

A Personality-based Framework for Utterance Generation in Dialogue Applications

François Mairesse and Marilyn Walker

Department of Computer Science, University of Sheffield, Sheffield S1 4DP, United Kingdom
{F.Mairesse,M.A.Walker}@sheffield.ac.uk

Abstract

Conversation is an essential component of social behaviour, one of the primary means by which humans express emotions, moods, attitudes and personality. Thus a key technical capability for dialogue applications, such as interactive narrative systems (INS), human robot interaction (HRI) and spoken dialogue systems (SDS), is the ability to support natural conversational interaction. However, system utterances in existing systems are typically handcrafted, leading to problems of portability and scalability. We propose a framework for automatically generating language projecting different personality traits based on the ‘Big Five’ model of personality. We show that our PERSONAGE generator can produce utterances with recognisable personality for all Big Five traits, according to human judges. We also test the ability of PERSONAGE to vary the characters’ personality in an existing interactive narrative system, showing that some forms of variation can be automatically obtained in a new domain, depending on the level of utterance representation.

Introduction

Conversation is an essential component of social behaviour, one of the primary means by which humans express emotions, moods, attitudes and personality. Thus a key technical capability for dialogue applications, such as interactive narrative systems (INS), human robot interaction (HRI) and spoken dialogue systems (SDS), is the ability to support natural conversational interaction. While INS clearly have the additional technical requirement of producing system outputs that are recognisable as manifesting differences in dramatic character, the quality of the interaction for any dialogue application is affected by the linguistic style of system utterances, even when the application is fundamentally task-oriented (Cassell & Bickmore, 2003; Isbister & Nass, 2000; Walker, Cahn, & Whittaker, 1997; Wang *et al.*, 2005).

In order to support natural conversational interaction, dialogue applications often apply natural language processing to the user’s input to allow users flexibility in what they say to the system (Mateas & Stern, 2003; Pieraccini & Levin, 1995). However, utterances produced by the system, in both INS and SDS, are often highly handcrafted, e.g. sound extracts pre-recorded by actors. Although this approach produces high quality utterances, it suffers from problems of

portability and scalability (Walker & Rambow, 2002), or what has been called the *authoring bottleneck* in the case of INS (Mateas, 2007).

We believe that these problems can be addressed by research on automated methods of utterance generation, based on a spoken language generation (SLG) architecture (Walker & Rambow, 2002). This research could ultimately lead to off-the-shelf tools that would support the creation of interesting dramatic characters and dialogue partners, without the need for a detailed knowledge of underlying SLG implementation details. But the requirement for variation in linguistic style that supports social behaviour suggests that what is needed is a framework for automatically generating character variation in a comprehensive and systematic way, i.e. a method or a theory that quantifies the most important individual differences in behaviour, and the way they affect system linguistic output in dialogic situations.

Philosophers and psychologists have tried to identify important dimensions of human behaviour for hundreds of years, beginning with Theophrastus in Ancient Greece (B.C. 371-287). However, we propose that a natural framework can be found in the somewhat more recent research on personality psychology, which suggests that behavioural differences of human individuals are best categorised in terms of *personality traits*, which describe consistent patterns in the way individuals behave, feel, and think, and which affect human behaviour in most situations (Allport & Odbert, 1936; Norman, 1963). In fact, Allport, the founder of trait theory, suggests that personality traits are intuitively appealing for modelling different aspects of human behaviour because of their pervasiveness in descriptions of dramatic and literary character (Allport, 1960):

Almost all the literature of character—whether [non-fiction] or fiction, drama or biography—proceeds on the psychological assumption that each character has certain *traits* peculiar to himself which can be defined through the narrating of typical episodes from life.

Allport’s approach was based on the *Lexical Hypothesis*, i.e. that any trait important for describing human behaviour has a corresponding lexical token. These lexical tokens are typically adjectives, e.g. *trustworthy*, *modest*, *friendly*, *spontaneous*, *talkative*, *dutiful*, *anxious*, *impulsive*, *vulnerable*, etc. Allport & Odbert (1936) collected 17,953 trait terms from English and identified 4,500 as ‘stable traits’.

	High	Low
Extraversion	warm, gregarious, assertive, sociable, excitement seeking, active, spontaneous, optimistic, talkative	shy, quiet, reserved, passive, solitary, moody, joyless
Emotional stability	calm, even-tempered, reliable, peaceful, confident	neurotic, anxious, depressed, self-conscious, oversensitive, vulnerable
Agreeableness	trustworthy, considerate, friendly, generous, helpful, altruistic	unfriendly, selfish, suspicious, uncooperative, malicious
Conscientiousness	competent, disciplined, dutiful, achievement striving, deliberate, careful, orderly	disorganised, impulsive, unreliable, careless, forgetful
Openness to experience	creative, intellectual, imaginative, curious, cultured, complex	narrow-minded, conservative, ignorant, simple

Table 1: Example adjectives associated with the Big Five traits.

Trait	Set	Example output utterance	Score
Extraversion	low	· Chimichurri Grill isn't as bad as the others.	1.00
	high	· I am sure you would like Chimichurri Grill, you know. The food is kind of good, the food is tasty, it has nice servers, it's in Midtown West and it's a Latin American place. Its price is around 41 dollars, even if the atmosphere is poor.	6.33
Emotional stability	low	· I am not sure! I mean, Ch-Chimichurri Grill is the only place I would recommend. It's a Latin American place. Err... its price is... it's damn ex-expensive, but it pr-pr-provides like, adequate food, though. It offers bad atmosphere, even if it features nice waiters.	4.00
	high	· Let's see what we can find on Chimichurri Grill. Basically, it's the best.	6.00
Agreeableness	low	· I mean, Chimichurri Grill isn't as bad as the others. Basically, the staff isn't nasty. Actually, its price is 41 dollars. It's damn costly.	2.00
	high	· You want to know more about Chimichurri Grill? I guess you would like it buddy because this restaurant, which is in Midtown West, is a Latin American place with rather nice food and quite nice waiters, you know, okay?	5.75
Conscientiousness	low	· I am not kind of sure pal. Err... Chimichurri Grill is the only place I would advise. It doesn't provide unfriendly service! This restaurant is damn expensive, its price is 41 dollars.	3.00
	high	· Let's see what we can find on Chimichurri Grill. I guess you would like it since this eating house, which offers sort of satisfying food and quite satisfactory waiters, is a Latin American eating place.	6.00
Openness to experience	low	· Err... I am not sure. Mhm... I mean, Chimichurri Grill offers like, nice food, so I would advise it, also the atmosphere is bad and its price is 41 dollars.	3.50
	high	· You want to know more about Chimichurri Grill? I believe you would love it, you know. I guess it's in Midtown West. Although this eating house's price is around 41 dollars, the food is rather satisfactory. This eating place, which provides kind of second-rate atmosphere, is a Latin American restaurant, alright?	5.00

Table 2: Example outputs of PERSONAGE for all Big Five dimensions for the input content plan in Table 2, with average judge ratings on the corresponding generation dimension. Personality ratings are on a scale from 1 to 7, with 1 = very low (e.g. introvert) and 7 = very high on the dimension (e.g. extravert).

This approach led to a great deal of subsequent work over the last century, and a lively debate over the definition of the most essential traits. However, in the last 50 years, a standard framework has emerged of the 'Big Five' personality traits (Goldberg, 1990; McCrae & Costa, 1987; Norman, 1963). These traits were obtained empirically by extracting the main components of a factor analysis over adjective descriptors, they are thus considered as the dimensions explaining the most variance of behaviour among people. Table 1 provides a description of each trait.

The Big Five provide a concise framework for characterising the projected personality of conversational agents. A second advantage of this framework is that a large number

of validated questionnaires can be used to evaluate human perceptions of system generated utterances (see evaluation section). Furthermore, psychologists have established the relation between the Big Five and other dimensions of expression variation, such as emotion. For example, there are strong relations between the extraversion and conscientiousness traits and the positive affects, and between neuroticism and disagreeableness and various negative affects (Watson & Clark, 1992). Finally, psychologists have documented a number of behavioural markers associated with each dimension, involving many aspects of communication such as language, speech, gesture and facial display. For example, previous research shows that extraverts talk more, louder,

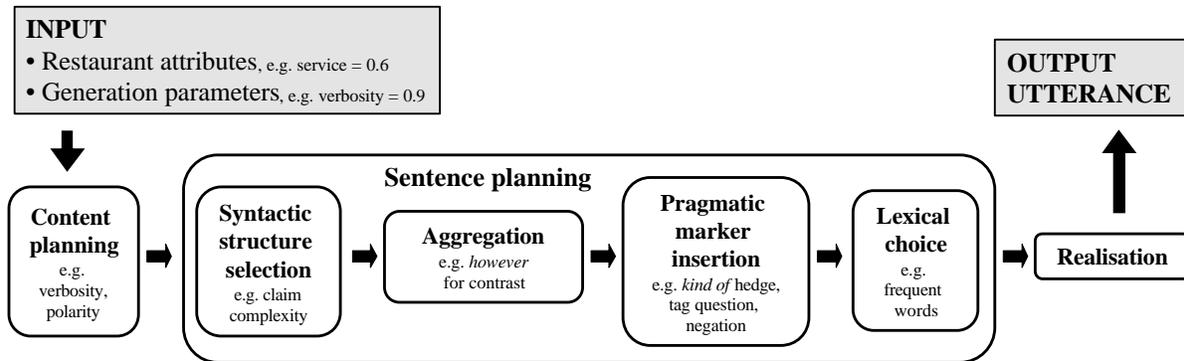


Figure 1: The architecture of the PERSONAGE generator.

with more repetitions, positive words, faster movements, a firmer handshake and a more attractive smile, whereas neurotics produce more self-references, negative emotion words and filled pauses (Chaplin *et al.*, 2000; Dong *et al.*, 1999; Doucet & Stelmack, 1997; Furnham, 1990; Gill & Oberlander, 2003; Pennebaker & King, 1999; Scherer, 1979). See how similar utterances can project different personality traits in Table 2. Thus, while there are other approaches to the generation of individual or stylistic differences (Bouayad-Agha, Scott, & Power, 2000; Cassell & Bickmore, 2003; Hovy, 1988; Walker, Cahn, & Whittaker, 1997), we would argue that the psychology literature provides a scientifically-motivated mapping between personality traits and various expressive behaviours, that can help system developers to project a desired personality successfully.

Below we describe our framework for generating language varying along the Big Five dimensions, and its application to two different domains: (1) restaurant recommendations and (2) interactive narrative in *Façade* (Mateas & Stern, 2003). The next section describes linguistic markers of personality that were previously identified in the psychology literature, and their implementation in PERSONAGE, a personality generator that was shown to produce recognisable variation along the extraversion scale (Mairesse & Walker, 2007). This paper extends our methodology to *all* Big Five traits, with a human evaluation reported in Section 3. Our results show that PERSONAGE produces recognisable personality variation for all personality traits. Additionally, Section 4 shows a dialogue from *Façade* modified by PERSONAGE to vary the characters' personality, showing that some types of stylistic variation can be automatically obtained in a new domain for different levels of utterance representation.

PERSONAGE: a personality generator

Our spoken language generator (SLG) consists of the following modules (Reiter & Dale, 2000):

1. Content planning: refine communicative goals, select and structure content;
2. Sentence planning: choose linguistic resources (lexicon, syntax) to achieve goals;
3. Realisation: use grammar (syntax, morphology) to generate surface utterances.

The high-level architecture of the base generator is illustrated in Figure 1, and its generation parameters are shown in Table 3. The first component is the content planner, which specifies the structure of the information to be conveyed by processing a content plan such as in Figure 2. The content plan is a tree in which the utterance's propositions—e.g. `assert(best(X))`—are combined using rhetorical relations (e.g. JUSTIFY). The resulting content plan tree is then processed by the sentence planner, which selects syntactic templates for expressing individual propositions and aggregates them to produce the utterance's full syntactic structure. The pragmatic marker insertion component then modifies the syntactic structure to produce various pragmatic effects, depending on the parameter values and the markers' insertion constraints. The next component selects the most appropriate lexeme for each content word, given the lexical selection parameters. Finally, the RealPro realiser (Lavoie & Rambow, 1997) converts the final syntactic structure into a string by applying surface grammatical rules, such as morphological inflection and function word insertion. See previous work for a detailed description of the implementation of each component (Mairesse & Walker, 2007).

Relations:	JUSTIFY (nuc:1, sat:2); JUSTIFY (nuc:1, sat:3); JUSTIFY (nuc:1, sat:4); JUSTIFY (nuc:1, sat:5); JUSTIFY (nuc:1, sat:6); JUSTIFY (nuc:1, sat:7)
Content:	1. <code>assert(best (Chimichurri Grill))</code> 2. <code>assert(is (Chimichurri Grill, cuisine (Latin American)))</code> 3. <code>assert(has (Chimichurri Grill, food-quality (decent)))</code> 4. <code>assert(has (Chimichurri Grill, service (decent)))</code> 5. <code>assert(has (Chimichurri Grill, atmosphere (bad)))</code> 6. <code>assert(is (Chimichurri Grill, price (41 dollars)))</code> 7. <code>assert(is (Chimichurri Grill, location(Midtown West)))</code>

Figure 2: A content plan for a recommendation.

Human perceptual evaluation of the Big Five traits

Methodology

Our primary hypothesis is that language generated by varying linguistic parameters identified by research on the Big Five can be recognised as expressing that specific trait. As a

NLG modules	Parameter	Description	
Content selection and structure	VERBOSITY	Controls the number of propositions in the utterance	
	RESTATEMENTS	Paraphrases an existing proposition, e.g. <i>'Chanpen Thai has great service, it has fantastic waiters'</i>	
	REPETITIONS	Repeats an existing propositions	
	CONTENT POLARITY	Controls the polarity of the propositions expressed, i.e. referring to negative or positive attributes	
	REPETITIONS POLARITY	Controls the polarity of the restated propositions	
	CONCESSIONS	Emphasises one attribute over another, e.g. <i>'even if Chanpen Thai has great food, it has bad service'</i>	
	CONCESSIONS POLARITY	Determines whether positive or negative attributes are emphasised	
	POLARISATION	Controls whether the expressed polarity is mild or extreme	
	POSITIVE CONTENT FIRST	Determines whether positive propositions—including the claim—are uttered first	
Syntactic structure selection	SELF-REFERENCES	Controls the number of first person pronouns in the templates	
	SYNTACTIC COMPLEXITY	Controls the syntactic complexity of the templates	
	TEMPLATE POLARITY	Controls the connotation of the claim, i.e. whether positive or negative affect is expressed	
Aggregation operations	RELATIVE CLAUSE	Aggregates propositions with a relative clause, e.g. <i>'Chanpen Thai, which has great service, has nice decor'</i>	
	WITH CUE WORD	Aggregates propositions using <i>with</i> , e.g. <i>'Chanpen Thai has great service, with nice decor'</i>	
	MERGE	Merges the subject and verb of two propositions, e.g. <i>'Chanpen Thai has great service and nice decor'</i>	
	CONJUNCTION CUE WORD	Joins two propositions using a conjunction, or a comma if more than two propositions	
	HOWEVER CUE WORD	Contrasts two propositions using the <i>however</i> cue word	
	ALTHOUGH CUE WORD	Concedes a proposition using <i>although</i> , e.g. <i>'Although Chanpen Thai has great service, it has bad decor'</i>	
	PERIOD	Leaves two propositions in their own sentences, e.g. <i>'Chanpen Thai has great service. It has nice decor.'</i>	
	...		
Pragmatic marker insertion	SUBJECT IMPLICITNESS	Makes the restaurant implicit by moving the attribute to the subject, e.g. <i>'the service is great'</i>	
	NEGATION	Negates a verb by replacing its modifier by its antonym, e.g. <i>'Chanpen Thai doesn't have bad service'</i>	
	SOFTENER HEDGES: ·SORT OF, KIND OF, LIKE, SOME-WHAT, QUITE, RATHER, I THINK THAT, IT SEEMS TO ME THAT	Inserts syntactic elements mitigating the strength of a proposition, e.g. <i>'Chanpen Thai has kind of great service'</i> or <i>'It seems to me that Chanpen Thai has rather great service'</i>	
	EMPHASISER HEDGES: ·REALLY, BASICALLY, ACTUALLY, JUST, EXCLAMATION	Inserts syntactic elements increasing the strength of a proposition, e.g. <i>'Chanpen Thai has really great service!'</i> or <i>'Basically, Chanpen Thai just has great service'</i>	
	ACKNOWLEDGMENT MARKERS: ·YEAH, RIGHT, OK, I SEE, WELL	Inserts a back-channel at the beginning of the utterance, e.g. <i>'Well, Chanpen Thai has great service'</i>	
	FILLED PAUSES: ·I MEAN, ERR, MMHM, YOU KNOW	Inserts syntactic elements expressing hesitancy, e.g. <i>'I mean, Chanpen Thai has great service, you know'</i> or <i>'Err... Chanpen Thai has, like, great service'</i>	
	EXPLETIVES	Inserts swear words, e.g. <i>'the service is damn great'</i>	
	NEAR-EXPLETIVES	Inserts near-swear words, e.g. <i>'the service is darn great'</i>	
	COMPETENCE MITIGATION	Mitigates the speaker's interest for the hearer's request, e.g. <i>'everybody knows that ...'</i>	
	TAG QUESTION	Inserts a tag question, e.g. <i>'the service is great, isn't it?'</i>	
	STUTTERING	Duplicates the first letters of a restaurant's name, e.g. <i>'Ch-ch-anpen Thai is the best'</i>	
	REQUEST CONFIRMATION	Begins the utterance with a confirmation of the restaurant's name, e.g. <i>'did you say Chanpen Thai?'</i>	
	INITIAL REJECTION	Begins the utterance with a mild rejection, e.g. <i>'I'm not sure'</i>	
	IN-GROUP MARKER	Refers to the hearer as a member of the same social group, e.g. <i>pal, mate and buddy</i>	
	Lexical choice	LEXICON FREQUENCY	Controls the average frequency of use of each content word, according to BNC frequency counts
		WORD LENGTH	Controls the average number of letters of each content word
VERB STRENGTH		Controls the strength of the selected verbs, e.g. <i>'I would suggest'</i> vs. <i>'I would recommend'</i>	

Table 3: PERSONAGE's generation parameters, implementation details can be found in (Mairesse & Walker, 2007).

proof of concept, we first test the hypothesis that personality can be made manifest in evaluative speech acts that *recommend* or *compare* domain entities, such as restaurants or movies (Isard, Brockmann, & Oberlander, 2006; Stent, Prasad, & Walker, 2004). The first instantiation of our generator targets a database of restaurants in New York City.¹ There are eight attributes for each restaurant: the name and address, scalar attributes for *price*, *food quality*, *atmosphere*, and *service* and categorical attributes for *neighbourhood* and *type of cuisine*.

Two expert judges evaluate a set of generated utterances as if they had been uttered by a friend responding in a di-

ologue to a request to recommend restaurants. These utterances systematically manipulate parameters related to each trait, e.g. one utterance is generated with the neurotic parameter setting, another with the emotionally stable parameters. We set ourselves the hard task of asking judges to make perception judges of personality on the basis of a single utterance, although in most applications, personality perceptions could emerge over the course of the dialogue. The judges rate the personality of each utterance by completing the Ten-Item Personality Inventory, as this instrument was shown to be psychometrically superior to a 'single item per trait' questionnaire (Gosling, Rentfrow, & Swann, 2003). The answers are averaged to produce a rating for each trait ranging from 1 (e.g. highly neurotic) to 7 (e.g. very emo-

¹An online demo is available at www.dcs.shef.ac.uk/cogsys/personage.html

tionally stable). The judges rated 200 utterances, grouped into 20 sets of 10 utterances generated from the same content plan, each set containing 2 utterances per trait, with parameter settings for both the low end and the high end of each dimension. Thus, there were a total of 40 utterances for each trait, with each half targeting one extreme of the trait dimension. Table 2 shows the 10 utterances generated from the content plan in Figure 2. The judges rated one randomly ordered set at a time, but viewed all 10 utterances in that set before rating them. To avoid replication of generation decisions, multiple outputs were generated by allowing each parameter setting to be normally distributed with a 15% standard deviation.

Results

To assess whether PERSONAGE generates language that can be recognised as projecting extreme personality, we compare the average ratings of the 20 utterances expressing the low end of each trait and the 20 utterances expressing the high end. Table 4 shows the average personality ratings for each set of utterances, on a scale from 1 to 7. Paired t-tests show that the judges can discriminate between both extreme utterance sets for each trait ($p < .001$). Introvert utterances have an average rating of 2.96 out of 7 while extravert utterances have an average rating of 5.98 (difference of 3.02). Additionally, if we divide the data into two equal-width bins around the neutral extravert rating (4 out of 7), then PERSONAGE’s utterance ratings fall in the bin predicted by the parameter set 89.2% of the time. Figure 3 shows the distribution of the extraversion ratings over the two sets.

The rightmost column in Table 4 shows that the judges agree significantly for all Big Five traits. The highest agreement occurs for extravert and introvert utterances ($r = .73$), and the lowest for utterances expressing conscientiousness ($r = .42$). Such a level of agreement reflects the difficulty of estimating the speaker’s personality through a single utterance.

Emotional stability is the most recognisable trait after extraversion, with a mean rating difference of 2.67 between neurotic and stable utterances. Openness to experience is the hardest trait to project in our domain, with a rating difference of 1.32 between the utterance sets. While the rating distributions of most traits tends to be more positive than the median, ratings for openness to experience are lower on average, showing that PERSONAGE’s utterances tend to project the image of a slightly narrow-minded speaker. While PERSONAGE’s utterances cover the full range for each personality trait (see Figure 3), a recalibration of the parameter settings to attenuate those biases is left to future work.

Generating personality in an interactive narrative

While the previous section shows that recognisable personality can be expressed in the information presentation domain, our intention is that PERSONAGE should be a tool that can be used in other domains, so most of the generation decisions it makes are domain-independent. In this section, we provide examples suggesting that PERSONAGE can project various personality types in a sample dialogue

Personality trait	Low	High	r
Extraversion	2.96	5.98	.73
Emotional stability	3.29	5.96	.67
Agreeableness	3.41	5.66	.54
Conscientiousness	3.71	5.53	.42
Openness to experience	2.89	4.21	.44

Table 4: Average personality ratings for the utterances generated with the low and high parameter settings for each trait on a scale from 1 to 7 (columns 2 and 3), as well as the interrater correlation r . The ratings of the two extreme utterance sets differ significantly for all traits ($p < .001$, two-tailed), and the correlations are all significant at the $p < .05$ level.

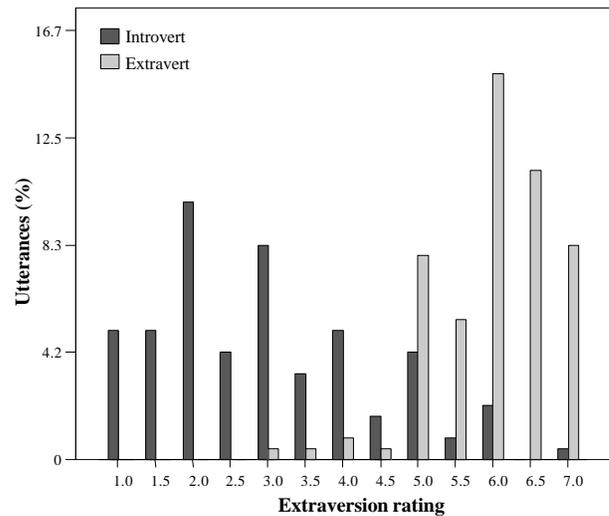


Figure 3: Distributions of individual ratings for introvert and extravert utterances.

from the Façade interactive narrative system (INS) (Mateas & Stern, 2003).

An important choice faced by INS developers is the level of abstraction of the manipulated language. Façade currently produces pre-recorded prompts, i.e. the variation is hand-coded in each utterance. We investigate deeper representations by manually converting dialogue turns into syntactic structures and modifying them using PERSONAGE’s pragmatic insertion marker and lexical choice components. Tables 5 and 6 show a series of modified dialogues between Trip and Grace, a married couple who the player is visiting for an evening. Each column represents the same initial dialogue, in which both speakers were assigned a different personality.

Although there is no generation at the content and aggregation levels, the amount of variation is still substantial, see Trip changing from an extravert in the third column of Table 5, i.e. ‘*Yeah, Grace’s decorating has been such like, a really great outlet for her, you know!*’, to a more serious conscientious person in the third column of Table 6, i.e. ‘*I see, ok, I suppose Grace’s decorating has been such a rather outstanding outlet for her*’. On the other end of the same

Speaker	Original dialogue	TRIP: extravert GRACE: introvert	TRIP: emotionally stable GRACE: neurotic	TRIP: agreeable GRACE: disagreeable
USER	· Do you like this place?	· Do you like this place?	· Do you like this place?	· Do you like this place?
TRIP	· Yeah, you know, Grace's decorating has been such a great outlet for her.	· Yeah, Grace's decorating has been such like, a really great outlet for her, you know!	· Basically, Grace's decorating has been such a kind of outstanding outlet for her.	· Well, yeah, Grace's decorating has been such a sort of outstanding outlet for her mate, you know, alright?
GRACE	· Outlet? What do you mean outlet?	· Outlet? Err... mmhm... what do you mean by an outlet?	· I mean, actually, I am not really sure. Outlet? Err... what do you mean by an ou-outlet?	· Outlet? What I mean, do you mean by an outlet?
TRIP	· And I am always encouraging her to buy. Only the best for us, right Grace?	· Actually, I am always encouraging her to buy buddy, you know! Only the best for us.	· Basically, I am always encouraging her to buy. Only the best for us.	· Ok, right, I am always encouraging her to buy, you know mate, okay? Only the best for us.
GRACE	· Trip, let's not get too carried away here.	· Well, right, let's not get too quite carried away here.	· Actually, I am not re-really sure! Err... I mean, let's not get too carried away here.	· Oh God I mean, let's not get too quite carried away here.
TRIP	· Aha! Great! I will fix us some drinks. Good news, I just fully loaded the bar, with the real quality.	· Oh gosh actually, I will fix us some drinks mate, wouldn't you? I fully just loaded the bar with really kind of real quality.	· I will fix us some drinks. Basically, I fully just loaded the bar with sort of real quality.	· Ok, oh I will fix us some drinks buddy, wouldn't you? I fully just loaded the bar with rather real quality, you know.
GRACE	· Trip, please... let's not go overboard with the drink preparation. The funny thing is, Trip doesn't even like the taste of alcohol.	· Well, i see, let's not go sort of overboard with the drink preparation. I guess the rather funny thing is, Trip doesn't even like the taste of alcohol.	· I don't know! Ac-ac-actually, let's not go really overboard with the drink preparation. I mean, the funny thing is err... Trip doesn't even like the taste of alcohol.	· Actually, basically, let's not go overboard with the drink preparation. I mean, the funny thing is mmhm... Trip doesn't even like the taste of alcohol.
TRIP	· How about you Grace, can I pour you a glass? I assume you want your usual... 'a lovely, very cold glass of Chardonnay'.	· How about you Grace. Can I pour you a glass? I assume you want like, a really lovely very cold glass of Chardonnay, pal, you know!	· How about you Grace. Can I pour you a glass? Basically, I assume you want a quite lovely very sort of cold glass of Chardonnay.	· How about you Grace. Can I pour you a glass? I assume you want a sort of lovely very quite cold glass of Chardonnay buddy, you know, alright?
GRACE	· Um, I'd like a simple glass of Chardonnay, thank you.	· Err... right, I would like a sort of simple glass of Chardonnay.	· I don't know! Err... I mean, I would like like, a really simple glass of Chardonnay.	· Actually, I mean, I would like a kind of simple glass of Chardonnay.

Table 5: Example of PERSONAGE's outputs in the interactive narrative domain for extraversion, emotional stability and agreeableness, based on a script from Mateas & Stern's *Façade* (2003). The sample dialogue was manually converted into a sequence of syntactic structures, and processed by PERSONAGE's pragmatic marker insertion and lexical choice components (Mairesse & Walker, 2007).

dimensions, Grace produces an introvert answer in Table 5, i.e. 'Outlet? Err... mmhm... what do you mean by an outlet?', whereas her unconscientious answer in Table 6 reflects hesitancy and carelessness, i.e. 'I don't know mate! Outlet? Mmhm... err... what do you mean by an outlet?'. While the perception of the personality in this new domain remains to be evaluated, we believe these example dialogues provide an argument for automating variation in INS, by manipulating deep utterance representations.

As *Façade*'s original characters already have specific personalities (see leftmost dialogue in Table 5), a first step was to remove any marker of personality to produce neutral syntactic structures. Personality is also marked at the content level, e.g. 'Trip, let's not get too carried away here', which can lead to unnatural utterances when inconsistent markers are introduced. As a consequence, we believe projected personality needs to be controlled at all levels of language gen-

eration to produce consistent outputs.

Discussion

Our results show that PERSONAGE can produce recognisable personality variation for all of the Big Five personality traits, while in previous work we only evaluated the projection of extraversion (Mairesse & Walker, 2007). Thus PERSONAGE is the first system to clearly demonstrate that the Big Five personality traits can be manifested by an automatic generator, to provide system utterances in dialogue applications that require social intelligence. In previous work, Walker, Cahn, & Whittaker (1997) describe a generator intended to affect perceptions of personality, based on Brown and Levinson's theory of politeness (Brown & Levinson, 1987). The generator uses some of the linguistic constructions implemented here, such as tag questions

Speaker	Original dialogue	TRIP: conscientious GRACE: unconscious	TRIP: open to experience GRACE: not open to experience
USER	· Do you like this place?	· Do you like this place?	· Do you like this place?
TRIP	· Yeah, you know, Grace's decorating has been such a great outlet for her.	· I see, ok, I suppose Grace's decorating has been such a rather outstanding outlet for her.	· I see, I suppose Grace's decorating has been such a sort of outstanding outlet for her, you know, you see?
GRACE	· Outlet? What do you mean outlet?	· I don't know mate! Outlet? Mhm... err... what do you mean by an outlet?	· I mean, basically, I am not sure. Outlet? Err... mhm... what do you mean by an outlet?
TRIP	· And I am always encouraging her to buy. Only the best for us, right Grace?	· Well, i see, I am always encouraging her to buy. Only the best for us.	· I see, mhm... I am always encouraging her to buy, you know, okay? Only the best for us.
GRACE	· Trip, let's not get too carried away here.	· Yeah, I mean, let's not get too rather carried away here mate!	· Err... mhm... I am not sure. I mean, let's not get too carried away here.
TRIP	· Aha! Great! I will fix us some drinks. Good news, I just fully loaded the bar, with the real quality.	· I see, I will fix us some drinks. I fully just loaded the bar with rather real quality.	· I am not kind of sure. I will fix us some drinks, you know, you see? I fully just loaded the bar with quite real quality.
GRACE	· Trip, please... let's not go overboard with the drink preparation. The funny thing is, Trip doesn't even like the taste of alcohol.	· I mean, I am not sure. Mhm... let's not go overboard with the drink preparation! The kind of funny thing is, Trip doesn't even like the taste of alcohol, pal.	· I am not sure. Err... I mean, let's not go overboard with the drink preparation. The funny thing is mhm... Trip doesn't even like the taste of alcohol.
TRIP	· How about you Grace, can I pour you a glass? I assume you want your usual... 'a lovely, very cold glass of Chardonnay'.	· How about you Grace. Can I pour you a glass? I assume you want a sort of lovely very quite cold glass of Chardonnay.	· How about you Grace. Can I pour you a glass? I assume you want a quite lovely very sort of cold glass of Chardonnay, you know, okay?
GRACE	· Um, I'd like a simple glass of Chardonnay, thank you.	· I mean, I am not sure. Err... I would like a kind of simple glass of Chardonnay, you know buddy.	· Mhm... I mean, I don't know. Err... I would like like, a simple glass of Chardonnay.

Table 6: Example of PERSONAGE's outputs in the interactive narrative domain for conscientiousness and openness to experience, based on a script from Mateas & Stern's *Façade* (2003).

and hedges, but it was never evaluated. Gupta, Walker, & Romano (2007) present and evaluate a generator based on politeness theory and show that humans predictably perceive differences in politeness. Research by André *et al.* (2000) and Piwek (2003) uses personality variables to affect the linguistic behaviour of conversational agents, but did not systematically manipulate parameters, and their generators were not evaluated. Reeves & Nass (1996) demonstrate that manipulations of personality affect many aspects of user's perceptions, but their experiments use handcrafted utterances, rather than generated utterances. Cassell & Bickmore (2003) show that extraverts prefer systems utilising discourse plans that include small talk. Isard, Brockmann, & Oberlander's CRAG-2 system (2006) overgenerates and ranks using ngram language models trained on a corpus labelled for all Big Five personality dimensions. However, CRAG-2 has no explicit parameter control, and it has yet to be evaluated.

There are several issues to be addressed before automatic generation can be applied successfully in a range of dialogue applications. One limitation is that so far we have focused only on textual language generation, whereas it would be possible to use the Big Five traits to control other aspects of expressive behaviour (e.g. gesture, dialogue strategy, speech rate, prosody and amplitude). A major issue is the reliance of automated generation on the quality of the text-to-speech

(TTS) engine. Only recently has research begun to build off-the-shelf high-quality synthesisers with a large range of speech variation in mind (Trouvain *et al.*, 2006). However there are many markers of personality in speech (Scherer, 1979), thus our approach could be applied to a customisable TTS engine, e.g. associating extraversion with higher amplitude and higher speech rate. While some speech markers can be controlled directly by the TTS engine, others require a careful implementation, such as the variation of pitch across an utterance or the prosody of individual words (e.g. filled pauses). Nevertheless, the current rate of progress in the field suggests that TTS engines will express realistic affect and personality in a short time frame.

Another issue is that PERSONAGE assumes that the communicative goal is given to it as input, and that the expression of this goal is supported by a generation dictionary. Thus PERSONAGE has not been integrated with a planner, and the initial dictionary must be handcrafted. Other research has started to address these limitations (Higashinaka, Prasad, & Walker, 2006; Riedl & Young, 2004).

References

- Allport, G. W., and Odbert, H. S. 1936. Trait names: a psychological study. *Psychological Monographs* 47(1, Whole No. 211):171-220.

- Allport, G. W. 1960. *Personality and social encounter*. Boston, MA: Beacon.
- André, E.; Rist, T.; van Mulken, S.; Klesen, M.; and Baldes, S. 2000. The automated design of believable dialogues for animated presentation teams. In S. Prevost, J. Cassell, J. S., and Churchill, E., eds., *Embodied conversational agents*. Cambridge, MA: MIT Press. 220–255.
- Bouayad-Agha, N.; Scott, D.; and Power, R. 2000. Integrating content and style in documents: a case study of patient information leaflets. *Information Design Journal* 9(2-3):161–176.
- Brown, P., and Levinson, S. 1987. *Politeness: Some universals in language usage*. Cambridge University Press.
- Cassell, J., and Bickmore, T. 2003. Negotiated collusion: Modeling social language and its relationship effects in intelligent agents. *User Modeling and User-Adapted Interaction* 13:89–132.
- Chaplin, W. F.; Phillips, J. B.; Brown, J. D.; Clanton, N. R.; and Stein, J. L. 2000. Handshaking, gender, personality, and first impressions. *Journal of Personality and Social Psychology* 79(1):110–117.
- Dong, J. K.; Jin, T. H.; Cho, H. W.; and Oh, S. C. 1999. The esthetics of the smile: a review of some recent studies. *International Journal of Prosthodontics* 12(1):9–19.
- Doucet, C., and Stelmack, R. M. 1997. Movement time differentiates extraverts from introverts. *Personality and Individual Differences* 23(5):775–786.
- Furnham, A. 1990. Language and personality. In Giles, H., and Robinson, W., eds., *Handbook of Language and Social Psychology*. Winley.
- Gill, A., and Oberlander, J. 2003. Perception of e-mail personality at zero-acquaintance: Extraversion takes care of itself; neuroticism is a worry. In *Proceedings of the 25th Annual Conference of the Cognitive Science Society*, 456–461.
- Goldberg, L. R. 1990. An alternative “description of personality”: The Big-Five factor structure. *Journal of Personality and Social Psychology* 59:1216–1229.
- Gosling, S. D.; Rentfrow, P. J.; and Swann, W. B. 2003. A very brief measure of the big five personality domains. *Journal of Research in Personality* 37:504–528.
- Gupta, S.; Walker, M.; and Romano, D. M. 2007. How rude are you?: Evaluating politeness and affect in interaction. In *Proceedings of ACII*, 203–217.
- Higashinaka, R.; Prasad, R.; and Walker, M. 2006. Learning to Generate Naturalistic Utterances Using Reviews in Spoken Dialogue Systems. In *Proceedings of ACL 2006*.
- Hovy, E. 1988. *Generating Natural Language under Pragmatic Constraints*. Lawrence Erlbaum Associates.
- Isard, A.; Brockmann, C.; and Oberlander, J. 2006. Individuality and alignment in generated dialogues. In *Proceedings of the 4th International Natural Language Generation Conference (INLG-06)*, 22–29.
- Isbister, K., and Nass, C. 2000. Consistency of personality in interactive characters: verbal cues, non-verbal cues, and user characteristics. *International Journal of Human-Computer Studies* 53(2):251 – 267.
- Lavoie, B., and Rambow, O. 1997. A fast and portable realizer for text generation systems. In *Proceedings of the 3rd Conference on Applied Natural Language Processing*, 265–268.
- Mairesse, F., and Walker, M. 2007. PERSONAGE: Personality generation for dialogue. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL)*, 496–503.
- Mateas, M., and Stern, A. 2003. Façade: An experiment in building a fully-realized interactive drama. In *Proceedings of the Game Developers Conference, Game Design track*.
- Mateas, M. 2007. The authoring bottleneck in creating AI-based interactive stories. In *Proceedings of the AAAI 2007 Fall Symposium on Intelligent Narrative Technologies*.
- McCrae, R. R., and Costa, P. T. 1987. Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology* 52:81–90.
- Norman, W. T. 1963. Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality rating. *Journal of Abnormal and Social Psychology* 66:574–583.
- Pennebaker, J. W., and King, L. A. 1999. Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology* 77:1296–1312.
- Pieraccini, R., and Levin, E. 1995. A learning approach to natural language understanding. In *Speech Recognition and Coding, New Advances and Trends, NATO ASI Series*. Springer Verlag. 139–155.
- Piwek, P. 2003. A flexible pragmatics-driven language generator for animated agents. In *Proceedings of EACL*.
- Reeves, B., and Nass, C. 1996. *The Media Equation*. University of Chicago Press.
- Reiter, E., and Dale, R. 2000. *Building Natural Language Generation Systems*. Cambridge University Press.
- Riedl, M., and Young, R. M. 2004. An intent-driven planner for multi-agent story generation. In *Proceedings of the 3rd International Conference on Autonomous Agents and Multi Agent Systems*.
- Scherer, K. R. 1979. Personality markers in speech. In Scherer, K. R., and Giles, H., eds., *Social markers in speech*. Cambridge University Press. 147–209.
- Stent, A.; Prasad, R.; and Walker, M. 2004. Trainable sentence planning for complex information presentation in spoken dialogue systems. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL)*.
- Trouvain, J.; Schmidt, S.; Schröder, M.; Schmitz, M.; and Barry, W. J. 2006. Modelling personality features by changing prosody in synthetic speech. In *Proceedings of Speech Prosody*.
- Walker, M., and Rambow, O. 2002. Spoken language generation. *Computer Speech and Language, Special Issue on Spoken Language Generation* 16(3-4):273–281.
- Walker, M.; Cahn, J. E.; and Whittaker, S. J. 1997. Improvising linguistic style: Social and affective bases for agent personality. In *Proceedings of the 1st Conference on Autonomous Agents*, 96–105.
- Wang, N.; Johnson, W. L.; Mayer, R. E.; Rizzo, P.; Shaw, E.; and Collins, H. 2005. The politeness effect: Pedagogical agents and learning gains. *Frontiers in Artificial Intelligence and Applications* 125:686–693.
- Watson, D., and Clark, L. A. 1992. On traits and temperament: General and specific factors of emotional experience and their relation to the five factor model. *Journal of Personality* 60(2):441–76.