other preliminary adjustments we have made to the model are the following:

- Tactics (4) and (5) were moved to the more polite APPROVAL 2
- Tactic (9) is no longer used for requests that are not face-threatening
- Tactic (19) was moved to the less polite AUTONOMY 2

However, it is clear that more adjustments are necessary.

## 5.2 Formality evaluation

The participants were also asked to rate 20 system utterances on a scale ranging from 1 (very formal) to 5 (very informal). Although the overall formality ranking of the utterances by the participants more or less corresponds with our model, the average formality ratings of the individual sentences generally do not match the formality ranges assigned to them by the model. The average ratings of most sentences came out around a ‘neutral’ value, with high standard deviations (0.94 on average). Apparently, formality judgements are very much a matter of personal ‘taste’. Also, apparently a few participants took politeness factors into account when rating formality. Asking the participants to rate individual words might have given more consistent results. However, this would not have provided us with information on the combined effect of these words in the utterances of the Virtual Guide.

## 5.3 Alignment evaluation

Finally, the participants were presented with three different versions of three short dialogues with the Virtual Guide. Three different user profiles were used for the generation of these dialogues. In dialogue 1, the user was informal and impolite, in dialogue 2 the user was formal and polite and in dialogue 3 the user was informal and polite. The three versions of each dialogue had the same user input but different degrees of alignment, leading to different system responses. The alignment settings used were 0%, 50% and 100% alignment. As an illustration, Figure 3 shows the 0% and 100% versions of dialogue 1. The other two dialogues were similar in length (3 to 4 pairs of dialogue acts) and in content.

The resulting nine dialogue versions were presented independently and in a random order to the participants, who were asked to rate each version on how well the Guide adapted its language to the user on a four-point scale. For dialogues 2 and 3 the relative ratings of the three versions reflected the degree of alignment used by the system; however for dialogue 1 the 50% alignment version was actually seen as aligning less than the 0% alignment version. This was probably caused by the fact that in the 50% version, after an imperative request from the user (“Show me the hall”) the Guide responded using the *Ability* tactic (“Is it possible for you to look at the map ...”). In the politeness evaluation described in Section 5.1, this tactic was rated as considerably more polite than its value in the model. As a consequence of this modelling error, for this utterance in the tested dialogue the system’s level of politeness did not match the user’s.

Finally, the participants were presented with all three versions of each dialogue and asked to choose which version they liked best (and explain their choice). This did not reveal clear overall preferences for one alignment setting or

<p>USER: Hi. (Hi)            GUIDE: Goeienavond. (Good evening)            USER: Toon me de zaal. (Show me the hall)            GUIDE: Je zou op de kaart moeten kijken waar ik de zaal heb aangegeven. (You should have a look at the map where I’ve indicated the hall)            USER: Doe. (See you)            GUIDE: Dag. (Good-bye)</p>
<p>USER: Hi. (Hi)            GUIDE: Hi. (Hi)            USER: Toon me de zaal. (Show me the hall)            GUIDE: Kijk op het kaartje waar ik de zaal heb aangestipt. (Look at the map where I’ve marked the hall)            USER: Doe. (See you)            GUIDE: Doe. (See you)</p>

**Figure 3: Dialogues with an informal, impolite user. Top: 0% alignment. Bottom: 100% alignment.**

the other, as some participants liked the way the Guide mirrored the user in the strong alignment settings, whereas others liked the fact that in the weak alignment versions, the Guide stayed formal and polite even when the user was not. As with formality, personal preference seems to play an important role here.

## 6. DISCUSSION AND FUTURE WORK

The Virtual Guide’s ‘job’ is to assist visitors navigating a virtual environment. In this paper we have presented a model for aligning the linguistic style of Guide’s utterances to the levels of politeness and formality detected in the utterances user. The system is fully implemented and functional. Using different parameter settings for the system’s initial levels of politeness and formality as well as the degree of alignment allows us to model different professional attitudes or personalities for the Guide. When using high politeness and no alignment, we can model a Guide that takes the saying ‘the customer is always right’ as its mission and tries to be as polite as possible, even when the user is not. Alternatively, higher degrees of alignment make for a less unflappable Guide who is influenced by the user’s behaviour: when the user is impolite, the Guide responds to this by being less polite as well (but never more impolite than the user). When the user is polite, the Guide is polite as well, but without becoming more polite than the user, as this could be interpreted as over-exaggerating and rude [15].

So far, we have only evaluated pre-generated system utterances and dialogues. Although the test results for politeness and alignment are somewhat promising, we cannot draw strong conclusions from them given the limited size of the experiments. For example, we used only one instance of each politeness tactic in our evaluation; other instances might have provided different results. The dialogues used in the alignment evaluation were quite short, and this may not have been sufficient to properly test the different alignment settings. Politeness and formality are currently treated as orthogonal dimensions in our model, and as such they were separately evaluated. However, in reality the two dimensions interact, and the evaluation results may have been influenced by this. Finally, the most important question, what effect the proposed alignment model has on users interacting

with the Virtual Guide, still remains to be investigated.

Unlike other work on politeness and linguistic style, our politeness model does not explicitly take the imposition ranking of different speech acts into account. This sometimes leads to the unnecessary use of politeness tactics in non-threatening Inform acts (see Section 5.1). However, as noted by Walker et al. [17], the face-threatening potential of a speech act does not solely depend on its type but also on its content. Some requests pose a stronger threat than others: compare a request to pass the salt to a request for money. In our system we also see examples of this in the shape of different Inform acts. Certain politeness tactics are appropriate for one Inform act (“It seems I didn’t understand you correctly”) but not for another (“It seems I have marked the exposition on the map”). So it seems that a more refined approach is needed here.

Achieving lexical and syntactic alignment (see Section 2.2) is not the main aim of our system, but it is a frequently occurring side-effect due to the use of the same politeness and formality values for both analysis and generation. For example, the user may greet the Guide with “Hi!”, which is considered as an informal greeting. In maximum alignment mode the Guide should respond with an equally informal greeting, which due to the use of the same shaded lexicon (and a limited number of synonyms in the same formality range) has a high chance of being “Hi” too. The second dialogue of Table 3 shows some examples of this effect.

In this paper, we have focused on text-based interaction, but the input and output modalities of the Virtual Guide also include speech and gesture. Vismans [15] notes the importance of intonation for politeness, and Rehm et al. [12] discuss the relation between gestures and verbal politeness strategies. Incorporating the use of these nonverbal modalities in our model is an interesting topic for future work.

## 7. ACKNOWLEDGMENTS

We thank Rieks op den Akker for his contributions to this research, and our anonymous reviewers for their useful comments on a previous version of this paper. This research was partially supported by the FP6 EU Network of Excellence Humaine and by NWO grant number 632.001.301.

## 8. REFERENCES

- [1] E. Andre, M. Rehm, W. Minker, and D. Buhler. Endowing spoken language dialogue systems with emotional intelligence. In *Affective Dialogue Systems*, LNCS 3068, pages 178–187, 2004.
- [2] J. Bateman and C. Paris. Adaptation to affective factors: architectural impacts for natural language generation and dialogue. In *Proceedings of the Workshop on Adapting the Interaction Style to Affective Factors at the 10th International Conference on User Modeling (UM-05)*, 2005.
- [3] P. Brown and S. C. Levinson. *Politeness - Some universals in language usage*. Cambridge University Press, 1987.
- [4] M. Fleischman and E. Hovy. Towards emotional variation in speech-based natural language generation. In *Proceedings of the Second International Natural Language Generation Conference (INLG’02)*, 2002.
- [5] A. Gill, A. Harrison, and J. Oberlander. Interpersonality: individual differences and interpersonal priming. In *Proceedings of the 26th Annual Conference of the Cognitive Science Society*, pages 464–469, 2004.
- [6] C. Guinn and R. Hubal. Extracting emotional information from the text of spoken dialog. In *Proceedings of the 9th International Conference on User Modeling*, 2003.
- [7] S. Gupta, M. A. Walker, and D. M. Romano. Generating politeness in task based interaction: An evaluation of the effect of linguistic form and culture. In *Proceedings of the Eleventh European Workshop on Natural Language Generation (ENLG-07)*, pages 57–64, 2007.
- [8] A. Isard, C. Brockmann, and J. Oberlander. Individuality and alignment in generated dialogues. In *Proceedings of the 4th International Conference on Natural Language Generation (INLG-06)*, pages 22–29, 2006.
- [9] L. Johnson, P. Rizzo, W. Bosma, M. Ghijsen, and H. van Welbergen. Generating socially appropriate tutorial dialog. In *Affective Dialogue Systems*, LNCS 3068, pages 254–264, 2004.
- [10] M. J. Pickering and S. Garrod. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27:169–226, 2004.
- [11] K. Porayska-Pomsta and C. Mellish. Modelling politeness in natural language generation. In *Proceedings of the Third International Conference on Natural Language Generation (INLG-04)*, LNAI 3123, pages 141–150, 2004.
- [12] M. Rehm and E. André. Informing the design of embodied conversational agents by analyzing multimodal politeness behaviors in human-human communication. In *Proceedings of the AISB Symposium for Conversational Informatics for Supporting Social Intelligence and Interaction*, pages 144–151, 2005.
- [13] M. Theune, D. Hofs, and M. van Kessel. The Virtual Guide: A direction giving embodied conversational agent. In *Proceedings of Interspeech 2007*, pages 2197–2200, 2007.
- [14] Van Dale Taalweb. Online Dutch dictionary. <http://www.vandale.nl/opzoeken/woordenboek/>.
- [15] R. Vismans. Beleefdheid, Nederlandse modale partikels en het ‘partikkeloze’ Engels. *Colloquium Neerlandicum*, 12:269–291, 1994.
- [16] R. Vismans. *Modal Particles in Dutch directives: a study in Functional Grammar*. IFOTT, Vrije Universiteit Amsterdam, 1994.
- [17] M. Walker, J. Cahn, and S. Whittaker. Linguistic style improvisation for lifelike computer characters. In *Entertainment and AI / A-Life, Papers from the 1996 AAAI Workshop.*, 1996. AAAI Technical Report WS-96-03.
- [18] M. White and T. Caldwell. EXEMPLARS: A practical, extensible framework for dynamic text generation. In *Proceedings of the Ninth International Workshop on Natural Language Generation (INLG-98)*, pages 266–275, 1998.