User Model User-Adap Inter (2010) 20:227–278
DOI 10.1007/s11257-010-9076-2

ORIGINAL PAPER

Towards personality-based user adaptation: psychologically informed stylistic language generation

François Mairesse · Marilyn A. Walker

Received: 18 June 2009 / Accepted in revised form: 3 June 2010 /
Published online: 3 July 2010

Abstract Conversation is an essential component of social behavior, one of the primary means by which humans express intentions, beliefs, emotions, attitudes and personality. Thus the development of systems to support natural conversational interaction has been a long term research goal. In natural conversation, humans adapt to one another across many levels of utterance production via processes variously described as linguistic style matching, entrainment, alignment, audience design, and accommodation. A number of recent studies strongly suggest that dialogue systems that adapted to the user in a similar way would be more effective. However, a major research challenge in this area is the ability to dynamically generate user-adaptive utterance variations. As part of a personality-based user adaptation framework, this article describes PERSONAGE, a highly parameterizable generator which provides a large number of parameters to support adaptation to a user's linguistic style. We show how we can systematically apply results from psycholinguistic studies that document the linguistic reflexes of personality, in order to develop models to control PERSONAGE's parameters, and produce utterances matching particular personality profiles. When we evaluate these outputs with human judges, the results indicate that humans perceive the personality of system utterances in the way that the system intended.

Keywords Natural language generation · Linguistic style · Personality · Individual differences · Big Five traits · Dialogue · Recommendation

F. Mairesse (✉)
Department of Engineering, University of Cambridge, Trumpington St., Cambridge CB2 1PZ, UK
e-mail: f.mairesse@eng.cam.ac.uk

M. A. Walker
Department of Computer Science, University of California Santa Cruz, 1156 N. High St. SOE-3, Santa Cruz, CA 95064, USA
e-mail: maw@soe.ucsc.edu

1 Introduction

Conversation is an essential component of social behavior, one of the primary means by which humans express intentions, beliefs, emotions, attitudes and personality. Thus systems to support natural conversational interaction have been a long term research goal (Grosz 1983; Power 1974; Cohen et al. 1982; Litman and Allen 1987; Zukerman and Litman 2001; Carberry 1989; Kobsa and Wahlster 1989; Finin et al. 1986). In natural conversation, humans adapt to one another across many levels of utterance production via processes variously described as linguistic style matching, entrainment, alignment, audience design, and accommodation (Niederhoffer and Pennebaker 2002; Brennan 1996; Levelt and Kelter 1982; Pickering and Garrod 2004; Brennan and Clark 1996; Nenkova et al. 2008; Giles et al. 1991). A number of recent studies strongly suggest that dialogue systems that adapted to the user in a similar way would be more effective (Murray 1997; Hayes-Roth and Brownston 1994; André et al. 2000; Cassell and Bickmore 2003; Tapus and Mataric 2008; Mott and Lester 2006; Hirschberg 2008; Reeves and Nass 1996; Brennan 1991; Forbes-Riley and Litman 2007; Forbes-Riley et al. 2008; Stenchikova and Stent 2007; Reitter et al. 2006). Several of these studies provide empirical evidence that adaptation to the conversational partner is also beneficial at the personality level through experiments using hand-crafted utterances designed intuitively to express a particular personality (Reeves and Nass 1996; Tapus and Mataric 2008). These experiments showed that users will spend more time on the task, or that their perceptions of a system's intelligence or competence increase when the systems match the user's personality. Although this *similarity-attraction* effect alone provides motivation for exploring methods for personality-based user adaptation, there is a case for a more general framework in which the system's personality is dependent on both the *user* and the *task*, as we discuss in more detail below.

Figure 1 illustrates how a generic adaptation capability for dialogue systems requires addressing three research problems: (a) acquiring relevant user traits (recognition), (b) deciding what traits should be conveyed by the system (adaptation), and

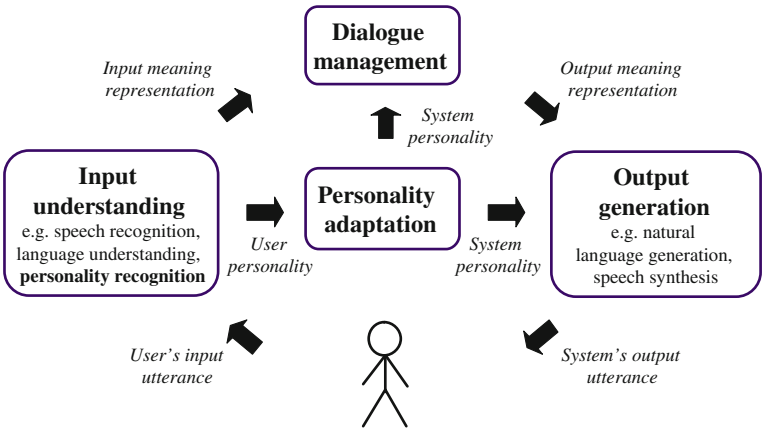


Fig. 1 High-level architecture of a dialogue system with personality-based user adaptation

(c) producing a consistent response matching those traits (generation). In our view, progress in this area has been stymied because without the generation component (c), there is no way to really test whether (a) and (b) are functional. Thus the latter task is the focus of this article, but here we will briefly discuss the first two.

How can we acquire user personality information and why do we think it is an important component of user adaptation? Our framework builds on the "Big Five" model of personality traits. The standard method in the Big Five framework is to assess personality with questionnaires (John et al. 1991; Costa and McCrae 1992; Gosling et al. 2003). The other possibility is to identify relevant behavioral cues, e.g. based on the user's interaction (Dunn et al. 2009) or the user's speech and language (Argamon et al. 2005; Oberlander and Nowson 2006; Mairesse et al. 2007). While personality questionnaires have a high predictive value and only need to be filled once by the user, they lack the objectivity of observer-reports and require a significant effort from the user. In other work, we developed automated methods for recognizing personality from user conversations (Mairesse et al. 2007). Results show that personality recognition models trained on content analysis and prosodic features are significant predictors of the speaker's level of Big Five personality traits such as extraversion, neuroticism, agreeableness and conscientiousness. Similar methods have also been applied successfully to textual content (Argamon et al. 2005; Oberlander and Nowson 2006). Thus, automated recognition methods provide a promising alternative to questionnaires.

Once the personality of the user has been assessed, findings from psychological studies can inform the *personality adaptation model* illustrated in Fig. 1. Table 1 suggests potential personality adaptation policies for different tasks, which remain to be evaluated. For example, we hypothesize that information presentation systems

Table 1 Hypothesized personality adaptation policies for various applications

Task	User type	System adaptation policy
Information presentation system	Novice	Extravert, agreeable
	Experienced	Converge towards user
Tutoring system	Any	Extravert, agreeable, conscientious
Telesales system	Any	Potent (extravert), match the company's brand
Video games/entertainment	Any	Character-based
Crucial information retrieval (e.g. finance)	Any	Conscientious, not extravert, not agreeable
Psychotherapy	Fearful (introvert, not open)	Aggressive (extravert, psychotic)
	Aggressive (extravert, psychotic)	Fearful or aggressive
Psychotherapist training system	Any	Neurotic
	Any	Aggressive (extravert, psychotic)

Specific traits are mapped to the Big Five or PEN framework (Norman 1963; Eysenck et al. 1985)

should produce extravert and agreeable language with novice users, and back-off to a similarity-attraction policy with more advanced users. Previous studies suggest that tutoring systems should be agreeable and extravert, based on findings associating the use of politeness forms with higher learning outcomes (Wang et al. 2005), as well as correlates between extraversion and the performance of human teachers (Rushton et al. 1987). Additionally, we hypothesize that systems providing crucial information—e.g. when requesting stock quotes or emergency advice—should produce outputs that are clear and concise. This suggests the need for a conscientious, introvert and non-agreeable operator, e.g. avoiding superfluous politeness forms. More generally, training systems should convey a large range of personalities. Examples include systems for training practitioners to interview anxious patients (Hubal et al. 2000), as well as systems training soldiers to gather information from uncooperative civilians through tactical questioning (Traum et al. 2007; Department of the Army 2006). Personality modeling has also found applications in virtual reality for psychotherapy, e.g. to reduce the patient's anxiety when interacting with aggressive personalities or when speaking in public (Slater et al. 2006). Task-dependent adaptation can also be beneficial to the *system* as opposed to the user. Furnham et al. (1999) report that potency (extraversion) correlates positively with sales figures and superior ratings, and that impulsivity is a significant performance predictor of telesales employees selling insurance. Such findings can guide system designers to optimize the personality conveyed by an automated sales agent. Additionally, a large body of marketing research shows that consumers associate brands with personality types, and that they tend to select brands conveying traits that are desirable to them (Aaker 1999; Fennis and Pruyn 2007; Plummer 1984). There is thus a strong incentive for companies to tailor the personality of their dialogue system to their target market (Fink and Kobsa 2000; Kobsa 2002).

Given the large number of applications benefiting from the projection of specific personality traits, there is a need for a re-usable and scalable computational framework for generating personality-rich dialogic utterances, which provides a wide selection of relevant parameters. In order to adapt to the user as humans adapt to one another in natural conversation, it is also important to be able to *dynamically* modify parameter values; if dialogue system utterances are handcrafted to portray a particular style, such parameters are not available, nor can they be modified in real time to adapt to a particular user. In this article, we describe a highly parameterizable generator PERSONAGE, which provides a large number of parameters to support adaptation to a user's linguistic style. In order to identify relevant parameters, we systematically organize and utilize findings from the psycholinguistic literature (see Table 27 in the Appendix). The generation parameters that we propose operate across many levels of linguistic production and can support adaptation of content selection, lexical choice, and selection of syntactic and rhetorical structure. They are well-specified in terms of generation decisions that manipulate well-defined syntactic and semantic representations used in many standard NLG architectures. Thus they could be easily implemented in other generators and in other domains, and provide a basis for systematically testing, across domains and applications, which types of stylistic variation affect user perceptions and how. One way to control all these parameters is via a model of the

user (Fink and Kobsa 2000; Kobsa and Wahlster 1989). In this paper, we show how we can build personality models using the psycholinguistic studies detailed in Table 27 to control PERSONAGE's parameters, and produce utterances adapted to particular user models of personality. When we evaluate these outputs with human judges, the results indicate that humans perceive the personality of system utterances in the way that the system intended.

Table 2 shows some example outputs of PERSONAGE. In Table 2 the *set* column indicates whether the output utterance was based on a personality model for the low end of a trait (introversion) versus the high end of a trait (extraversion). The examples in Table 2 manipulate parameters such as verbosity (verbal fluency, study SP65 in Table 27), polarity of the content selected (polarity, study TH87 in Table 27), and the occurrence of hedges or markers of tentativeness (content analysis category counts, studies PK99, OG06, and ME06 in Table 27). The *score* column of Table 2 shows the average score of three judges when asked to assess the personality of the speaker of the utterance using the Ten Item Personality Inventory of Gosling et al. (2003). We explain in detail in this paper how we generate such utterances and how we collect human judgments to evaluate our framework.

Previous research on the generation of linguistic variation includes both rule-based and statistical approaches, as well as hybrid methods that combine rule-based linguistic knowledge with statistical methods (Langkilde and Knight 1998; Langkilde-Geary 2002). This includes work on variation using parameters based on pragmatic effects (Hovy 1988; Fleischman and Hovy 2002), stylistic factors such as formality, sentence length, and syntactic structure (Power et al. 2003; Bouayad-Agha et al. 2000; DiMarco and Hirst 1993; Green and DiMarco 1996; Paiva and Evans 2005; Belz 2005; Walker et al. 2002; Paris and Scott 1994), emotion (Cahn 1990), lexical choice (Inkpen and Hirst 2004), user expertise or confidence (Porayska-Pomsta and Mellish 2004; Wang et al. 2005; DiMarco and Hirst 1993; Forbes-Riley and Litman 2007; Forbes-Riley et al. 2008), a theory of linguistic politeness (Brown and Levinson 1987; Walker et al. 1997; Gupta et al. 2007, 2008; Wilkie et al. 2005; Porayska-Pomsta and Mellish 2004; Wang et al. 2005), theories of personality (Ball and Breese 1998; André et al. 2000; Loyall and Bates 1995; Isard et al. 2006), and individual differences and preferences for both style and content (Stent et al. 2004; Walker et al. 2007; Reiter and Sripada 2002; Lin 2006; Belz 2008). While there are strong relations between these different notions of style, and the types of linguistic variation associated with personality factors, here we limit our detailed discussion of prior work to personality generation. In Sect. 6, we will discuss how, in future work, PERSONAGE could be used to generate different types of stylistic variation.

Previous work on personality generation has primarily been associated with embodied conversational agents (ECAs). This research is very useful for identifying applications of personality generation, and showing how to integrate personality generation at the textual level with other modalities such as gesture and prosody (Rehm and André 2008). While we do not know of any studies using ECAs that evaluate whether the personality of generated utterances is perceived by human users as the ECA intended, some of this work has shown an effect on task-related metrics, such as user satisfaction or perceptions of system competence (Reeves and Nass 1996; Isbister and Nass 2000; Tapus and Mataric 2008). In contrast to our approach, the work on ECAs has typically

Table 2 Example outputs of PERSONAGE with average judges ratings on the corresponding personality dimension (see Sect. 4)

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	Chimichurri Grill is a Latin American restaurant, also it's located in Midtown West. It has quite friendly waiters. It offers adequate food. I imagine you would appreciate it	2.92
	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. Mmhm… 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

Personality ratings are on a scale from 1 to 7, with 1 = very low (e.g. introvert) and 7 = very high (e.g. extravert)

modeled the generation task using templates, which have been labeled as expressing a particular personality, rather than by manipulating parameters within modules of the NLG pipeline.

Loyall and Bates (1995) is one of the first papers to suggest the use of personality models for language generation in ECAs. They present a model where personality

factors are integrated with emotions, intentions and desires, and use template-based generation indexed by personality variables. Ball and Breese (1998) model the effect of the agent's personality (i.e. dominance and friendliness) and emotions (i.e. valence and arousal) on its behavior. The personality values affect a layer of variables determining the paraphrase template to be selected by the system, such as the language strength, positivity and terseness. Scripted dialogue is another venue for modeling the personality of multiple conversational agents. André et al. (2000) provide a system where the agents' utterances can be modified by selecting different values for extraversion, agreeableness and openness to experience (Rehm and André 2008; André et al. 2000). Templates are annotated with intermediary variables (e.g. force) which in turn are associated with the personality traits (e.g. extravert agents use more forceful language, and they show more initiative in dialogue), and with gesture and facial expression. Lester et al. (1999b,a) use handcrafted models of personality and emotion in pedagogical applications to teach children about science, and suggest that children become much more engaged in learning when the pedagogical agents exhibit colorful personalities and express emotions. Ruttkay et al. (2004) suggest that personality is an important design variable for developing embodied conversational agents. The NECA system is a multimodal language generator that models pragmatic effects and personality (Piwek 2003), in which information about the character's personality is passed from one module to the other in order to produce consistent behavior across modes (e.g. language, speech and gesture), while the way personality affects language is encoded in a generation grammar. Cassell and Bickmore (2003) extend their REA real estate agent with smalltalk generation capabilities, which is hypothesized to increase the user's trust in the system. Interestingly, they observe large perceptual variations between user groups with different personalities. Extravert users feel that they know REA better if she produces social language, resulting in a more satisfying interaction. On the other hand, introvert users are much less affected by REA's smalltalk, and rate that version of REA lower.

The most closely related work to this paper is the CRAG- 2 system, which extends HALOGEN's methodology (Langkilde-Geary 2002) to model personality and alignment in dialogue (Isard et al. 2006; Brockmann 2009). CRAG- 2 ranks a set of candidate utterances based on a linear combination of n-gram models, including a general-domain model trained on conversations from the Switchboard corpus, and models trained on a corpus of weblogs labeled with the author's personality. The system models linguistic alignment using a cache language model that primes particular syntactic forms on the basis of the conversational partner's previous utterance. This work is the first to combine personality control and alignment within the same framework. Brockmann (2009) shows that the alignment model affects personality perceptions of agreeableness and reduces the overall interaction quality, thus illustrating the trade-off between (a) benefits of the similarity attraction effect and (b) task-dependent personality requirements such as those presented in Table 1. The main difference between the current work and the CRAG- 2 system lies in (a) the motivation for choosing generation parameters, (b) the range of parameters controlled, and (c) the generation methodology being used. First, while the variation in CRAG- 2 is produced by 10 heuristic parameters at the realization level (i.e., inserting pragmatic markers such as 'I mean', 'well' or 'basically'),

this article puts forward a systematic framework consisting of 67 *psychologically-motivated* parameters that can be used to generate language manifesting personality at *all* levels of the generation process. As the number of generation parameter increases, the interaction between surface realization rules becomes prohibitive. The modularity offered by the NLG pipeline (Reiter and Dale 2000) allows the combination of many global (e.g., sentence planning) and local (e.g., lexical choice) parameters to produce a larger range of outputs than template-based approaches. Secondly, the CRAG- 2 system uses an overgenerate and rank approach, by reranking utterances based on surface features which can result from multiple generation decisions. The computational cost of the overgeneration phase grows exponentially with the number of control parameters. As a result, the size of the parameter space that can be explored by such methods is limited, especially for real-time dialogues.

Section 2 describes the PERSONAGE base generator and all of its parameters. Section 3 presents a method for controlling these parameters, by developing personality models that target the ends of each personality trait scale. In order to test our framework, we instantiate it in a particular discourse situation and domain, namely producing recommendations in the restaurant domain. Sections 4 and 5 describe our evaluation experiment and present the results. Our evaluation metric is based on a standard personality measurement instrument (Gosling et al. 2003). Section 5.1 reports results showing that the judges agree significantly on their perceptions. Section 5.2 presents the correlations between PERSONAGE's linguistic parameters and personality ratings in order to test precisely which parameters are affecting user perceptions, and to test whether findings from previous work generalize to our domain and discourse situation. Results show that linguistic reflexes documented in naturally occurring genres can be manipulated in a language generator and that those reflexes in many cases have the same effect on perception of personality. We sum up and discuss future work in Sect. 6.

2 The PERSONAGE generator

The Big Five model is based on the observation that, when talking about a close friend, one can usually produce a large number of descriptive adjectives (Allport and Odbert 1936). This observation is described as the *Lexical Hypothesis*, i.e. that any trait important for describing human behavior has a corresponding lexical token, which is typically an adjective, such as *trustworthy, modest, friendly, spontaneous, talkative, dutiful, anxious, impulsive, vulnerable*. The Lexical Hypothesis led to a consensus that there are essential traits, known as the *Big Five* personality traits (Norman 1963; Peabody and Goldberg 1989; Goldberg 1990; Revelle 1991; Bouchard and McGue 2003). These traits (see Table 3) are *extraversion*, *emotional stability*, *agreeableness*, *conscientiousness* and *openness to experience*.

The Big Five model has several advantages as the basis of a computational framework for generating variation in linguistic style. There are a large number of useful prior studies (Pennebaker and King 1999; Thorne 1987; Mehl et al. 2006; Oberlander and Gill 2006), that carefully document correlations between Big Five traits and linguistic behavior (measured via lexical category, word or syntactic structure counts). These correlations suggest a large number of relevant parameters for generation.

Table 3 Example adjectives associated with the Big Five 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, friendly, considerate, generous, helpful, altruistic	Unfriendly, selfish, suspicious, uncooperative, malicious
Conscientiousness	Competent, disciplined, dutiful, achievement striving, deliberate, careful, orderly	Disorganized, impulsive, unreliable, careless, forgetful
Openness to experience	Creative, intellectual, imaginative, curious, cultured, complex	Narrow-minded, conservative, ignorant, simple

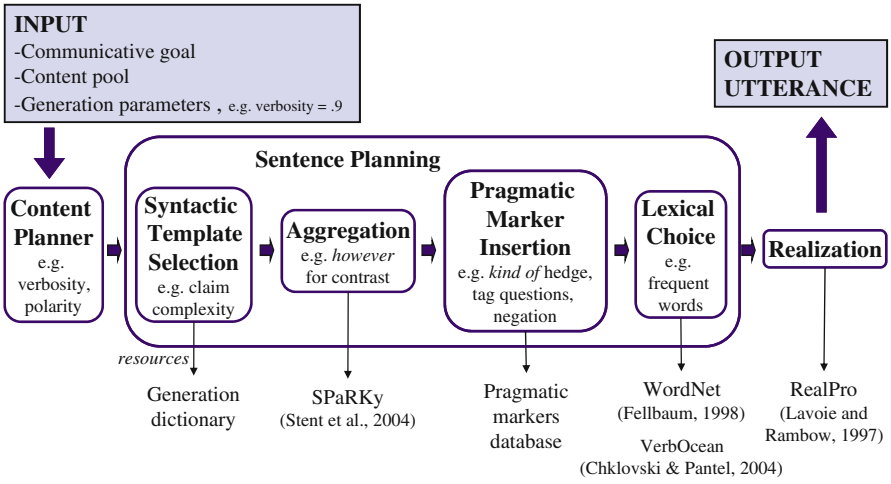


Fig. 2 The architecture of the PERSONAGE base generator

One important contribution of this paper is our survey of these studies, and our proposals for generation parameters that can affect these lexical category, word or syntactic structure counts. Another advantage is that prior work on the Big Five model uses validated personality surveys to assess personality traits in humans (e.g. McCrae and Costa 1987; John and Srivastava 1999; Gosling et al. 2003). Rather than inventing our own assessment methods for potentially ill-defined stylistic variations, we use these same surveys to evaluate our computational model of personality generation, and verify that the personality we intend to project is perceived correctly.

Figure 2 specifies PERSONAGE's architecture and gives examples of parameters introduced in each module in order to produce and control linguistic variation. This architecture is based on standard NLG pipeline architecture (Reiter and Dale 2000; Kittredge et al. 1991; Walker and Rambow 2002; Walker et al. 2007); we know of no other work that exploits the modular nature of this architecture to target

personality-based variation. PERSONAGE builds on the SPARKY sentence planner (Stent et al. 2004; Walker et al. 2007), which produces comparisons and recommendations of restaurants in New York City. The inputs are (1) a content plan representing a high-level communicative goal (speech act); (2) a content pool that can be used to achieve that goal, and (3) a set of parameter values for the generation parameters that we define below. PERSONAGE's content pool is based on a database of restaurants in New York City, with associated scalar values representing evaluative ratings for six attributes: *food quality*, *service*, *cuisine*, *location*, *price* and *atmosphere*.[1] In a dialogue system, the content plan is provided by the dialogue manager. Figure 2 also shows how PERSONAGE uses multiple general, domain-independent, online lexical resources, such as WordNet and VerbOcean (Fellbaum 1998; Chklovski and Pantel 2004).

Here we introduce all of the parameters that PERSONAGE controls and explain how the families of parameters are modularized in terms of the standard NLG architecture. In Sect. 3, we will present personality models for each Big Five personality trait; the parameters of these personality models will be organized in terms of the architecture we present here. Section 2.1 discusses the first module shown in Fig. 2. This component is called the *content planner*; it is responsible for selecting and structuring the information to be conveyed. The resulting content plan tree is then processed by the *sentence planner*, which selects syntactic structural templates for expressing individual propositions (Sect. 2.2), and aggregates them to produce the utterance's full syntactic structure (Sect. 2.3). The pragmatic marker insertion component then modifies the syntactic structure locally to produce various pragmatic effects, depending on the markers' insertion constraints (Sect. 2.4). The lexical choice component selects the most appropriate lexeme for each content word, given the lexical selection parameters (Sect. 2.5). Finally, the RealPro realizer (Lavoie and Rambow 1997) converts the final syntactic structure into a string by applying surface grammatical rules, such as morphological inflection and function word insertion.[2] To make PERSONAGE domain-independent, the input parameter values are normalized between 0 and 1 for continuous parameters, and to 0 or 1 for binary parameters, e.g. a VERBOSITY parameter of 1 maximizes the utterance's verbosity given the input, regardless of the actual number of propositions expressed.

2.1 Content planning

The input to the generation process is a *content plan*, a high level structure reflecting the communicative goal of the utterance. The content plan combines together propositions expressing information about individual attributes using *rhetorical relations* from Rhetorical Structure Theory, as in other generators (Mann and Thompson 1988; Scott and de Souza 1990; Marcu 1996; Moore and Paris 1993). Two types of

[1] The attribute values used in the present work are derived from Zagat Survey's ratings.

[2] In a typical dialogue system, the output of the realizer is annotated for prosodic information by the prosody assigner, before being sent to the text-to-speech engine to be converted into an acoustic signal. PERSONAGE does not currently express personality through prosody, although studies of how personality is expressed in speech (Scherer 1979) could be used to develop such parameters for PERSONAGE.

Relations:	JUSTIFY (N:1, S:2); JUSTIFY (N:1, S:3); JUSTIFY (N:1, S:4); JUSTIFY (N:1, S:5); JUSTIFY (N:1, S:6); JUSTIFY (N:1, S:7)
Content:	1. assert(best (*Chanpen Thai*)) 2. assert(is (*Chanpen Thai*, cuisine (*Thai*))) 3. assert(has (*Chanpen Thai*, food-quality (.8))) 4. assert(has (*Chanpen Thai*, atmosphere (.6))) 5. assert(has (*Chanpen Thai*, service (.8))) 6. assert(is (*Chanpen Thai*, price (*24 dollars*))) 7. assert(is (*Chanpen Thai*, location (*Midtown West*)))

Fig. 3 An example content plan for a recommendation. *N* nucleus, *S* satellite

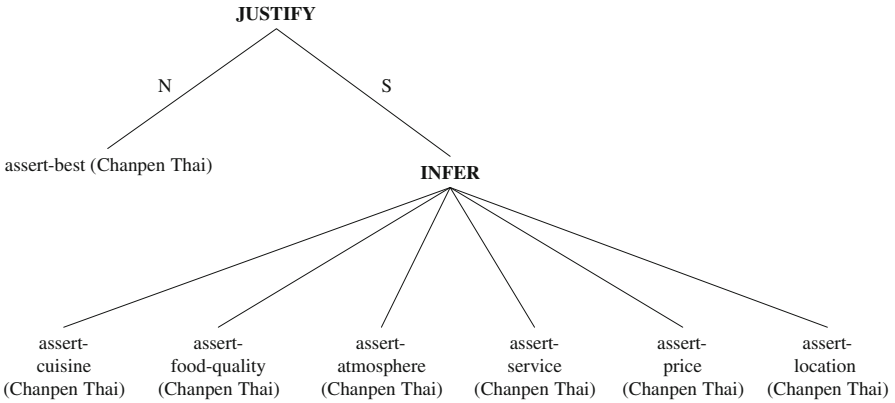


Fig. 4 An example content plan tree for a recommendation for Chanpen Thai, using all the restaurant attributes. *N* nucleus, *S* satellite

communicative goals are supported in PERSONAGE: *recommendation* and *comparison* of restaurants. Figure 3 shows an example content plan for a recommendation.

The content plan is automatically converted into an equivalent tree structure, as illustrated in Fig. 4. This tree structure is referred to as the *content plan tree*. Each recommendation content plan contains a claim (nucleus) about the overall quality of the selected restaurant(s), supported by a set of satellite propositions describing their attributes. The propositions—the leaves in the content plan tree—are assertions labeled *assert-attribute(selection name)* in Fig. 4. Claims can be expressed in different ways, such as RESTAURANT NAME is the best, while the attribute satellites follow the pattern RESTAURANT NAME has MODIFIER ATTRIBUTE NAME, as in *Le Marais has good food*.

Twelve content planning parameters are shown in Table 4 and discussed below. These parameters influence the size of the content plan tree, the content ordering, the rhetorical relations used, and the polarity of the propositions expressed. The correlational studies discussed in Sect. 3 suggests potential relationships between personality traits and a number of generation decisions at the content plan level.

Content size Certain personality types tend to be more verbose, e.g. extraverts are more talkative than introverts (Cope 1969; Furnham 1990; Pennebaker and King 1999; Dewaele and Furnham 1999; Mehl et al. 2006). However because this finding is simply based on word count, it is not clear whether this involves the production of

Table 4 Content planning parameters

Parameters	Description
VERBOSITY	Control the number of propositions in the utterance
RESTATEMENTS	Paraphrase an existing proposition, e.g. *'X has great Y, it has fantastic Z'*
REPETITIONS	Repeat an existing proposition
CONTENT POLARITY	Control the polarity of the propositions expressed, i.e. referring to negative or positive attributes
REPETITION POLARITY	Control the polarity of the restated propositions
CONCESSIONS	Emphasize one attribute over another, e.g. *'even if X has great Z, it has bad Y'*
CONCESSION POLARITY	Determine whether positive or negative attributes are emphasized
POLARIZATION	Control whether the expressed polarity is neutral or extreme
POSITIVE CONTENT FIRST	Determine whether positive propositions are uttered first
REQUEST CONFIRMATION	Begin the utterance with a confirmation of the request, e.g. *'did you say X?'*
INITIAL REJECTION	Begin the utterance with a rejection, e.g. *'I'm not sure'*
COMPETENCE MITIGATION	Express the speaker's negative appraisal of the hearer's request, e.g. *'everybody knows that …'*

more content, or just being redundant and wordy. Thus several parameters relate to the amount and type of content produced.

The VERBOSITY parameter controls the number of propositions selected from the content plan. The parameter value defines the ratio of propositions that are kept in the final content plan tree, while satisfying constraints dependent on the communicative goal: a recommendation must include a claim, and a comparison must include a pair of contrasted propositions. For example, the low extraversion utterance in Table 2 has a low VERBOSITY value, while the high extraversion utterance has high VERBOSITY, and expresses most of the items in the content plan.

The REPETITION parameter adds an exact repetition: the proposition node is duplicated and linked to the original content by a RESTATE rhetorical relation. The continuous parameter value (between 0 and 1) is mapped linearly to the number of repetitions in the content plan tree, i.e. between 0 and a domain-specific maximum (set to 2 in our domain). In Table 12, utterance 6 contains a repetition for the food quality attribute. The RESTATEMENT parameter adds a paraphrase to the content plan, obtained from the generation dictionary (see Sect. 2.2). If no paraphrase is found, one is created automatically by substituting content words with the most frequent WordNet synonym (see Sect. 2.5).

Polarity Polarity parameters bias the type of propositions that are selected to achieve the communicative goal, and control whether positive or negative information is most salient in the utterance. This models findings that some personality types are more positive, e.g. they try to find something positive to say, while other traits tend to engage in more "problem talk" and expressions of dissatisfaction (Thorne 1987; Pennebaker and King 1999; Oberlander and Gill 2006). See Table 4 for definitions of the CONTENT POLARITY, REPETITION POLARITY, CONCESSIONS, CONCESSION POLARITY, and POLARIZATION parameters.

To support the CONTENT POLARITY parameter, propositions are defined as positive or negative. In our domain, propositions expressing attributes that received low ratings from users in Zagat surveys are defined as negative, although there are potentially many ways to assign positive and negative polarities to propositions (Wiebe 1990; Fleischman and Hovy 2002; Higashinaka et al. 2007). The claim in a recommendation is assigned a maximally positive polarity of 1, while the *cuisine* and *location* attributes are set at neutral polarity.[3] Then, the value of the CONTENT POLARITY parameter controls whether the content is mostly negative (e.g. *'Chanpen Thai has mediocre food'*), neutral (e.g. *'Le Marais is a French restaurant'*), or positive (e.g. *'Babbo has fantastic service'*).

From the filtered set of propositions, the POLARIZATION parameter determines whether the final content includes attributes with extreme scalar values (e.g. *'Chanpen Thai has fantastic staff'* versus *'Chanpen Thai has decent staff'*). The REPETITIONS POLARITY parameters controls whether repetitions and paraphrases, if introduced, repeat and emphasize the positive content, or the negative content.

Rhetorical structure also affects the perceived polarity of an utterance, e.g. compare *'even if the food is good, it's expensive'* to *'even if the food is expensive, it's good'*. The CONCESSIONS parameter controls whether two propositions with different polarity are presented objectively, or if one is foregrounded and the other backgrounded. If two opposed propositions are selected for a concession, a CONCEDE relation is inserted between them. The CONCESSION POLARITY parameter controls if the positive content is conceded (*'even if the food is good, it's expensive'*) or the negative content *'even if the food is expensive, it's good'*).

Content ordering Although extraverts use more positive language (Thorne 1987; Pennebaker and King 1999), the position of content affects the persuasiveness of an argument (Carenini and Moore 2000). The POSITIVE CONTENT FIRST parameter controls whether positive propositions—including the claim—appear first or last. The INITIAL REJECTION, REQUEST CONFIRMATION and COMPETENCE MITIGATION parameters are defined in Table 4 are, from a theoretical perspective, content planning parameters. However, since they only affect the beginning of the utterance, we implemented them along with the insertion of pragmatic markers as described in Sect. 2.4.

2.2 Syntactic template selection

Once the content planner has determined *what* will be talked about, the remaining components control *how* the information is to be conveyed. The first phase of sentence planning looks in the generation dictionary for the set of syntactic elementary structures stored for each proposition in the content plan. PERSONAGE manipulates syntactic dependency tree representations inspired by Melčuk's Meaning-Text Theory (1988), and referred to as Deep Syntactic Structures (DSyntS), Fig. 5 shows two DSyntS expressing the recommendation claim. The DSyntS are stored in a handcrafted generation dictionary, currently containing 18 DSyntS: 12 for the recommendation claim and one per attribute. Some attribute DSyntS contain variables that are instantiated

[3] An alternative would be to use individual user models to assign positive and negative polarities to categorical attributes as well (Walker et al. 2004; Carenini and Moore 2006; Ardissono et al. 2003).

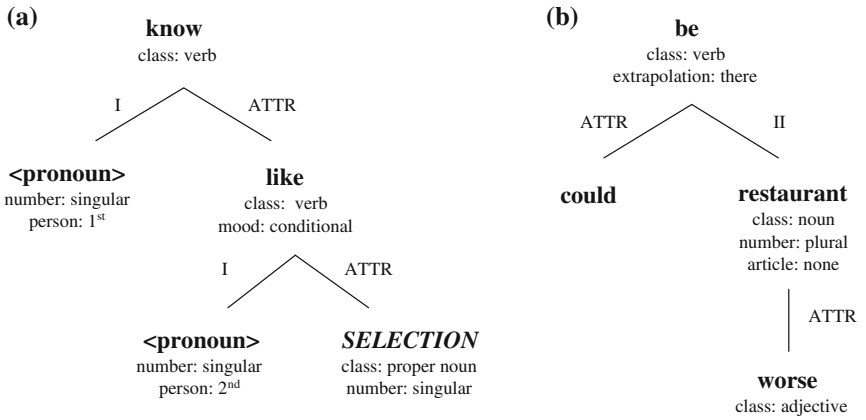


Fig. 5 Two example DSyntS for a recommendation claim. The lexemes are in bold, and the attributes below indicate non-default values in the RealPro realizer. Branch labels indicate dependency relations, i.e. I = subject, II = object and ATTR = modifier. Lexemes in italic are variables that are instantiated at generation time. Realizations: (**a**) *I know you would like SELECTION*; (**b**) *There could be worse restaurants*

based on the input restaurant (e.g. polarity adjectives in Sect. 2.5, see adjective *good* in Fig. 6a). These DSynts representations can be combined using domain-independent general-purpose linguistic operations to make more complex DSyntS (complex utterance structures) in order to produce a wide range of variation. The DSyntS can be converted to an output string using the RealPro surface realizer, which is also based on Melcuk's theory (Lavoie and Rambow 1997). The DSyntS contain variables that are filled at generation time, such as the restaurant's name or cuisine. See Fig. 5.

Table 5 shows the PERSONAGE parameters that control the selection of DSyntS from the generation dictionary. The DSyntS selection process first assigns each candidate DSyntS to a point in a three-dimensional space, characterizing the DSyntS' syntactic complexity, number of self-references and polarity. Parameter values are normalized over all candidate DSyntS, and the DSyntS with the shortest Euclidean distance to the target value vector is selected. The following three dimensions (i.e., parameters) affect personality perceptions.

Syntactic complexity Furnham (1990) suggests that introverts produce more complex constructions: the SYNTACTIC COMPLEXITY parameter controls the number of subordinate clauses of the DSyntS chosen to represent the claim, based on Beaman's definition of syntactic complexity (1984).[4] For example, the claim in Fig. 5a is rated as more complex than the one in Fig. 5b, because the latter has no subordinate clause.

Self-references Extraverts and neurotics make more self-references (Pennebaker and King 1999). The SELF- REFERENCES parameter controls whether the claim is made in the first person (based on the speaker's own experience), or whether the claim is reported as objective or information obtained elsewhere. The SELF- REFERENCES value is computed from the DSyntS by counting the number of first person pronouns. For example, the DSyntS in Fig. 5a contains one self-reference, while that in Fig. 5b does not.

[4] The syntactic complexity is computed as the number of verb nodes in the DSyntS, which is equivalent to the number of subordinate clauses in the final utterance.

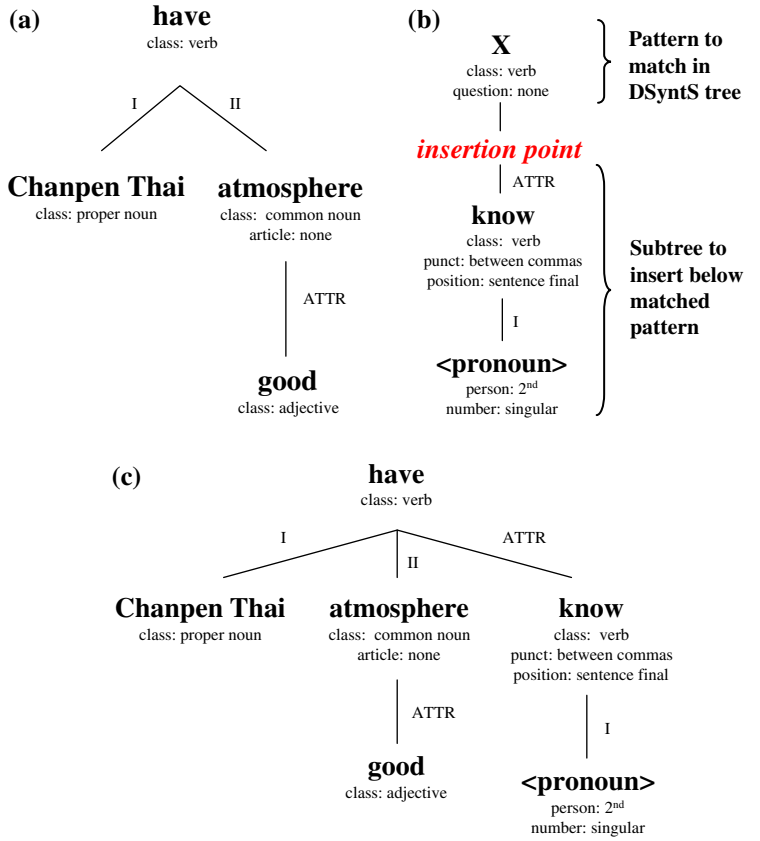


Fig. 6 Illustration of the pragmatic marker insertion process for the hedge *you know* in the DSyntS *'Chanpen Thai has good atmosphere'*. (**a**) Example input DSyntS realized as *'Chanpen Thai has good atmosphere'*; (**b**) Syntactic representation of the insertion constraints for the pragmatic marker *you know*; (**c**) Modified DSyntS after the insertion of the pragmatic marker below the main verb matching the pattern defined in (**b**)'s root node

Table 5 Syntactic template selection parameters

Parameters	Description
SYNTACTIC COMPLEXITY	Control the syntactic complexity (e.g. syntactic embedding)
SELF- REFERENCES	Control the number of first person pronouns
TEMPLATE POLARITY	Control the syntactic structure's connotation (positive or negative)

Polarity While polarity can be expressed by content selection and structure, it can also be directly associated with the DSyntS. The TEMPLATE POLARITY parameter determines whether the claim has a positive or negative connotation (Wiebe 1990; Hovy 1988; Fleischman and Hovy 2002). While automated methods for opinion extraction could be used in the future to annotate the generation dictionary (Pang et al. 2002; Hu and Liu 2004; Higashinaka et al. 2007; Wiebe 1990; Wilson et al. 2004; Riloff et al. 2005), at present DSyntS are manually annotated for polarity. An example claim

with low polarity can be found in Fig. 5b, i.e. *'There could be worse restaurants'*, while the claim in Fig. 5a is rated more positively.

2.3 Aggregation

Previous work suggests that personality affects the aggregation process, e.g. introverts prefer complex syntactic constructions, long pauses and rich vocabulary (Furnham 1990; Siegman and Pope 1965; Scherer 1979). The aggregation component combines elementary DSyntS into larger syntactic structures in order to produce a large variety of sentences from a small number of elementary DSyntS. As in Rambow et al. (2001), the aggregation process randomly selects a *clause-combining operation*, for each rhetorical relation in the content plan tree. Table 6 shows the PERSONAGE parameters that control the selection of clause-combining operators in the sentence planner, and summarizes the effect of each operation on the final utterance. For example, poor food quality can be contrasted with good atmosphere using cue words such as *however*,

Table 6 Aggregation parameters

Parameters	Description
PERIOD	Leave two propositions in their own sentences, e.g. *'X has great Y. It has nice Z.'*
RELATIVE CLAUSE	Join propositions with a relative clause, e.g. *'X, which has great Y, has nice Z'*
WITH CUE WORD	Aggregate propositions using *with*, e.g. *'X has great Y, with nice Z'*
CONJUNCTION	Join propositions using a conjunction, or a comma if more than two propositions
MERGE	Merge the subject and verb of two propositions, e.g. *'X has great Y and nice Z'*
ALSO CUE WORD	Join two propositions using *also*, e.g. *'X has great Y, also it has nice Z'*
CONTRAST- CUE WORD	Contrast two propositions using *while, but, however, on the other hand*, e.g. *'While X has great Y, it has bad Z'*, *'X has great Y, but it has bad Z'*
WHILE CUE WORD	Contrast two propositions using *while*, e.g. *'While X has great Y, it has bad Z'*
HOWEVER CUE WORD	Contrast two propositions using *however*, e.g. *'X has great Y. However, it has bad Z'*
ON THE OTHER HAND CUE WORD	Contrast two propositions using *on the other hand*, e.g. *'X has great Y. On the other hand, it has bad Z'*
JUSTIFY- CUE WORD	Justify a proposition using *because, since, so*, e.g. *'X is the best, since it has great Y'*
BECAUSE CUE WORD	Justify a proposition using *because*, e.g. *'X is the best, because it has great Y'*
SINCE CUE WORD	Justify a proposition using *since*, e.g. *'X is the best, since it has great Y'*
SO CUE WORD	Justify a proposition using *so*, e.g. *'X has great Y, so it's the best'*
CONCEDE- CUE WORD	Concede a proposition using *although, even if, but/though*, e.g. *'Although X has great Y, it has bad Z'*, *'X has great Y, but it has bad Z though'*
ALTHOUGH CUE WORD	Concede a proposition using *although*, e.g. *'Although X has great Y, it has bad Z'*
EVEN IF CUE WORD	Concede a proposition using *even if*, e.g. *'Even if X has great Y, it has bad Z'*
BUT/THOUGH CUE WORD	Concede a proposition using *but/though*, e.g. *'X has great Y, but it has bad Z though'*
MERGE WITH COMMA	Restate a proposition by repeating only the object, e.g. *'X has great Y, nice Z'*
OBJECT ELLIPSIS	Replace part of a repeated proposition by an ellipsis, e.g. *'X has …it has great Y'*

or *but*. PERSONAGE augments the SPARKY clause-combining operations (Stent et al. 2004; Walker et al. 2007), with additional operations for the RESTATE and CONCEDE rhetorical relations. For more detail see (Mairesse 2008).

2.4 Pragmatic marker insertion

Psychological studies identify many pragmatic markers of personality that affect the utterance locally, and can be implemented as context-independent syntactic transformations. Table 7 describes all of the pragmatic marker insertion parameters and provides examples. For example, parameters in Table 7 include negations, tentative/softening hedges (e.g. *maybe*, *kind of*) and filled pauses (Pennebaker and King 1999; Siegman and Pope 1965; Scherer 1979), expletives, emphasizer hedges (e.g. *really*) and exclamation marks (Oberlander and Gill 2004b; Mehl et al. 2006). There are FILLED PAUSES and STUTTERING markers because introverts and neurotics produce more filled pauses and disfluencies (Scherer 1979; Weaver 1998; Scherer 1981). Neuroticism is associated with frustration and acquiescence, which are modeled with

Table 7 Pragmatic marker insertion parameters

Parameters	Description
SUBJECT IMPLICITNESS	Make the presented object implicit by moving its attribute to the subject, e.g. *'the Y is great'*
NEGATION	Negate a verb by replacing its modifier by its antonym, e.g. *'X doesn't have bad Y'*
SOFTENER HEDGES	Insert syntactic elements (*sort of*, *kind of*, *somewhat*, *quite*, *around*, *rather*, *I think that*, *it seems that*, *it seems to me that*) to mitigate the strength of a proposition, e.g. *'X has kind of great Y'* or *'It seems to me that X has rather great Y'*
EMPHASIZER HEDGES	Insert syntactic elements (*really*, *basically*, *actually*, *just*) to strengthen a proposition, e.g. *'X has really great Y'* or *'Basically, X just has great Y'*
ACKNOWLEDGMENTS	Insert an initial back-channel (*yeah*, *right*, *ok*, *I see*, *oh*, *well*), e.g. *'Ok, X has great Y'*
FILLED PAUSES	Insert syntactic elements expressing hesitancy (*I mean*, *err*, *mmhm*, *like*, *you know*), e.g. *'Err…X has, like, great Y'*
EXCLAMATION	Insert an exclamation mark, e.g. *'X has great Y!'*
EXPLETIVES	Insert a swear word, e.g. *'the Y is damn great'*
NEAR EXPLETIVES	Insert a near-swear word, e.g. *'the Y is darn great'*
TAG QUESTION	Insert a tag question, e.g. *'the Y is great, isn't it?'*
STUTTERING	Duplicate parts of a content word, e.g. *'X has gr-gr-great Y'*
IN- GROUP MARKER	Refer to the hearer as a member of the same social group, e.g. *pal*, *mate* and *buddy*
PRONOMINALIZATION	Replace references to the object by pronouns, as opposed to proper names or the reference *this restaurant*
REQUEST CONFIRMATION	Begin the utterance with a confirmation of the request, e.g. *'did you say X?'*
INITIAL REJECTION	Begin the utterance with a rejection, e.g. *'I'm not sure'*
COMPETENCE MITIGATION	Express the speaker's negative appraisal of the hearer's request, e.g. *'everybody knows that …'*

the EXPLETIVES and ACKNOWLEDGMENT parameters (Weaver 1998; Oberlander and Gill 2004b). Syntactic pattern matching controls the insertion of context-independent markers, while some parameters require more complex processing. Weaver (1998) shows that extraverts are more sympathetic to other people—i.e. they show more concern—although this sympathy is not related to empathy, as they are not more inclined to feel other people's feelings. Concern for the user can be expressed in the information presentation domain by emphasizing the user's request using the REQUEST CONFIRMATION parameter.

Syntactically embedded markers Some pragmatic markers are inserted in the syntactic structure of an utterance (*like, you know, sort of*), and their insertion must respect particular syntactic constraints. Our approach is to add to the generation dictionary the syntactic representations that characterize each pragmatic marker. For each marker, the insertion process involves traversing the aggregated DSyntS to identify *insertion points* satisfying the syntactic constraints specified in the database.

Figure 6 illustrates the matching and insertion process for the hedge *you know*. Each pragmatic marker dictionary entry consists of a syntactic pattern to be matched in the DSyntS, such as the root node in Fig. 6b, and an insertion point element corresponding to the location in the DSyntS where the insertion should be made. Given the input DSyntS in Fig. 6a *'Chanpen Thai has good atmosphere'*, the verb *to have* is matched with the root node of the structure in Fig. 6b, and thus the subtree below the insertion point is inserted under Fig. 6a's root node. The resulting DSyntS is in Fig. 6c. This DSyntS is realized as *'Chanpen Thai has good atmosphere, you know'*.

This approach supports modifying utterances at the syntactic level rather than at the surface level, which allows the RealPro surface realizer to do what it was designed for, namely to enforce grammaticality. For example, pragmatic markers are added without controlling the final word order, while positional constraints can be enforced when required, e.g. the *position* attribute in Fig. 6b specifies that *you know* should be in sentence final position. Similarly, while the *punct* attribute specifies that the marker must appear between commas—irrespectively of its position in the utterance, the realizer ensures that the sentence is punctuated correctly by removing commas preceding the final period. We believe that the DSyntS representation is also what enables us to insert many different pragmatic markers into the same utterance while carrying out syntactic transformations such as negation insertion, while still ensuring that we produce grammatical outputs.[5]

PERSONAGE implements a binary generation parameter for the pragmatic markers in Table 8 using the insertion mechanism described above. At generation time, syntactic patterns are randomly chosen among markers with parameter values set to 1, and matched against the aggregated DSyntS. The insertion process ends when there are no markers left in the database, or when the number of successful insertions is above a threshold.

Other markers The remaining pragmatic markers (marked with an asterisk in Table 8) require more complex syntactic processing and are implemented independently.

[5] The EVEN IF CUE WORD, MERGE WITH COMMA, and OBJECT ELLIPSIS operations were added to increase the range of pragmatic effects

Table 8 Pragmatic markers implemented in PERSONAGE, with insertion constraints and example realizations

Marker	Constraints	Example
General		
NEGATION*	Adjective modifier + antonym	Chanpen Thai doesn't have bad atmosphere
EXCLAMATION	Sentence-final punctuation	Chanpen Thai has good atmosphere!
IN- GROUP MARKER	Clause-final adjunct, e.g. *pal*, *mate* and *buddy*	Chanpen Thai has good atmosphere pal
SUBJECT IMPLICITNESS*	Requires a DSyntS of the form *NOUN has ADJ NOUN*	The atmosphere is good
TAG QUESTION*	None	Chanpen Thai has good atmosphere, doesn't it?
STUTTERING*	Selection name	Ch-Chanpen Thai has good atmosphere
EXPLETIVES	Adjective modifier (*damn*, *bloody*)	Chanpen Thai has damn good atmosphere
	Clause-initial adjunct (*oh god*)	Oh god Chanpen Thai has good atmosphere
NEAR EXPLETIVES	Adjective modifier (*darn*)	Chanpen Thai has darn good atmosphere
	Clause-initial adjunct (*oh gosh*)	Oh gosh Chanpen Thai has good atmosphere
REQUEST	None	You want to know more about Chanpen Thai?
CONFIRMATION*		Let's see… Chanpen Thai
		Let's see what we can find on Chanpen Thai
		Did you say Chanpen Thai?
INITIAL REJECTION*	None	I don't know
		I'm not sure
		I might be wrong
COMPETENCE MITIGATION	Main verb is subordinated to new clause (*everybody knows that* and *I thought everybody knew that*)	Everybody knows that Chanpen Thai has good atmosphere
	Clause-initial adjunct (*come on*)	Come on, Chanpen Thai has good atmosphere
Softeners		
KIND OF	Adjective modifier	Chanpen Thai has kind of good atmosphere
SORT OF	Adjective modifier	Chanpen Thai has sort of good atmosphere
SOMEWHAT	Adjective modifier with verb *to be*	The atmosphere is somewhat good
QUITE	Adjective modifier	Chanpen Thai has quite good atmosphere
RATHER	Adjective modifier	Chanpen Thai has rather good atmosphere
AROUND	Numeral modifier	Chanpen Thai's price is around $44
SUBORDINATE	Main verb is subordinated to hedge clause, e.g. *I think that* and *it seems (to me) that*	It seems to me that Chanpen Thai has good atmosphere

Table 8 continued

Marker	Constraints	Example
Filled pauses		
LIKE	Verb modifier	Chanpen Thai has, like, good atmosphere
ERR	Clause-initial adjunct	Err… Chanpen Thai has good atmosphere
MMHM	Clause-initial adjunct	Mmhm… Chanpen Thai has good atmosphere
I MEAN	Clause-initial adjunct	I mean, Chanpen Thai has good atmosphere
YOU KNOW	Clause-final adjunct	Chanpen Thai has good atmosphere, you know
Emphasizers		
REALLY	Adjective modifier	Chanpen Thai has really good atmosphere
BASICALLY	Clause-initial adjunct	Basically, Chanpen Thai has good atmosphere
ACTUALLY	Clause-initial adjunct	Actually, Chanpen Thai has good atmosphere
JUST	Pre-verbal modifier of *to have*	Chanpen Thai just has good atmosphere
	Post-verbal modifier of *to be*	The atmosphere is just good
Acknowledgments		
YEAH	Clause-initial adjunct	Yeah, Chanpen Thai has good atmosphere
WELL	Clause-initial adjunct	Well, Chanpen Thai has good atmosphere
OH	Clause-initial adjunct	Oh, Chanpen Thai has good atmosphere
RIGHT	Clause-initial adjunct	Right, Chanpen Thai has good atmosphere
OK	Clause-initial adjunct	Ok, Chanpen Thai has good atmosphere
I SEE	Clause-initial adjunct	I see, Chanpen Thai has good atmosphere

An asterisk indicates that the pragmatic marker requires specific processing and was not implemented through pattern matching and insertion

proximal deictic expressions are a way to express involvement and empathy (Brown and Levinson 1987), e.g. *'this restaurant has great service'*. Referring expression generation in PERSONAGE is based on a simple algorithm which identifies as potential anaphoric expressions any restaurant name that follows a previous reference to it, e.g. *'Chanpen Thai is the best, it has great service'*. Then, a PRONOMINALIZATION parameter controls whether referring expressions are realized as personal pronouns or proximal demonstrative phrases. by specifying the ratio of pronouns to other types of referring expressions. The RealPro surface realizer automatically selects the personal pronoun based on the selection's DSyntS node; inserting a demonstrative phrase requires replacing the selection's lexeme with a generic noun (e.g. *restaurant*) and setting the determiner to a demonstrative.

Negations indicate both introversion and a lack of conscientiousness (Pennebaker and King 1999; Mehl et al. 2006), a NEGATION parameter allows PERSONAGE to insert a negation while preserving the initial communicative goal. An adjective modifying a verb or its object is randomly selected from the DSyntS, and its antonym is retrieved from WordNet (Fellbaum 1998). If the query is successful, the adjective's lexeme is

replaced by the antonym and the governing verb is negated,[6] e.g. *'Chanpen Thai has good atmosphere'* becomes *'Chanpen Thai doesn't have bad atmosphere'*. Adjectives in the domain are manually sense-tagged to ensure that they can be substituted by their antonym. Also, a maximum of one negation can be inserted to prevent the utterance from sounding unnatural.

Heylighen and Dewaele (2002) found that extraverts use more implicit language than introverts. A SUBJECT IMPLICITNESS parameter thus determines whether predicates describing restaurant attributes are expressed with the restaurant's name in the subject, or with the attribute itself by making the reference to the restaurant implicit (e.g. *'Chanpen Thai has good atmosphere'* versus *'the atmosphere is good'*). The syntactic transformation involves shifting the object attribute to the subject, while promoting the adjective below the main verb, and changing the main verb's lexeme to *to be*. Hence, the transformation requires an input DSyntS matching the template *NOUN has ADJECTIVE NOUN*.

Speech disfluencies are associated with anxiety and neuroticism (Scherer 1981), so we introduce a STUTTERING parameter that modifies the lexeme of a randomly selected proper noun by repeating the first two letters two or three times, e.g. *'Ch-Ch-Chanpen Thai'*. Only selection names are repeated as they are likely to be new to the speaker, the stuttering can therefore be interpreted as non-pathological. Allowing disfluencies to affect any word requires determining what words can be altered, which involves deep psycholinguistic modeling that is beyond the scope of this work.

PERSONAGE also implements politeness markers such as rhetorical questions. The TAG QUESTION parameter processes the DSyntS by (1) duplicating a randomly selected verb and its subject; (2) negating the verb; (3) pronominalizing the subject; (4) setting the verb to the interrogative form and (5) appending the duplicated subtree as a sentence-final adjunct, e.g. *'Chanpen Thai has great food'* results in the insertion of *'doesn't it?'*. The duplicated verb is generally not realized,[7] i.e. only the negated auxiliary appears in the tag question. Additionally, whenever the subject is a first person pronoun, the verb is set to the conditional form and a second person pronoun is inserted, producing *'I would recommend Chanpen Thai, wouldn't you?'*. If the tag question insertion is unsuccessful, e.g. due to an extrapolated subject *'there is'*, a default tag question is appended, producing either *'you see?'*, *'alright?'* or *'okay?'*.

As mentioned above, the REQUEST CONFIRMATION, INITIAL REJECTION and COMPETENCE MITIGATION parameters are content level parameters that we implement as pragmatic markers, by inserting a full DSyntS at the beginning of the utterance, randomly chosen from the dictionary of such markers. The INITIAL REJECTION parameter reduces the level of confidence of the speaker over the utterance's informational content, by beginning the utterance with either *'I don't know'*, *'I'm not sure'* or *'I might be wrong'*. The REQUEST CONFIRMATION parameter produces an implicit confirmation, which both redresses the hearer's positive face through grounding and emphasizes the system's uncertainty about the user's request, e.g. *'you want to know more about*

[6] At the DSyntS level the negation is represented as an attribute of the verb element, the actual inflection is done by RealPro in the realization phase.

[7] The verb *to be* is an exception.

Chanpen Thai?'. In order to convey disagreeableness, a COMPETENCE MITIGATION parameter also presents the user's request as trivial by embedding it as a subordinate clause, e.g. *'everybody knows that Chanpen Thai has good service'*. See Table 8 for additional examples of confirmation and competence mitigation DSyntS.

2.5 Lexical choice

Lexical features related to personality include word length, word frequency and verb strength. In addition, lexical choice is crucial to successful individual adaptation in dialogue systems (Brennan 1996; Lin 2006). Thus, PERSONAGE allows many different lexemes to be expressed for each content word, depending on input parameter values. See Table 9.

The lexical selection component processes the DSyntS by sequentially modifying each content word. For each lexeme in the DSyntS, the corresponding WordNet synonyms are mapped to a multi-dimensional space defined by the lexeme's length, frequency of use and strength, using machine-readable dictionaries. The values along each dimension are normalized over the set of synonyms, and the synonym that is the closest to the target parameter values (in terms of Euclidean distance) is selected. Although word-sense disambiguation techniques could be used in the future, content words are manually sense-tagged to ensure that the synonyms are interchangeable in the dialogue domain. Figure 7 illustrates the lexical choice process using the word length and word frequency dimensions, resulting in the selection of *cheap* over *inexpensive* because its length (5 letters) and its normalized frequency (1.0) are closer to the desired target values, i.e. a 6 letter word (normalized length of $\frac{6-5}{11-5} = .17$) with a normalized frequency of .7.

To enrich the pool of synonyms from Wordnet, adjectives extracted by Higashinaka et al. (2007) from a corpus of restaurant reviews and their synonyms are added to the synonym set of each attribute modifier. Table 10 lists the extracted adjective sets for the food quality attribute, ordered by polarity. Synonym selection jointly controls the average normalized frequency of use, word length and verb strength in each DSyntS.

Frequency of use Introvert and emotionally stable speakers use a richer vocabulary (Dewaele and Furnham 1999; Gill and Oberlander 2003). We model this with a LEXICON FREQUENCY parameter that selects lexical items associated with a particular part of speech using the frequency count in the British National Corpus, in order to support the selection of unusual low-frequency words.

Table 9 Lexical choice parameters

Parameters	Description
LEXICON FREQUENCY	Control the average frequency of use of each content word (e.g. according to frequency counts from a corpus)
LEXICON WORD LENGTH	Control the average number of letters of each content word
VERB STRENGTH	Control the strength of the verbs, e.g. *'I would suggest'* versus *'I would recommend'*

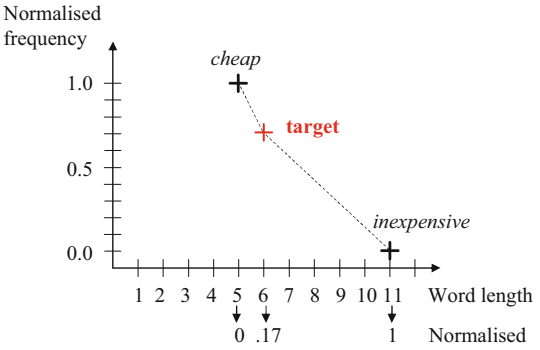


Fig. 7 Illustration of the lexical selection process between the synonyms *cheap* and *inexpensive* with two input dimensions

Table 10 Adjectives and polarity ratings (5 = very positive) for the food quality attribute, extracted from a corpus of restaurant reviews by Higashinaka et al. (2007)

Polarity	Adjectives
1	Awful, bad, terrible, horrible, horrendous
2	Bland, mediocre, bad
3	Decent, acceptable, adequate, satisfying
4	Good, flavourful, tasty, nice
5	Excellent, delicious, great, exquisite, wonderful, legendary, superb, terrific, fantastic, outstanding, incredible, delectable, fabulous, tremendous, awesome, delightful, marvellous

Word length Mehl et al. (2006) show that observers associate long words with agreeableness, conscientiousness and openness to experience. Thus we introduce a LEXICON WORD LENGTH parameter to control the number of letters of the selected synonym.

Verb strength Verb synonyms, such as *appreciate*, *like* and *love*, differ in terms of their connotative strength (Wilson et al. 2004; Inkpen and Hirst 2004). This variation is controlled in PERSONAGE through the VERB STRENGTH parameter, which orders each verb's synonym set according to the *stronger-than* semantic relation in the VERBOCEAN database (Chklovski and Pantel 2004). The process is illustrated in Fig. 8 for synonyms of the verb *to know*. The ordered synonyms are mapped to equidistant points in the [0, 1] interval to produce the final parameter value, i.e. the weakest verb is associated with 0.0 and the strongest with 1.0. This mapping is based on the assumption that the magnitude of the *stronger-than* relation is constant between contiguous synonyms, i.e. the verb strength is uniformly distributed over the synonym set.

The lexical choice parameters described above associate each candidate synonym with three values, and the one with the closest values to the target is selected. Since values are normalized over the members of the synonym set, all dimensions have the same weight in the selection process.[8] For example, consider the input DSyntS expressing *'I know you would like Chanpen Thai'*; a low VERB STRENGTH parameter

[8] An exception is that verb selection is only affected by the VERB STRENGTH parameter, to ensure that its effect is perceptible in the output utterance.

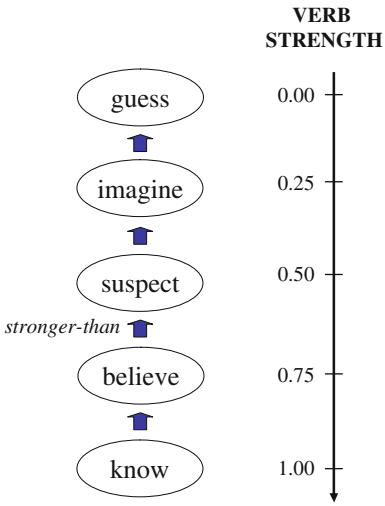


Fig. 8 Determination of the VERB STRENGTH parameter values for synonyms of the verb *to know*, based on the *stronger-than* semantic relation in VERBOCEAN

value produces *'I guess you would like Chanpen Thai'*, while a high value yields *'I know you would love Chanpen Thai'*. Similarly, a proposition realized as *'this place has great ambiance'* is converted into *'this restaurant features fantastic atmosphere'* given high LEXICON WORD LENGTH and VERB STRENGTH parameter values, together with a low LEXICON FREQUENCY value.

3 Personality models

The PERSONAGE base generator described above can produce *thousands* of utterances for any input content plan; this variation needs to be *controlled* to achieve particular communicative goals. Table 11 provides examples of utterances generated when the PERSONAGE parameters are randomly varied. Random variation can result in inconsistent personality cues that may sometimes, but not reliably, be recognized as manifesting a particular personality. See the personality ratings in Table 11. Table 23, discussed below, clearly illustrates the differences in the reliability of human perceptions for randomly generated utterances versus those generated using personality models.

In the appendix, Table 27 summarizes the studies that motivate our parameters. An important thing to realize about these studies is that they are typically based on lexical or syntactic tokens or types that can be counted, and the results consist of correlations between these counts and personality traits. In order to use such results, we explore various parameters that could have caused an increase or decrease in a particular count, e.g. a high word count can be associated with the expression of more content and/or more repetitions and redundancies. Thus, one of the important contributions of this work is that we test whether and when findings from other language genres (see the *Language source* column in Table 27) generalize to the production of a single utterance presenting information to the user in a controlled discourse situation, such as a recommender dialogue system.

Table 11 Example outputs of PERSONAGE with random parameter settings (random utterances), together with average user ratings from the evaluation experiment in Sect. 4

#	Content plan	PERSONAGE's output	Ratings
1.a	Compare(Kin Khao, Tossed)	Kin Khao and Tossed are bloody outstanding. Kin Khao just has rude staff. Tossed features sort of unmannered waiters, even if the food is somewhat quite adequate	Extraversion = 4.75 Emotional stability = 6.00 Agreeableness = 3.75 Conscientiousness = 4.00 Openness to experience = 3.25
1.b	Compare(Kin Khao, Tossed)	Err… I am not really sure. Tossed offers kind of decent food. Mmhm… however, Kin Khao, which has quite ad-ad-adequate food, is a thai place. You would probably enjoy these restaurants	Extraversion = 4.00 Emotional stability = 4.00 Agreeableness = 5.50 Conscientiousness = 4.75 Openness to experience = 5.00
3.a	Recommend (Chimichurri Grill)	Obviously, oh God Chimichurri Grill has like, quite adequate food, so it's the only re-restaurant I would consider, and it's a latin american place and located in Midtown West with ra-ra-rather friendly staff	Extraversion = 4.50 Emotional stability = 3.75 Agreeableness = 5.25 Conscientiousness = 4.00, Openness to experience = 4.50
3.b	Recommend (Chimichurri Grill)	I see, oh Chimichurri Grill is a latin american place with sort of poor atmosphere. Although it doesn't have rather nasty food, its price is 41 dollars. I suspect it's kind of alright	Extraversion = 2.50 Emotional stability = 4.50 Agreeableness = 3.50 Conscientiousness = 4.75 Openness to experience = 4.25

Below we present tables for each trait that summarize the findings from these studies and show how we develop personality models from them. To do so, each finding (correlation) is first mapped to one or more parameters (generation decisions) of the PERSONAGE generator described in Sect. 2. Second, the personality model for each trait is expressed via parameter trends specified in terms of `high` or `low` settings for particular parameters relevant to that trait, based on the direction and the magnitude of the correlations, e.g. see Table 13 for extraversion. Third, at generation time, these parameter trends are mapped to extreme parameter values to maximize their impact on the utterance, with $low = 0.0$ and $high = 1.0$ for most continuous and binary parameters However, if PERSONAGE always expressed a trait using identical parameter settings (e.g. neurotic), then identical generation decisions (e.g. hesitancy markers) could lead to excessive repetitions of particular linguistic forms. Therefore, parameter values are randomized before generation, according to a normal distribution, with a 15% standard deviation around their predefined value.

3.1 Extraversion

Extraverts tend to engage in social interaction, they are enthusiastic, risk-taking, talkative and assertive, while introverts are more reserved and solitary. Eysenck et al. (1985) suggest that this trait is associated with a lack of internal arousal: extraverts are thus seeking additional external stimulation, while introverts avoid it. The extraversion trait has received the most attention in linguistic studies. There are three reasons for this: (1) the extraversion dimension is often seen as the most important, since it

explains the most variance among the adjective descriptors from which the Big Five factors are derived (Goldberg 1990), (2) extraversion is present in most other personality frameworks—e.g. Eysenck et al.'s PEN model (Psychoticism, Extraversion and Neuroticism; 1985); and (3) extraversion may have the most influence on language, because it is strongly associated with talkativeness and enthusiasm (Furnham 1990).

The findings about linguistic markers of extraversion are summarized in Table 13, together with the generation parameters that represent our hypotheses about how each finding can be mapped into the PERSONAGE framework. Most generation parameters are based on study results, however some are derived from hypotheses about how a specific trait affects language (indicated by a single asterisk). The right-most columns (e.g. *Intro* and *Extra*) contain the parameter values for expressing each end of the personality dimension, i.e. either introversion or extraversion. As mentioned above, parameter values are specified in terms of *low* and *high* settings, and then mapped to normalized scalar values between 0 and 1.

Table 12 shows examples generated by PERSONAGE using the extraversion personality model specified by the parameter settings in Table 13. Mairesse (2008) includes many such examples with more detailed discussion. Utterance 9 in Table 12 conveys introversion through a low VERBOSITY parameter value resulting in 2 propositions, whereas the 6 propositions in utterance 16 are perceived as more extravert. Verbosity can also be expressed by repeating the same content, as in utterance 13 with a restatement of the price information (*its price is 12 dollars, it's cheap*). Utterance 9 in Table 12 provides an example negative claim, while utterance 13 contains a positive claim. Long unfilled pauses can be generated by using the PERIOD aggregation operation, as in utterance 9. Introvert parameters favor operations that we hypothesize to result in more formal language, as in *on the other hand* in utterance 1. Extravert aggregation produces longer sentences with simpler constructions and informal cue words, as in utterance 13. The NEGATION parameter settings and an example of a subordinate hedge are illustrated in utterance 1.

Table 12 Example outputs of PERSONAGE for the extraversion personality model

#	Content plan	End	PERSONAGE's output	Score
1	Compare (Lemongrass Grill, Monsoon)	Low	I think that Lemongrass Grill features mediocre ambience. Monsoon doesn't, on the other hand, have nasty ambiance	3.50
5	Compare (Lemongrass Grill, Monsoon)	High	Yeah, Lemongrass Grill's price is 22 dollars, even if the ambience is just really poor. Monsoon is low-cost and the atmosphere is nice	5.67
9	Recommend (Amy's Bread)	Low	Amy's Bread's price is 12 dollars. It isn't as bad as the others	2.50
13	Recommend (Amy's Bread)	High	I am sure you would like Amy's Bread. Basically, its price is 12 dollars, it's cheap, you know, the food is good and the servers are friendly	6.50
16	Recommend (Bond Street)	High	Yeah, Bond Street is the best place. The atmosphere is good, it has nice service and it's a japanese and sushi place. Even if it's expensive, you know, the food is great	6.67

Personality ratings are on a scale from 1 to 7, with 1 = very low and 7 = very high

Table 13 Summary of language cues for extraversion, with corresponding generation parameters

Introvert findings	Extravert findings	Studies	Parameters	Intro	Extra
Content planning:					
Single topic	Many topics, higher verbal output	FU90, PK99, DF99, ME06, CO69	VERBOSITY	low	high
Strict selection	Think out loud	FU90*	RESTATEMENTS	low	high
			REPETITIONS	low	high
Problem talk, dissatisfaction, negative emotion words	Pleasure talk, agreement, compliment, positive emotion words	PK99, TH87	CONTENT POLARITY	low	high
			REPETITION POLARITY	low	high
			CONCESSION POLARITY	low	high
			POSITIVE CONTENT FIRST	low	high
Not sympathetic	Sympathetic, concerned about hearer (but not empathetic)	WE98	REQUEST CONFIRMATION	low	high
Syntactic template selection:					
Elaborated constructions	Simple constructions	FU90*	SYNTACTIC COMPLEXITY	high	low
Problem talk	Pleasure talk	PK99	TEMPLATE POLARITY	low	high
Aggregation:					
Few conjunctions	Many conjunctions	OG04a	CONJUNCTION, BUT, ALSO CUE WORD	low	high
Many unfilled pauses	Few unfilled pauses	SC79, SP65	PERIOD	high	low
Many uses of *although*	Few uses of *although*	OG04b	ALTHOUGH CUE WORD	high	low
Formal language	Informal language	FU90*, HD02	RELATIVE CLAUSE	high	low
Pragmatic marker insertion:					
Many nouns, adjectives, prepositions (explicit)	Many verbs, adverbs, pronouns (implicit)	HD02	SUBJECT IMPLICITNESS	low	high
Many negations	Few negations	PK99	NEGATION	high	low
Many tentative words (e.g. *maybe*, *guess*)	Few tentative words	PK99	SOFTENER HEDGES: ·SORT OF, SOMEWHAT, QUITE, RATHER, I THINK THAT, IT SEEMS THAT, IT SEEMS TO ME THAT	high	low
Formal language	Informal language	FU90*, HD02	·KIND OF, LIKE	low	high
			ACKNOWLEDGMENTS:		
			·YEAH	low	high
			·RIGHT, OK, I SEE, WELL	high	low
Few swear words	Many swear words	ME06	NEAR EXPLETIVES	low	high
Many unfilled pauses	Few unfilled pauses	SC79, SP65	FILLED PAUSES: · ERR, I MEAN, MMHM, YOU KNOW	high	low
Realism	Exaggeration (e.g. *really*)	OG04b*	EMPHASIZER HEDGES: ·REALLY, BASICALLY, ACTUALLY, JUST	low	high
			EXCLAMATION	low	high
Not sympathetic	Sympathetic, concerned about hearer, minimize positive face threat	WE98	TAG QUESTION	low	high
Few words related to humans	Many words related to humans (e.g. *man*, *pal*)	NO06	IN-GROUP MARKER	low	high
Lexical choice:					
Rich vocabulary	Poor vocabulary	FU90*, DF99	LEXICON FREQUENCY	low	high
Longer words	Shorter words	ME06	LEXICON WORD LENGTH	high	low
Realism	Exaggeration	*	VERB STRENGTH	low	high

3.2 Emotional stability

Emotional stability—or neuroticism—is the second most studied personality trait; it is part of most existing frameworks of personality, such as the Big Five and the PEN model (Norman 1963; Eysenck et al. 1985). Neurotics tend to be anxious, negative and oversensitive, while emotionally stable people are calm and even-tempered. Eysenck et al. (1985) suggest that this dimension is related to activation thresholds in the nervous system, i.e. neurotics turn more easily into a 'fight-or-flight' state when facing danger, resulting in an increase of their heart beat, muscular tension, level of sweating, etc. In order to increase the number of relevant findings, we also include studies focusing on short-lived emotions that are symptomatic of the personality trait (Watson and

Table 14 Example outputs of PERSONAGE for the emotional stability personality model

#	Content plan	End	PERSONAGE's output	Score
1	Compare (Acacia, Marinella)	Low	I might be wrong. I might approve Acacia and Marinella. Err… Acacia provides like, acceptable food. It's in Midtown! It's a bistro and french place. Actually, I mean, Marinella is in Manhattan and an it-it-italian place	3.50
4	Compare (Caffe Cielo, Trattoria Spaghetto)	High	Let's see, Caffe Cielo and Trattoria Spaghetto… Caffe Cielo offers kind of acceptable food and Trattoria Spaghetto just has sort of satisfying food. Basically, I guess they're outstanding restaurants	5.75
5	Recommend (Cent anni)	Low	I am not really sure. Cent'anni is the only restaurant I would recommend. It's an italian place. It offers bad at-at-atmosphere, but it features like, nice waiters, though. It provides good food. I mean, it's bloody expensive. Err… its price is 45 dollars	3.50
8	Recommend (Chimichurri Grill)	High	Let's see what we can find on Chimichurri Grill. Basically, it's the best	6.00

Personality ratings are on a scale from 1 to 7, with 1 = very low and 7 = very high

Clark 1992), e.g. markers of anxiety are considered as valid markers of neuroticism.[9] Table 15 summarizes the linguistic cues for emotional stability and the hypothesized personality models, and Table 14 provides example utterances generated using the personality models. See Mairesse (2008) for more detail. Some example parameters are illustrated in the primarily negative and neutral content with negative content repeated and foregrounded in utterances 1 and 5. The high STUTTERING parameter is also seen in utterances 1 and 5. Weaver (1998) shows that neuroticism is associated with frustration and acquiescence, which we model respectively with high EXPLETIVES and ACKNOWLEDGMENTS parameter values (e.g. *bloody,damn* in utterance 5. We hypothesize that neurotics are more likely to exaggerate when presenting information, based on the impulsiveness facet of that trait. They are thus associated with high EMPHASIZER HEDGES parameter values (e.g. *really*, *actually*) as in utterance 1. Table 14 also shows how neuroticism is conveyed through a high VERBOSITY parameter value, e.g. utterance 5 describes 6 restaurant attributes, whereas utterance 8 refers to only mentions the claim.

3.3 Agreeableness

Agreeable people are generous, optimistic, emphatic, interested in others, and they make people feel comfortable, whereas disagreeable people are self-interested, and do not see others positively. Agreeableness has not been studied as much as extraversion and emotional stability, as it has only emerged with the Big Five framework. As with neuroticism, we include markers of emotions related to agreeableness in our study, such as anger. Table 17 summarizes the linguistic cues for agreeableness and the

[9] The term 'anxiety' is sometimes used to describe either an emotion or a permanent trait, the former is then referred to as *state anxiety* and the latter as *trait anxiety*.

Table 15 Summary of language cues for emotional stability, with corresponding generation parameters

Neurotic findings	Stable findings	Studies	Parameters	Neuro	Emot
Content planning:					
Problem talk, dissatisfaction	Pleasure talk, agreement, compliment	PK99	CONTENT POLARITY	low	high
			REPETITION POLARITY	low	high
			CONCESSION POLARITY	low	high
Direct claim	Inferred claim	*	POSITIVE CONTENT FIRST	high	low
High verbal productivity	Low verbal productivity	SI78	VERBOSITY	high	low
Many lexical repetitions	Few lexical repetitions	OG04b, SC81	REPETITIONS	high	low
Polarized content	Neutral content	*	POLARIZATION	high	low
Stressed	Calm	*	REQUEST CONFIRMATION	low	high
			INITIAL REJECTION	high	low
Syntactic template selection:					
Many self-references	Few self-references	PK99, ME06, OG04b	SELF-REFERENCES	high	low
Problem talk	Pleasure talk	PK99	TEMPLATE POLARITY	low	high
Aggregation:					
Low use of 'punct *which*'	High use of 'punct *which*'	OG04b	RELATIVE CLAUSE	low	high
Many conjunctions	Few conjunctions	OG04a	MERGE	high	low
Few short silent pauses	Many short silent pauses	SI78	CONJUNCTION	low	high
Low use of 'punct *so*'	High use of 'punct *so*'	OG04b	JUSTIFY - SO CUE WORD	low	high
Low use of clause final *also*	High use of clause final *also*	OG04b	INFER - ALSO CUE WORD	low	high
Many inclusive words (e.g. *with*, *and*)	Few inclusive words	OG04b, GO03	WITH CUE WORD	high	low
High use of final *though*	Low use of final *though*	OG04a	CONCEDE - BUT/THOUGH CUE WORD	high	low
Many long silent pauses	Few long silent pauses	SI78	PERIOD	high	low
Many 'non-ah' disfluencies (omission)	Few 'non-ah' disfluencies	SC81**	RESTATE - OBJECT ELLIPSIS	high	low
Pragmatic marker insertion:					
Many pronouns, few articles	Few pronouns, many articles	PK99, OG04a	SUBJECT IMPLICITNESS	low	high
			PRONOMINALIZATION	high	low
			SOFTENER HEDGES:		
Few tentative words	Many tentative words	ME06	·SORT OF, SOMEWHAT, QUITE, RATHER, IT SEEMS THAT,IT SEEMS TO ME THAT, KIND OF	low	high
Many self-reference	Few self-references	PK99, ME06, OG04b	·I THINK THAT	high	low
Many filled pauses (apprehensive)	Few filled pauses	SC79, WE98	FILLED PAUSES:	high	low
			· ERR, I MEAN, MMHM, LIKE		
			ACKNOWLEDGMENTS:		
More acquiescence	Few acquiescence	WE98	·YEAH, RIGHT, OK	high	low
Many self references	Few self references	PK99, ME06, OG04b	·I SEE	high	low
High use of 'punct *well*'	Low use of 'punct *well*'	OG04b	·WELL	high	low
Exaggeration	Realism	*	EMPHASIZER HEDGES:		
			·REALLY, ACTUALLY,	high	low
			·BASICALLY, JUST	low	high
Many rhetorical interrogatives	Few rhetorical interrogatives	*	TAG QUESTION	high	low
Frustration	Less frustration	WE98	EXPLETIVES	high	low
Many 'non-ah' disfluencies (stutter)	Few 'non-ah' disfluencies	SC81**	STUTTERING	high	low
Lexical choice:					
Many frequent words	Few frequent words	OG04b, GO03	LEXICON FREQUENCY	high	low
High concreteness	Low concreteness	OG04b, GO03			
Exaggeration	Realism	*	VERB STRENGTH	high	low

hypothesized personality models, and Table 16 provides example utterances generated using the personality models.

In the information presentation domain, the optimism of agreeable persons is associated with the presentation of positive information first (POSITIVE CONTENT FIRST parameter), while their empathy is conveyed by asking for explicit confirmations as illustrated in utterances 5 and 13 in Table 16, such as *Did you say Cent'anni?* (REQUEST CONFIRMATION parameter). We associate disagreeableness with a high INITIAL REJECTION parameter value, and with a high COMPETENCE MITIGATION parameter (Infante 1995). Disagreeableness biases the selection of syntactic templates towards negatively-connotated templates as in utterance 10 in Table 16. Our disagreeable personality model has a high EXPLETIVES parameter value (see *damn* in utterance 1). Agreeable people use in-group markers and tag questions (see Brown and Levinson

Table 16 Example outputs of PERSONAGE for the agreeableness personality model

#	Content plan	End	PERSONAGE's output	Score
1	Compare (Acacia, Marinella)	Low	Acacia and Marinella are damn alright. I mean, actually, Acacia doesn't provide nasty food. Obviously, it's located in Midtown. Basically, Marinella is in Manhattan	3.25
5	Compare (Acacia, Marinella)	High	Did you say Acacia and Marinella? I imagine you would appreciate them, you see? It seems to me that Marinella provides kind of satisfactory food, also it's an italian place mate, but Acacia offers sort of acceptable food, you know	6.00
10	Recommend (Chimichurri Grill)	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
13	Recommend (Cent anni)	High	Did you say Cent'anni? I imagine you would appreciate it, you see? It seems that this eating place, which provides sort of good food and rather acceptable service, you know, is in Manhattan mate	6.25

Table 17 Summary of language cues for agreeableness, with the corresponding generation parameters

Disagreeable findings	Agreeable findings	Studies	Parameters	Disag	Agree
Content planning:					
Problem talk, dissatisfaction	Pleasure talk, agreement, compliment	PK99, ME06	CONTENT POLARITY	low	high
			REPETITION POLARITY	low	high
			CONCESSION POLARITY	low	high
			POSITIVE CONTENT FIRST	low	high
Fewer empathy	More empathy	*	REQUEST CONFIRMATION	low	high
Many personal attacks (competence)	Few personal attacks	IN95	COMPETENCE MITIGATION	high	low
Many commands, global rejections	Few commands, global rejections	IN95	INITIAL REJECTION	high	low
Syntactic template selection:					
Problem talk	Pleasure talk	PK99, ME06	TEMPLATE POLARITY	low	high
Few self-references	Many self-references	PK99, ME06	SELF-REFERENCES	high	low
Aggregation:					
Many pauses	Few pauses	SI78	PERIOD	high	low
Pragmatic marker insertion:					
Many articles	Few articles	PK99, ME06	SUBJECT IMPLICITNESS	high	low
Many negations	Few negations	ME06	NEGATION	high	low
Many swear words	Few swear words	ME06, IN95	EXPLETIVES	high	low
No politeness form	Minimize negative face threat	*	SOFTENER HEDGES: ·SORT OF, SOMEWHAT, QUITE, RATHER, IT SEEMS THAT, IT SEEMS TO ME THAT, AROUND, KIND OF	low	high
Few insight words	Many insight words (e.g. *see*, *think*)	ME06	·I THINK THAT	low	high
No politeness form	Minimize positive face threat	*	ACKNOWLEDGMENTS: ·YEAH, RIGHT, OK, WELL	low	high
Few insight words	Many insight words	ME06	·I SEE	low	high
No politeness form	Minimize negative face threat	*	EMPHASIZER HEDGES: ·REALLY, BASICALLY, ACTUALLY, JUST	high	low
			FILLED PAUSES:		
No politeness form	Minimize positive face threat	*	·YOU KNOW	low	high
			TAG QUESTION	low	high
			IN-GROUP MARKER	low	high
Lexical choice:					
Few frequent words	Many frequent words	NO06	LEXICON FREQUENCY	low	high
Shorter words	Longer words	ME06, NO06	LEXICON WORD LENGTH	low	high

1987) (e.g. *mate* in utterances 5 and 13 in Table 16). Agreeable speakers are modeled with high LEXICON WORD LENGTH and LEXICON FREQUENCY parameter values, e.g. *satisfactory, acceptable* in utterances 5 and 13 in Table 16.

3.4 Conscientiousness

Conscientiousness is related to the control of one's impulses, resulting in careful, self-disciplined, and success-driven people on the one side, and impulsive, disorganized, and laid-back individuals on the other. Similarly to agreeableness, recent work has studied linguistic correlates of conscientiousness, however it has not been researched as extensively as extraversion.

Table 19 summarizes the linguistic cues for conscientiousness and the hypothesized personality models. Table 18 provides example utterances generated using the personality models. The high REQUEST CONFIRMATION parameter value in Table 19, is illustrated in utterance 14 (*Let's see what we can find*). Unconscientious speakers' high rejection is illustrated with *I am not sure, I might be wrong* in utterances 2 and 9. Because of their thoroughness, we also hypothesize that conscientious speakers use a more formal language, thus producing more formal discourse connectives, e.g. relative clauses (utterance 14). Unconscientious speakers produce more swear words and negations (utterances 2 and 9) (Pennebaker and King 1999; Mehl et al. 2006). We also associate conscientiousness with the use of formal softener hedges and acknowledgment markers (e.g. *rather*, *quite* and *sort of* in utterances 6 and 14).

Table 18 Example outputs of PERSONAGE for the conscientiousness personality model

#	Content plan	End	PERSONAGE's output	Score
2	Compare (Caffe Cielo, Trattoria Spaghetto)	Low	I might be wrong! I mean, there could be worse places. Err… Caffe Cielo is an italian place. Even if it offers like, nice food, the atmosphere isn't good. Trattoria Spaghetto is an italian place and the atmosphere is damn bad, this restaurant provides bad atmosphere	4.50
6	Compare (Caffe Cielo, Trattoria Spaghetto)	High	Let's see, Caffe Cielo and Trattoria Spaghetto… They're rather outstanding. I guess Trattoria Spaghetto offers sort of acceptable food, also it's an italian eating house. Caffe Cielo, which provides quite satisfactory food, is an italian eating place	5.75
9	Recommend (Cent anni)	Low	I am not kind of sure! I mean, Cent'anni's price is 45 dollars, so this restaurant is the only place that is any good, it's damn expensive, this restaurant has nice waiters though mate and the atmosphere isn't good	2.50
14	Recommend (Chimichurri Grill)	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

Table 19 Summary of language cues for conscientiousness, with the corresponding generation parameters

Unconscientious findings	Conscientious findings	Studies	Parameters	Unc	Consc
Content planning:					
Few positive emotion words, many negative emotion words	Many positive emotion words (e.g. *happy*, *good*), few negative emotion words (e.g. *hate*, *bad*)	PK99, ME06	CONTENT POLARITY	low	high
			REPETITION POLARITY	low	high
			CONCESSION POLARITY	low	high
Less perspective	More perspective	*	CONCESSIONS	low	high
Less careful	Check that information is conveyed correctly	*	REQUEST CONFIRMATION	low	high
More vague	Straight to the point	*	RESTATEMENTS	high	low
			REPETITIONS	high	low
			INITIAL REJECTION	high	low
Syntactic template selection:					
Few positive affect	Some positive affect	PK99, ME06	TEMPLATE POLARITY	low	high
Aggregation:					
Many exclusive words (e.g. *but*, *without*)	Few exclusive words	PK99	CONTRAST - ANY CUE WORD	high	low
Many causation words (e.g. *because*, *hence*)	Few causation words	PK99	JUSTIFY - ANY CUE WORD	high	low
Informal	Formal	*	ALTHOUGH, WHILE, SINCE, HOWEVER CUE WORD	low	high
			RELATIVE CLAUSE	low	high
Pragmatic marker insertion:					
Many negations	Few negations	PK99	NEGATION	high	low
Many swear words	Few swear words	ME06	EXPLETIVES	high	low
			NEAR EXPLETIVES	high	low
Many references to friends (e.g. *pal*, *buddy*)	Few references to friends	NO06	IN-GROUP MARKER	high	low
Many disfluencies, filler words	Few disfluencies, filler words	ME06	FILLED PAUSES:	high	low
			· ERR, I MEAN, MMHM, LIKE		
			SOFTENER HEDGES:		
Few insight words	Many insight words	ME06	· I THINK THAT	low	high
Informal	Formal	*	·KIND OF	high	low
			·SOMEWHAT, RATHER, SORT OF, QUITE	low	high
			ACKNOWLEDGMENTS:		
			·I SEE, WELL	low	high
			·YEAH, KIND OF	high	low
Impulsive	Not impulsive	*	EXCLAMATION	high	low
Lexical choice:					
Many frequent words	Few frequent words	*	LEXICON FREQUENCY	high	low
Shorter words	Longer words	ME06	LEXICON WORD LENGTH	low	high

3.5 Openness to experience

Openness to experience is the trait generating the most controversy in the personality psychology community. It is the weakest factor among the Big Five traits, and it has also been referred to as *intellect* or *culture* in some studies (Peabody and Goldberg 1989). People that are open to experience are usually creative, sophisticated, intellectual, curious and interested in art. A lack of openness is thus associated with narrow-mindedness, unimaginativeness and ignorance. As with previous traits, we include findings concerning dimensions that are related to openness to experience/intellect, such as socio-economic background. Table 21 summarizes the linguistic cues for openness to experience and the hypothesized personality models, We associate open speakers with high values for all polarity parameters in Table 21. As speakers from a higher socio-economic background produce more elaborations (Siegman 1978), we associate openness to experience with a higher VERBOSITY and a lower INITIAL REJECTION parameter value, based on the assumption that culture and intellect correlate with socio-economic status. For template selection, we hypothesize that open speakers produce more complex syntactic structures through subordination, resulting in, e.g., more relative clauses.

Table 20 provides example utterances generated using the personality models. The high RELATIVE CLAUSE parameter can be seen in utterances 5 and 14 in Table 20. Our

Table 20 Example outputs of PERSONAGE for the openness to experience personality model

#	Content plan	End	PERSONAGE's output	Score
1	Compare (Acacia, Marinella)	Low	I mean, I might be darn wrong. Mmhm… err … I might suggest Acacia and Marinella. Acacia offers like, nice food, nice food, and Marinella is located in Manhattan	2.5
5	Compare (Acacia, Marinella)	High	You want to know more about Acacia and Marinella? It seems to me that they're kind of outstanding, you know. Acacia, which features rather satisfactory food, is a bistro and french eating house and in Midtown while Marinella is an italian eating place and in Manhattan, you see?	5.25
9	Recommend (Cent anni)	Low	Mmhm…err… I don't know. I mean, I would consider Cent'anni because it has like, good food	2.5
14	Recommend (Chimichurri Grill)	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

open personality model generates implicit references with a high SUBJECT IMPLICITNESS parameter value—e.g. producing *'the food is satisfactory'* as opposed to *'it has satisfactory food'* in utterance 14. Many researchers have suggested that high openness is correlated with a high WORD LENGTH and a low LEXICON FREQUENCY parameter value (Pennebaker and King 1999; Mehl et al. 2006). See *satisfactory* and *second-rate* in utterance 5 and 14.

4 Evaluation experimental design

As we discussed above, there has been considerable prior work on the linguistic expression of stylistic effects (Power et al. 2003; Bouayad-Agha et al. 2000; Hovy 1988; DiMarco and Hirst 1993; Paiva and Evans 2005; Isard et al. 2006). However, there have been relatively fewer evaluations of whether humans perceive the variation as the system intended (Cahn 1990; Fleischman and Hovy 2002; Porayska-Pomsta and Mellish 2004; Cassell and Bickmore 2003; Rambow et al. 2001; Brockmann 2009). Since the expressive effect of linguistic variation—e.g. style, emotion, mood and personality—can only be measured subjectively, an advantage of the Big Five framework is its standard questionnaires for testing the perception of personality (John et al. 1991; Costa and McCrae 1992; Gosling et al. 2003).

Our evaluation of PERSONAGE exploits these questionnaires by asking human judges to rate the personality of a set of generated utterances by completing the Ten-Item Personality Inventory (TIPI) (Gosling et al. 2003). The TIPI instrument minimizes the number of judgments required to elicit personality ratings. To test whether personality

Table 21 Summary of language cues for openness to experience, with corresponding generation parameters

Non-open findings	Open findings	Studies	Parameters	Non-op	Open
Content planning:					
Few positive emotion words	Many positive emotion words (e.g. *happy*, *good*)	NO06	CONTENT POLARITY	low	high
			REPETITION POLARITY	low	high
			CONCESSION POLARITY	low	high
Low meaning elaboration	High meaning elaboration	SI78*,**	VERBOSITY	low	high
			INITIAL REJECTION	high	low
Less perspective	More perspective	*	CONCESSIONS	low	high
Few politeness forms	Many politeness forms	*	REQUEST CONFIRMATION	low	high
Syntactic template selection:					
Few positive emotion words	Many positive emotion words	NO06	TEMPLATE POLARITY	low	high
Many self-references	Few self-references	PK99	SELF-REFERENCES	high	low
Simple construction	Complex constructions	*	SYNTACTIC COMPLEXITY	low	high
Aggregation:					
Few exclusive words	Many exclusive words (e.g. *but*, *without*)	PK99	CONTRAST - ANY CUE WORD	low	high
Many causation words (e.g. *because*, *hence*)	Few causation words	PK99	JUSTIFY - ANY CUE WORD	high	low
Few inclusive words	Many inclusive words (e.g. *with*, *and*)	NO06	WITH CUE WORD	low	high
			CONJUNCTION	low	high
			MERGE	low	high
Simple construction	Complex constructions	*	RELATIVE CLAUSE	low	high
Many planning errors	Few planning errors	*	RESTATE - OBJECT ELLIPSIS	high	low
Pragmatic marker insertion:					
Few articles, many third person pronouns	Many articles, few third person pronouns	PK99, ME06	SUBJECT IMPLICITNESS	low	high
			PRONOMINALIZATION	high	low
Few tentative words	Many tentative words (e.g. *maybe*, *guess*)	PK99, ME06	SOFTENER HEDGES: ·SORT OF, SOMEWHAT, QUITE, RATHER, IT SEEMS THAT, IT SEEMS TO ME THAT, AROUND, KIND OF	low	high
Few insight words	Many insight words (e.g. *think*, *see*)	PK99, ME06	· I THINK THAT	low	high
			ACKNOWLEDGMENTS: ·I SEE	low	high
Many filler words and within-utterance pauses	Few filler words and within-utterance pauses	*,SI78**	FILLED PAUSES: ·ERR, I MEAN, MMHM, LIKE	high	low
Few politeness forms	Many politeness forms	*	TAG QUESTION	low	high
			NEAR EXPLETIVES	high	low
Lexical choice:					
More frequent words, lower age of acquisition	Less frequent words, higher age of acquisition	*, NO06	LEXICON FREQUENCY	high	low
Shorter words	Longer words	PK99, ME06, NO06	LEXICON WORD LENGTH	low	high
Milder verbs	Stronger, uncommon verbs	*	VERB STRENGTH	low	high

can be recognized from a small sample of linguistic output, and to localize the effect of varying particular parameters, the judges evaluated the speaker's personality on the basis of a *single* utterance, i.e. ignoring personality perceptions that could emerge over the course of a dialogue. The judges rated the utterances as if they had been uttered by a friend responding in a dialogue to a request to recommend restaurants. In addition, the judges evaluated the naturalness of each utterance on the same scale. Naturalness was defined in the experimental instructions to mean how likely an utterance is to have been uttered by a real person.

The judges were researchers in psychology, history and anthropology who were familiarized with Big Five trait theory by being provided with associated lists of trait adjectives from Goldberg (1990), as exemplified by the adjectives shown with each trait in Table 3. Because of the high number of control parameters, a large number of utterances was needed to reveal any significance. We thus restricted the number of judges to three in order to ensure consistency over our dataset. The judges were not familiar with language generation engines, nor were they given any information about which linguistic reflexes are associated with different traits. The judges were alone in their own offices when they produced the ratings via the online TIPI, and they did not know each other or discuss their intuitions or judgments with one another. It took the judges approximately 10 to 14 hours, over several weeks of elapsed time to complete

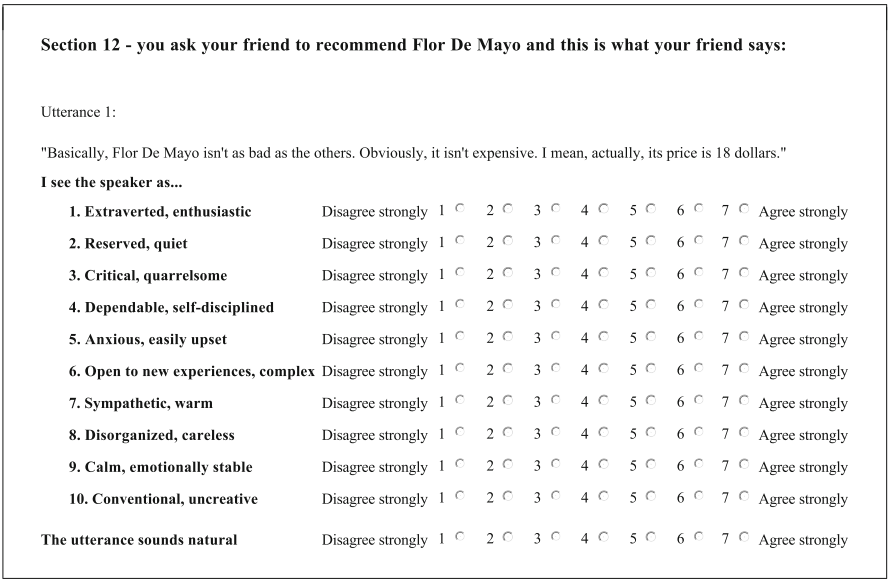
the TIPI for all utterances. The judgments from the three judges were averaged for each utterance to produce a rating for each trait ranging from 1 (e.g. highly neurotic) to 7 (e.g. very stable).

Our main hypothesis is that the personality models specified and illustrated with example output utterances in Tables 13, 14, to 21, can be used to control PERSONAGE's generation process, and the user's perception of the system's personality. There are two personality models for each trait.

However, if we only test utterances produced with the personality models, we cannot tell which parameters in each model are responsible for the judge's perceptions, because the same linguistic cues would be consistently used to convey a given personality. In other words, because the cues covary, it is not possible to identify which cue—or utterance feature—is responsible. Due to the high cost of collecting the personality judgments (each utterance has 11 associated questions), and the large number of parameters, it is not possible to systematically vary each parameter. Therefore, in our evaluation, we combine a sample of utterances generated with random parameter settings with utterances generated using the personality models, and then examine correlations between generation decisions and personality ratings on the random sample. This evaluation method was chosen because of its similarity with a large range of existing correlational studies between personality and human language (Scherer 1979; Furnham 1990; Pennebaker and King 1999).

Because extraversion is the most important of the Big Five traits (Goldberg 1990), three judges evaluated PERSONAGE in a first experiment focusing strictly on that trait (Mairesse and Walker 2007). After positive results were obtained for extraversion, two judges evaluated the four remaining traits in a second experiment. The judges rated a total of 240 utterances based on *personality models* (i.e., predefined parameter values based on Tables 13 and 15), and 320 *random* utterances generated with uniformly distributed parameter values.[10] The personality models parameter values were normally distributed with a 15% standard deviation to increase the range of variation for a given trait. There were 80 utterances generated using personality models for the extraversion experiment and 160 with personality models for the evaluation of the other four traits. Utterances were grouped into 20 sets of utterances generated from the same content plan. Each set contained two utterances per trait (four for extraversion), generated with parameter settings for both the low end and the high end of each dimension, and four random utterances. In each set, the personality model utterances were randomly ordered and mixed with utterances generated with random parameters. The judges rated one randomly ordered set at a time, but viewed all utterances in that set before rating them. All questionnaires were filled online. Figure 9 shows the online TIPI adapted to the evaluation of personality in our domain. A total of 40 utterances were rated for each trait (80 for extraversion), with each half targeting one extreme of the dimension. Multiple outputs were generated for each content plan, trait, and personality model by allowing each parameter setting to be normally distributed with a 15% standard deviation.

[10] The 320 random utterances were rated for extraversion, and half of them were also rated for the remaining traits.



Section 12 - you ask your friend to recommend Flor De Mayo and this is what your friend says:

Utterance 1:

"Basically, Flor De Mayo isn't as bad as the others. Obviously, it isn't expensive. I mean, actually, its price is 18 dollars."

I see the speaker as...

1. Extraverted, enthusiastic	Disagree strongly	1 ○	2 ○	3 ○	4 ○	5 ○	6 ○	7 ○	Agree strongly
2. Reserved, quiet	Disagree strongly	1 ○	2 ○	3 ○	4 ○	5 ○	6 ○	7 ○	Agree strongly
3. Critical, quarrelsome	Disagree strongly	1 ○	2 ○	3 ○	4 ○	5 ○	6 ○	7 ○	Agree strongly
4. Dependable, self-disciplined	Disagree strongly	1 ○	2 ○	3 ○	4 ○	5 ○	6 ○	7 ○	Agree strongly
5. Anxious, easily upset	Disagree strongly	1 ○	2 ○	3 ○	4 ○	5 ○	6 ○	7 ○	Agree strongly
6. Open to new experiences, complex	Disagree strongly	1 ○	2 ○	3 ○	4 ○	5 ○	6 ○	7 ○	Agree strongly
7. Sympathetic, warm	Disagree strongly	1 ○	2 ○	3 ○	4 ○	5 ○	6 ○	7 ○	Agree strongly
8. Disorganized, careless	Disagree strongly	1 ○	2 ○	3 ○	4 ○	5 ○	6 ○	7 ○	Agree strongly
9. Calm, emotionally stable	Disagree strongly	1 ○	2 ○	3 ○	4 ○	5 ○	6 ○	7 ○	Agree strongly
10. Conventional, uncreative	Disagree strongly	1 ○	2 ○	3 ○	4 ○	5 ○	6 ○	7 ○	Agree strongly
The utterance sounds natural	Disagree strongly	1 ○	2 ○	3 ○	4 ○	5 ○	6 ○	7 ○	Agree strongly

Fig. 9 Online version of the Ten-Item Personality Inventory (TIPI) scale (from Gosling et al. 2003), used in our experiments, adapted to the evaluation of personality in generated utterances by judges. The header was modified to ask a judge to evaluate the personality of the speaker, rather than his own personality

5 Experimental results

Section 5.1 reports results showing that the judges agree significantly on their perceptions and that the personality models are perceived as intended. Section 5.2 presents the correlations between parameters used to generate the random utterances and judge's ratings, in order to test generalizations from other genres.

5.1 Personality perceptions

Judges evaluated the naturalness of each utterance, i.e. to what extent it could have been uttered by a human. Results in Table 22 show that the utterances were seen as moderately natural on average, with a mean rating of 4.59 out of 7 for the personality model utterances. Table 22 shows that extravert utterances are rated as the most natural, with an average rating above 5.5 out of 7. Introvert utterances are also perceived as natural, with ratings close to 5. On the other hand, utterances expressing neuroticism or a lack of conscientiousness are rated as moderately unnatural, with average scores below 3.5. A comparison between Tables 25 and 22 suggests a correlation between naturalness and generation accuracy, however it is not clear whether (1) poor personality recognition is a consequence of unnatural utterances, or whether (2) the projection of inconsistent personality cues causes the low naturalness scores, or whether (3) extreme traits are likely to be perceived as unnatural because they are not commonly observed. Table 22 also indicates that the random utterances are rated as less natural than the

Table 22 Average naturalness ratings for the utterance sets generated with the personality models and the random utterances

Personality trait	Low	High	Random
Extraversion	4.93	5.78	4.75
Emotional stability	3.43	4.63	4.72
Agreeableness	3.63	5.56	4.76
Conscientiousness	3.33	5.33	4.61
Openness to experience	3.98	3.85	3.86

Table 23 Krippendorf's α for the rule-based and random utterances

Parameter set	Personality model	Random	All
Extraversion	.72	.27	.40
Emotional stability	.52	.13	.26
Agreeableness	.51	.19	.29
Conscientiousness	.38	.22	.29
Openness to experience	.43	.24	.32

Agreement results under the *All* column were computed over the full dataset. Results for the full dataset (*All*) include cross-trait judgments such as extraversion ratings for utterances with neurotic parameters

utterances generated using personality models. An independent sample t-test shows that this difference is marginally significant ($p = .075$, two-tailed).

Table 23 reports the *inter-rater agreement* over the 3 judges.[11] We use Krippendorf's α coefficient for interval scales, which is suitable for our experiments as (a) it does not depend on the number of judges; (b) it allows the use of continuous scales; and (c) it can handle different numbers of ratings per judge (Krippendorf 1980). We use 5000 rounds of bootstrapping to estimate α's distribution. Column *Personality model* evaluates agreement over the 40 utterances generated with personality models (80 for extraversion), while column *Random* evaluates agreement over 160 random utterances (320 for extraversion). Results show that the level of agreement is higher for utterances generated by personality models (e.g., $\alpha = .72$ for extraversion) than for random utterances ($\alpha = .27$ for extraversion), possibly because random generation decisions are more likely to produce utterances projecting inconsistent personality cues, which may be interpreted in different ways by the judges. The agreement is also higher for extraversion than for any other trait, probably because it the trait which is conveyed the most strongly through language (Pennebaker and King 1999; Mehl et al. 2006). Possibly for the same reason, the lowest level of agreement over personality models is found for conscientiousness and openness to experience ($\alpha = .38$ and $\alpha = .42$). Since personality models are derived from studies focusing on a large range of domains and linguistic genres, it is likely that some of the hypothesized markers do not carry over to our domain. Such markers are therefore likely to confuse the judges. An analysis of the impact of individual parameters on the judges ratings is presented in Sect. 5.2.

[11] There were 4 judges for extraversion, as one judge took part in both extraversion experiments.

Table 24 Average personality ratings for the utterances generated with the low and high personality models for each trait on a scale from 1 to 7

Personality trait	Low	High
Extraversion	2.96	5.98
Emotional stability	3.29	5.96
Agreeableness	3.41	5.66
Conscientiousness	3.71	5.53
Openness to experience	2.89	4.21

The ratings of the two extreme utterance sets differ significantly for all traits ($p < .001$, two-tailed)

Table 25 Generation accuracy (in %) for the utterance sets generated with the low and high parameter settings for each trait

Personality trait	Low	High	Overall
Extraversion	82.5	100.0	91.3
Emotional stability	80.0	100.0	90.0
Agreeableness	70.0	100.0	85.0
Conscientiousness	60.0	100.0	80.0
Openness to experience	90.0	55.0	72.5
All utterances	85.0		

An utterance is correctly recognized if its average rating falls in the half of the scale predicted by its parameter setting. Neutral ratings (4 out of 7) are counted as misrecognitions

Table 24 compares the average ratings of the 20 utterances expressing the low end of each trait and the 20 utterances expressing the high end (40 for extraversion). Paired t-tests show that the judges can discriminate between both extreme utterance sets for each trait ($p < .001$), e.g. utterances predicted to be perceived as introvert received an average rating of 2.96 out of 7, but utterances predicted to be perceived as extravert received an average rating of 5.98. Five traits, with mean rating differences ranging from 1.32 for openness to 3.02 for extraversion. Openness to experience is the hardest trait to convey in our domain, with a rating difference of 1.32 between the utterance sets. This difference, however, is still largely significant ($p < .001$) despite the small number of ratings.

We can also compute *generation accuracy* by splitting ratings into two bins around the neutral rating (4 out of 7), and counting the percentage of utterances with an average rating falling in the bin predicted by its personality model, e.g. an introvert utterance with an average rating of 3.5 would be classified as correctly generated. Since personality models aim to produce utterances manifesting extreme traits, neutral ratings (as well as ratings falling in the wrong bin) count as errors. Table 25 summarizes generation accuracies for both traits, showing that PERSONAGE produces an average accuracy of 85%, i.e. a large majority of the utterances were recognized correctly.

Extraversion is the easiest trait to project in our domain, with ratings covering the full range of the scale and an overall accuracy of 91.3% over both utterance sets (See Table 25). While all emotionally stable utterances were perceived correctly,

20% of the neurotic utterances were rated as neutral or moderately stable: the ratings' distribution of neurotic utterances is slightly biased towards the positive end of the scale. The parameter settings for agreeableness produce utterances covering the largest range of ratings after extraversion (from 1.5 to 6.5), although 30% of the disagreeable utterances were rated as neutral or agreeable. On the other hand, all agreeable utterances were perceived correctly. Unconscientiousness is more difficult to model, as only 60% of the utterances generated with the corresponding parameter setting were perceived as unconscientious, with no average rating below 2.5 out of 7. However, all conscientious utterances have ratings in the positive end of the scale. Openness to experience is the most difficult dimension to evaluate, as misinterpretations occurred for both ends of the scale (10% for non-open utterances, and 45% for open utterances), yielding an average accuracy of 72.5% for this dimension.

Table 25 shows that the positive ends of 4 traits out of 5 are modeled with high precision, while parameter settings for the low ends—typically associated with a low desirability—produce more misrecognized utterances. Openness to experience is the only exception, with a higher accuracy for narrow-minded utterances. This overall trend can be explained by a bias of the judges towards the positive end, as suggested by the overall distributions of ratings (Mairesse 2008). It could also be a consequence of a bias in PERSONAGE's predefined parameter settings, that could be attenuated by recalibrating the parameter values. Finally, it is likely that some aspects of personality cannot be conveyed through language only, or that more than a single utterance is required.

5.2 Correlational analysis of random utterances

As discussed above, the utterances generated from personality models do not allow us to understand which generation decisions are responsible for the judge's ratings. In other words, because the same linguistic cues are consistently used to convey a given personality, it is not possible to identify which cue—or utterance feature—is responsible for observed discrepancies between the target personality and the judges' ratings. Thus we apply the same method as used in psycholinguistic studies: we test correlations between linguistic markers that can be counted and the Big Five traits. Table 26 presents the correlations for the random utterances between the average judges' ratings and generation decisions for each Big Five trait ($p < .01$).[12] Generation decision features are labeled with the parameter's name prefixed with its component in the NLG architecture. The correlations indicate that some generation decisions have higher impact (magnitude of r). The values of `yes` and `no` in the prediction *Pred* column indicate which hypotheses are confirmed, i.e. which predictions from other language genre carry over to our domain. Interestingly, some results also contradict our hypotheses, as indicated by `opp` in the *Pred* columns in Table 26. Although only correlations at the $p < .01$ level are reported, it is important to note that the level of significance

[12] The values used for the generation decision correlations are the actual decisions that were taken in each utterance rather than input parameter values.

Table 26 Correlations between generation decision features and average personality ratings at the $p < .01$ level (* = Bonferroni-corrected $p < .05$)

Generation decision features	*r*	Pred
Extraversion		
PRAGMATIC MARKER- EXCLAMATION	0.34*	Yes
AGGREGATION- INFER	0.21*	No
CONTENT PLANNING- VERBOSITY	0.19*	Yes
CONTENT PLANNING- REQUEST CONFIRMATION: YOU WANT TO KNOW	0.16	Yes
AGGREGATION- CONJUNCTION	0.16	Yes
CONTENT PLANNING- SYNTACTIC COMPLEXITY	0.15	Opp
PRAGMATIC MARKER- FILLED PAUSE: ERR	−0.23*	Yes
Emotional stability		
LEXICAL CHOICE- LEXICON WORD LENGTH	0.25	No
PRAGMATIC MARKER- IN- GROUP MARKER: PAL	0.22	No
PRAGMATIC MARKER- IN- GROUP MARKER	0.20	No
PRAGMATIC MARKER- TAG QUESTION: ALRIGHT	−0.21	Yes
PRAGMATIC MARKER- EXPLETIVES: DAMN	−0.21	Yes
PRAGMATIC MARKER- FILLED PAUSE: ERR	−0.22	Yes
AGGREGATION- RESTATE	−0.23	Yes
CONTENT PLANNING- REPEATED NEGATIVE CONTENT	−0.26	Yes
LEXICAL CHOICE- LEXICON FREQUENCY	−0.28*	Yes
Agreeableness		
CONTENT PLANNING- CONTENT POLARITY	0.49*	Yes
CONTENT PLANNING- POSITIVE CONTENT	0.37*	Yes
PRAGMATIC MARKER- IN- GROUP MARKER	0.33*	Yes
CONTENT PLANNING- POLARIZATION	0.25	No
PRAGMATIC MARKER- IN- GROUP MARKER: PAL	0.24	Yes
LEXICAL CHOICE- LEXICON WORD LENGTH	0.21	Yes
AGGREGATION- CONCEDE- EVEN IF NS	−0.21	No
AGGREGATION- CONTRAST- PERIOD	−0.25	Yes
CONTENT PLANNING- VERBOSITY	−0.28*	No
PRAGMATIC MARKER- FILLED PAUSE: ERR	−0.28*	No
CONTENT PLANNING- CONCESSIONS	−0.29*	No
AGGREGATION- CONCEDE	−0.29*	No
CONTENT PLANNING- REPEATED NEGATIVE CONTENT	−0.32*	Yes
AGGREGATION- CONTRAST- PERIOD	−0.41*	Yes
CONTENT PLANNING- NEGATIVE CONTENT	−0.53*	Yes
Conscientiousness		
PRAGMATIC MARKER- IN- GROUP MARKER	0.23	Opp
CONTENT PLANNING- CONTENT POLARITY	0.21	Yes
PRAGMATIC MARKER- TAG QUESTION	−0.22	No

Table 26 continued

Generation decision features	*r*	Pred
CONTENT PLANNING- - REPETITIONS	−0.22	Yes
CONTENT PLANNING- - CONCESSIONS	−0.22	Opp
CONTENT PLANNING- - NEGATIVE CONTENT	−0.31∗	Yes
Openness to experience		
CONTENT PLANNING- - TEMPLATE POLARITY	0.23	Yes
PRAGMATIC MARKER- - IN- GROUP MARKER	0.22	No

The *Pred* column indicates whether the relation was predicted by the findings in Table 27 (*opp* = predicted opposite relation)

is potentially overestimated due to the large number of significance tests performed, i.e. one pairwise correlation test for each of the 67 hypothesized parameters. Table 26 thus also provides a lower bound on the true significance level by indicating results significant at the $p < .05$ significance level after applying Bonferroni correction for 67 repeated correlation tests. While this conservative correction reduces the risk of spurious findings (type I error), it also increases the risk of failing to observe a significant effect (type II error). See Mairesse (2008) for correlations at lower significance levels.

While most correlations are low, they are in the same range as personality studies on human language (Mehl et al. 2006). Nevertheless, we find that many parameters correlate significantly with personality. Table 26 shows that exclamation marks are the strongest indicators of extraversion—with a correlation of .34 with the average ratings—which is indicative of the assertiveness facet of that trait. As shown by the literature, verbosity is also associated with extraversion with a correlation of .19 (Furnham 1990; Mehl et al. 2006), however the use of the INFER rhetorical relation—joining propositions together without emphasis—produces a higher association, suggesting that extraverts do not put pieces of information into perspective.[13] While extraverts talk about a large number of attributes, their verbosity is also expressed through explicit confirmations ($r = .16$), possibly because they are cues to the expression of sympathy previously associated with that trait (Weaver 1998). Extravert utterances also contain more conjunctions (Oberlander and Gill 2004b). Table 26 show that filled pause *err* is the strongest indicator of introversion, with a correlation of $-.23$, confirming previous findings from Scherer (1979) and Siegman and Pope (1965).

The correlations indicate that neuroticism is associated with the use of short, frequent words ($r = .25$ and $r = -.28$), which confirms previous results on emails (Gill and Oberlander 2003). Interestingly, in-group markers indicate emotional stability (especially *pal*), while filled pauses (i.e. *err*) and repetitions indicate neuroticism. The latter are used to convey the apprehension previously associated with that trait (Scherer 1979; Weaver 1998). As in other genres, negative content and swear words

[13] To improve readability throughout this paper, the perception of the judges regarding a personality type is referred to as a characteristic of individuals that possess that personality trait, e.g. *extravert utterances*, *extravert speakers* and *extraverts* are used interchangeably.

are also associated with a lack of stability (Pennebaker and King 1999), with a stronger association for the expletive *damn* ($r = -.21$).

Table 26 shows that agreeableness is the trait presenting the highest correlation with language generation decisions. Polarity is the most important indicator of agreeableness ($r = .49$), especially when repeating negative content to project disagreeableness ($r = -.53$). The second most important marker of disagreeableness is the use of the PERIOD operation for contrasting propositions ($r = -.41$), which can be perceived as a long, unfilled pause. As suggested by the literature on politeness (Brown and Levinson 1987), in-group markers also project agreeableness ($r = .33$), especially *pal*. Negative content is strongly associated with a lack of conscientiousness ($r = -.31$), as well as concessions, repetitions, and tag questions ($r = -.22$). Interestingly, in-group markers are found to correlate positively with conscientiousness ($r = .23$), contradicting previous findings on weblogs (Nowson 2006). We only find two parameters correlating with openness to experience at the $p < .01$ level, namely positively-connotated recommendations and in-group markers. Openness to experience is the hardest trait to model in our domain, with correlations below .24, possibly because (a) it is the most controversial of the Big Five traits (Goldberg 1990), and (b) it is by definition conveyed more strongly through ones long-term actions rather than language.

This correlational analysis provides insight into which generation parameters help the judges to discriminate between various traits. The knowledge of strong markers of personality is useful for controlling the generation process. More importantly, these correlations show clearly that the findings from other genres of language that we summarized in Table 27 in the Appendix generalize to our domain. Interestingly, we also find that many new markers emerge, while some results contradict our hypotheses (i.e. indicated by *opp* in the *Pred* columns). Future work could thus enhance PERSONAGE's rule-based approach based on the correlations presented here, by taking domain-specific information into account to refine the predefined parameter settings derived from psychological studies.

6 Discussion and conclusion

We present and evaluate PERSONAGE, a highly parameterizable generator that produces outputs that are reliably perceived by human judges as expressing Big Five personality traits. We believe that such a generation capability is a necessary step towards personality-based user adaptation. This paper makes four contributions:

1. We present a systematic review of psycholinguistic findings, organized by the NLG reference architecture;
2. We propose a mapping from these findings to generation parameters for each NLG module and a real-time implementation of a generator using these parameters.[14] Our parameters are defined in terms of well-defined operations on standard semantic and syntactic representations, which should therefore be replicable in other systems;

[14] An online demo is available at http://nlds.soe.ucsc.edu/personage.

3. We present an evaluation experiment showing that we can use personality models based on psycholinguistic findings to control the parameters, in order to produce recognizable linguistic variation for all Big Five personality traits;
4. We analyze the correlations between judges' ratings of personality and PERSONAGE generation decisions, showing which linguistic reflexes of personality generalize from naturally-occurring genres to our application domain.

Our evaluation shows that human judges reliably interpret PERSONAGE's personality cues. While there has been considerable prior work on the linguistic expression of stylistic effects (Power et al. 2003; Bouayad-Agha et al. 2000; Hovy 1988; DiMarco and Hirst 1993; Paiva and Evans 2005; Isard et al. 2006), many of the parameters that are systematically and replicably implemented in PERSONAGE, such as hedges, negation insertion, tag questions and polarity, have never been implemented within the standard NLG architecture. Many of our parameters are not only useful for generating language expressing personality, but could also be used for other types of affective generation, such as politeness (Walker et al. 1997; Porayska-Pomsta and Mellish 2004; Gupta et al. 2008; Wang et al. 2005), or formality (DiMarco and Hirst 1993), if the appropriate models for controlling these parameters were developed. For example, hedges and tag questions can convey politeness or status (Prince et al. 1982; Lakoff 1973a,b; Brown and Levinson 1987; Brennan and Ohaeri 1994, 1999).

Another novel aspect of PERSONAGE is the idea of indexing and selecting content by polarity. In every personality model we tested, content selection according to polarity has a significant effect on human perceptions of personality. This content selection mechanism suggests that there are many potential ways to index and discriminate content, in order to make different versions of a story, a play, or a tutorial, or indeed any dialogue. For example, other work implicitly distinguishes content according to how "personal" particular questions or statements might be in a conversational context (Mateas and Stern 2003; Cassell and Bickmore 2003), or how "threatening" a teacher's criticism might be, using ideas from politeness theory (Porayska-Pomsta and Mellish 2004; Wang et al. 2005). Thus the idea of content that is interchangeable and selectable according to particular social or pragmatic criteria is potentially very powerful.

To our knowledge, the only other generation system to be evaluated in a similar vein is CRAG- 2, a system generating movie review dialogues (Brockmann 2009). In a first experiment, Brockmann presents human judges with dialogues combining utterances selected from an annotated corpus. Results show that the judges perceive variations between extravert and introvert utterances correctly ($p < .001$), however results for emotional stability are not significant. Interestingly, the introduction of n-gram language models for re-ranking paraphrases generated from logical forms produces non-significant results for both traits.[15] Brockmann hypothesizes that this is a consequence of the bias of n-gram language models towards shorter utterances. Although future work should investigate other data-driven methods for stylistic generation, these results suggest that controlling the target personality from within

[15] Agreeableness is the only trait that is perceived correctly above chance level, however that trait is not evaluated in the first experiment.

the generation process is beneficial both in terms of perceptual accuracy and efficiency.

There are a number of issues that deserve further research. We examined only the effect of manipulations of linguistic form, and tested these manipulations by asking judges to read the generated utterances. However, prior research suggests that personality affects dialogue strategy, prosody, and gesture (Vogel and Vogel 1986; Scherer 1979). Our approach could be extended to the parameterization of these other modules.

Another limitation of this work is that we treat personality as a discrete phenomenon, with personality models controlling generated utterances expressing either the low or the high end of each personality trait, and only one trait at a time. This capability can be used for dialogue system adaptation in systems supporting a limited range of user models, or other applications that do not require fine-grained variation of the generation output, e.g. artificial characters with static behavior. However, the wide range of individual differences reflected by the literature on the Big Five (Allport and Odbert 1936; Norman 1963; Goldberg 1990) as well as recent work in medical research (Marcus et al. 2006) suggest that personality varies continuously. This continuity is also reflected by the continuous scales used in personality psychology instruments (John et al. 1991; Costa and McCrae 1992; Gosling et al. 2003). In other work, we investigate methods for producing language targeting any arbitrary value on the Big Five dimensions (Mairesse and Walker 2008b).

Additionally, our approach currently does not offer fine-grained control of the control of various pragmatic markers, but this might be needed to increase naturalness. For example, previous work suggests constraints on the placement of cue words that we do not capture, such as avoiding the repetition of the same cue within a single turn (Moser and Moore 1995; Di Eugenio et al. 1997).

Finally, while PERSONAGE's generation decisions were implemented with domain independence in mind, the effort required to port PERSONAGE to new application domains remains to be evaluated. While our approach can trivially be extended to other information presentation domains by modifying the generation dictionary (e.g., to present information about hotels, films, trains, etc.), extending our approach to new communicative goals (e.g., requests) is likely to require new syntactic transformation rules. However, we believe that such rules can be implemented once for all for a large range of communicative goals and re-used across applications.

Our long term goal is to adapt to the user's personality and linguistic style *during* dialogic interaction. We have started to explore PERSONAGE's generalization capabilities, for interactive drama systems and personal assistants (Mairesse and Walker 2008a; Walker et al. 1997). Personalization is often an important technical requirement for such applications (Murray 1997; Hayes-Roth and Brownston 1994; Mott and Lester 2006). Gill et al. (2004) show that entrainment can be measured by personality variables, and other authors have shown that entrainment takes place in naturally occurring dialogue at all levels of linguistic production (Darves and Oviatt 2002; Stenchikova and Stent 2007; Reitter et al. 2006; Nenkova et al. 2008; Isard et al. 2006; Schober and Brennan 2003). In other work, we have developed models and techniques for recognizing the user's personality from conversational data (Mairesse et al. 2007); these models could be used to produce a system that models similarity-attraction (Byrne

and Nelson 1965; Nass and Lee 2001) and task-specific personality adaptation, based on the adaptation policies outlined in Sect. 1. PERSONAGE provides many parameters that can be dynamically varied in real time; this is essential for adapting to the user during a conversation. In future work, we plan to use PERSONAGE to evaluate the effect of lexical, syntactic, and personality-based adaptation on various dialogue system tasks.

Acknowledgements This research was carried out at the University of Sheffield, where the authors were supported under a Vice Chancellor's studentship and Walker's Royal Society Wolfson Research Merit Award.

Appendix

See Table 27

Table 27 Psychological studies on the identification of personality markers in language

Study ref	Authors	Language source	Cues	Assessment method	Personality dimensions
CO69	Cope (1969)	spoken	output size, type-token ratio	self-report	extraversion
DF99	Dewaele and Furnham (1999)*	spoken	various	self-report	extraversion
FU90	Furnham (1990)*	spoken	speech, linguistic markers	self-report	extraversion, type A behavior, self-monitoring
GO03	Gill and Oberlander (2003)	emails	part-of-speech n-gram counts	self-report	extraversion, neuroticism
HD02	Heylighen and Dewaele (2002)	essays, oral examinations	measure of formality	self-report	extraversion
IN95	Infante (1995)*	spoken	communicative behavior	emotion induction	verbal aggressiveness
ME06	Mehl et al. (2006)	daily-life conversations	content-analysis category counts	observer, self-report	Big Five traits
NO06	Nowson (2006)	blogs	content-analysis category and n-gram counts	self-report	Big Five traits
OG04a	Oberlander and Gill (2004a)	emails	part-of-speech n-grams	self-report	extraversion, neuroticism, psychoticism
OG04b	Oberlander and Gill (2004b)	emails	content-analysis category and n-gram counts	self-report	extraversion, neuroticism
OG06	Oberlander and Gill (2006)	emails	content-analysis category and n-gram counts	self-report	extraversion, neuroticism, psychoticism
PK99	Pennebaker and King (1999)	essays	content-analysis category counts	self-report	Big Five traits
SC79	Scherer (1979)*	spoken	speech markers	self-report, emotion induction	extraversion, emotional stability, anxiety *inter alia*
SC81	Scherer (1981)*	spoken	speech markers	various	stress, anxiety
SI78	Siegman (1978)*	spoken	speech markers	various	socio-economic background, extraversion, anxiety, anger, *inter alia*
SP65	Siegman and Pope (1965)	spoken	verbal fluency	self-report	extraversion
TH87	Thorne (1987)	spoken	polarity, focus	self-report	extraversion
WE98	Weaver (1998)	questionnaires	communicative behavior	self-report	extraversion, neuroticism, psychoticism

References

Aaker, J.L.: The malleable self: The role of self-expression in persuasion. J. Mark. Res. **36**(1), 45–57 (1999)

Allport, G.W., Odbert, H.S.: Trait names: A psycho-lexical study. Psychol. Monogr. **47**(1, Whole No. 211), 171–220 (1936)

André, E., Rist, T., van Mulken, S., Klesen, M., Baldes, S.: The automated design of believable dialogues for animated presentation teams. In: Prevost, S., Cassell, J.S.J., Churchill, E. (eds.) Embodied Conversational Agents, pp. 220–255. MIT Press, Cambridge (2000)

Ardissono, L., Goy, A., Petrone, G., Segnan, M., Torasso, P.: Intrigue: personalized recommendation of tourist attractions for desktop and hand held devices. Appl. Artif. Intell. **17**(8), 687–714 (2003)

Argamon, S., Dhawle, S., Koppel, M., Pennebaker, J.W.: Lexical predictors of personality type. In: Proceedings of the Joint Annual Meeting of the Interface and the Classification Society of North America, St. Louis, MO (2005)

Ball, G., Breese, J.: Emotion and personality in a conversational character. In: Proceedings of the Workshop on Embodied Conversational Characters, pp. 83–86. Lake Tahoe, CA (1998)

Beaman, K.: Coordination and subordination revisited: Syntactic complexity in spoken and written narrative discourse. In: Tannen, D., Freedle, R. (eds.) Coherence in Spoken and Written Discourse, pp. 45–80. Ablex, Norwood, NJ (1984)

Belz, A.: Statistical generation: Three methods compared and evaluated. In: Proceedings of the 10th European Workshop on Natural Language Generation, pp. 15–23. Aberdeen (2005)

Belz, A.: Automatic generation of weather forecast texts using comprehensive probabilistic generation-space models. Nat. Lang. Eng. **14**(4), 431–455 (2008)

Bouayad-Agha, N., Scott, D.R., Power, R.: Integrating content and style in documents: A case study of patient information leaflets. Inf. Des. J. **9**(2–3), 161–176 (2000)

Bouchard, T.J., McGue, M.: Genetic and environmental influences on human psychological differences. J. Neurobiol. **54**, 4–45 (2003)

Brennan, S.E.: Conversations with and through computers. User Model. User-Adap. Inter. **1**, 67–86 (1991)

Brennan, S.E.: Lexical entrainment in spontaneous dialog. In: Proceedings of the International Symposium on Spoken Dialogue. pp. 41–44. Philadelphia, PA (1996)

Brennan, S.E., Clark, H.H.: Conceptual pacts and lexical choice in conversation. J. Exp. Psychol. Learn. Mem. Cogn. **22**, 1482–1493 (1996)

Brennan, S., Ohaeri, J.: Effects of message style on users' attributions toward agents. In: Proceedings of the Conference on Human Factors in Computing Systems, pp. 281–282. Boston, MA (1994)

Brennan, S., Ohaeri, J.: Why do electronic conversations seem less polite? The costs and benefits of hedging. In: Proceedings of the International Joint Conference on Work Activities, Coordination, and Collaboration, pp. 227–235. San Francisco, CA (1999)

Brockmann, C.: Personality and Alignment Processes in Dialogue: Towards a Lexically-Based Unified Model. PhD thesis, School of Informatics, University of Edinburgh (2009)

Brown, P., Levinson, S.: Politeness: Some Universals in Language Usage. Cambridge University Press, Cambridge (1987)

Byrne, D., Nelson, D.: Attraction as a linear function of proportion of positive reinforcements. J. Pers. Soc. Psychol. **1**, 659–663 (1965)

Cahn, J.E.: The Generation of Affect in Synthesized Speech. J. Am. Voice I/O Soc. **8**, 1–19 (1990)

Carberry, S.: Plan recognition and its use in understanding dialogue. In: Kobsa, A., Wahlster, W. (eds.) User Models in Dialogue Systems, pp. 133–162. Springer Verlag, Berlin (1989)

Carenini, G., Moore, J.D.: A strategy for generating evaluative arguments. In: Proceedings of International Conference on Natural Language Generation, pp. 47–54. Mitzpe Ramon, Israel (2000)

Carenini, G., Moore, J.D.: Generating and evaluating evaluative arguments. Artif. Intel. J. **170**(11), 925–952 (2006)

Cassell, J., Bickmore, T.: Negotiated collusion: Modeling social language and its relationship effects in intelligent agents. User Model. User-Adap. Inter. **13**, 89–132 (2003)

Chklovski, T., Pantel, P.: VERBOCEAN: Mining the web for fine-grained semantic verb relations. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 33–40. Barcelona (2004)

Cohen, P.R., Perrault, C.R., Allen, J.F.: Beyond question answering. In: Lehnert, W., Ringle, M. (eds.) Strategies for Natural Language Processing, pp. 245–274. Lawrence Erlbaum Associates, Hillsdale, NJ (1982)

Cope, C.: Linguistic structure and personality development. J. Couns. Psychol. **16**, 1–19 (1969)
Costa, P.T., McCrae, R.R.: NEO PI-R Professional Manual. Psychological Assessment Resources, Odessa, FL (1992)
Darves, C., Oviatt, S.: Adaptation of users' spoken dialogue patterns in a conversational interface. In: Proceedings of the International Conference on Spoken Language Processing (ICSLP), pp. 561–564. Denver, CO (2002)
Department of the Army: Police intelligence operations. Field Manual FM 3-19.50. Appendix D: Tactical Questioning (2006)
Dewaele, J.-M., Furnham, A.: Extraversion: The unloved variable in applied linguistic research. Lang. Learn. **49**(3), 509–544 (1999)
Di Eugenio, B., Moore, J.D., Paolucci, M.: Learning features that predict cue usage. In: Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics (ACL), pp. 80–87. Madrid (1997)
DiMarco, C., Hirst, G.: A computational theory of goal-directed style in syntax. Comput. Liguist. **19**(3), 451–499 (1993)
Dunn, G., Wiersema, J., Ham, J., Aroyo, L.: Evaluating interface variants on personality acquisition for recommender systems. In: Proceedings of the International Conference on User Modeling, Adaptation, and Personalization, pp. 259–270. Trento (2009)
Eysenck, S.B.G., Eysenck, H.J., Barrett, P.: A revised version of the psychoticism scale. Pers. Indivd. Differ. **6**(1), 21–29 (1985)
Fellbaum, C.: WordNet: An Electronic Lexical Database. MIT Press, Cambridge, MA (1998)
Fennis, B.M., Pruyn, A.T.H.: You are what you wear: Brand personality influences on consumer impression formation. J. Bus. Res. **60**(6), 634–639 (2007)
Finin, T.W., Joshi, A.K., Webber, B.L.: Natural language interactions with artificial experts. Proc. IEEE **74**(7), 921–938 (1986)
Fink, J., Kobsa, A.: Review and analysis of commercial user modeling servers for personalization on the world wide web. User Model User-Adap. Inter. **10**(2), 209–249 (2000)
Fleischman M., Hovy E.: Towards emotional variation in speech-based natural language generation. In: Proceedings of the International Conference on Natural Language Generation, pp. 57–64, Harriman, NY (2002)
Forbes-Riley, K., Litman, D.J.: Investigating human tutor responses to student uncertainty for adaptive system development. In: Lecture Notes in Computer Science, vol. 4738, pp. 678–689 (2007)
Forbes-Riley, K., Litman, D.J., Rotaru, M.: Responding to student uncertainty during computer tutoring: an experimental evaluation. In: Lecture Notes in Computer Science, vol. 5091, pp. 60–69 (2008)
Furnham, A.: Language and personality. In: Giles, H., Robinson, W. (eds.) Handbook of Language and Social Psychology, Winley, UK (1990)
Furnham, A. , Jackson, C.J., Miller, T.: Personality, learning style and work performance. Pers. Individ. Differ. **27**, 1113–1122 (1999)
Giles, H., Coupland, N., Coupland, J. : Accommodation theory: Communication, context, and consequence. In: Giles, H., Coupland, N., Coupland, J. (eds.) Contexts of Accommodation: Developments in Applied Sociolinguistics, chap. 1, Cambridge University Press, Cambridge (1991)
Gill, A.J., Oberlander, J.: 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, pp. 456–461, Boston, MA (2003)
Gill, A.J., Harrison, A.J., Oberlander, J.: Interpersonality: Individual differences and interpersonal priming. In: Proceedings of the 26th Annual Conference of the Cognitive Science Society, pp. 464–469. Chicago, IL (2004)
Goldberg, L.R.: An alternative "description of personality": The Big-Five factor structure. J. Pers. Soc. Psychol. **59**, 1216–1229 (1990)
Gosling, S.D., Rentfrow, P.J., Swann, W.B.: A very brief measure of the big five personality domains. J. Res. Pers. **37**, 504–528 (2003)
Green, S. J., DiMarco, C.: Stylistic decision-making in natural language generation. In: Lecture Notes in Computer Science, Trends in Natural Language Generation An Artificial Intelligence Perspective, vol. 1036, pp. 125–143 (1996)
Grosz B.J.: TEAM: A transportable natural language interface system. In: Proceedings of the Conference on Applied Natural Language Processing, pp. 39–45, Santa Monica, CA (1983)

Gupta, S., Walker, M.A., Romano, D.M.: How rude are you?: Evaluating politeness and affect in interaction. In: Proceedings of ACII, pp. 203–217. Lisbon (2007)
Gupta, S., Walker, M.A., Romano, D.M.: POLLy: A conversational system that uses a shared, representation to generate action and social language. In: Proceedings of the Third International Joint Conference on Natural Language Processing (IJCNLP), pp. 203–217. Hyderabad (2008)
Hayes-Roth, B., Brownston, L.: Multiagent collaboration in directed improvisation. Technical Report KSL 94-69, Knowledge Systems Laboratory. Stanford University (1994)
Heylighen, F., Dewaele, J.-M.: Variation in the contextuality of language: An empirical measure. Context Context Spec. Issue Found. Sci. **7**(3), 293–340 (2002)
Higashinaka, R., Walker, M.A., Prasad, R.: An unsupervised method for learning generation lexicons for spoken dialogue systems by mining user reviews. ACM Trans. Speech Lang. Process. **4**(4), 8 (2007)
Hirschberg, J.: Speaking more like you: Lexical, acoustic/prosodic, and discourse entrainment in spoken dialogue systems. In: Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue, pp. 128. Columbus, OH (2008)
Hovy, E.: Generating Natural Language Under Pragmatic Constraints. Lawrence Erlbaum Associates, USA (1988)
Hu, M., Liu, B.: Mining and summarizing customer reviews. In: Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining (KDD), pp. 168–177. Seattle, WA (2004)
Hubal, R.C., Kizakevich, P.N., Guinn, C.I., Merino, K.D., West, S.L.: The virtual standardized patient: Simulated patient-practitioner dialogue for patient interview training. In: Westwood, J.D.,Hoffman, H.M., Mogel, G.T., Robb, R.A., Stredney, D. (eds.) Envisioning Healing: Interactive Technology and the Patient-Practitioner Dialogue, IOS Press, Amsterdam (2000)
Infante, D.A.: Teaching students to understand and control verbal aggression. Commun. Edu. **44**(1), 51–63 (1995)
Inkpen, D.Z., Hirst, G.: Near-synonym choice in natural language generation. In: Nicolov, N., Bontcheva, K., Angelova, G., Mitkov, R. (eds.) Recent Advances in Natural Language Processing III, pp. 141–152. John Benjamins Publishing Company, Amsterdam/Philadelphia (2004)
Isard, A., Brockmann, C., Oberlander, J.: Individuality and alignment in generated dialogues. In: Proceedings of the 4th International Natural Language Generation Conference (INLG), pp. 22–29. Sydney (2006)
Isbister, K., Nass, C.: Consistency of personality in interactive characters: Verbal cues, non-verbal cues, and user characteristics. Int. J. Hum.-Comput. Stud. **53**(2), 251–267 (2000)
John, O.P., Donahue, E.M., Kentle, R.L.: The "Big Five" Inventory: Versions 4a and 5b. Technical report. Berkeley: University of California, Institute of Personality and Social Research (1991)
John, O.P., Srivastava, S.: The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In: Pervin, L.A., John, O.P. (eds.) Handbook of Personality Theory and Research, Guilford Press, New York (1999)
Kittredge, R., Korelsky, T., Rambow, O.: On the need for domain communication knowledge. Comput. Intell. **7**(4), 305–314 (1991)
Kobsa, A.: Personalized hypermedia and international privacy. Commun. ACM **45**(5), 64–67 (2002)
Kobsa, A., Wahlster, W. (eds.): User Models in Dialog Systems. Springer Verlag, Berlin (1989)
Krippendorf, K.: Content Analysis: An Introduction to its Methodology. Sage Publications, Beverly Hills, CA (1980)
Lakoff, R.: Language and woman's place. Lang. Soc. **2**, 45–79 (1973a)
Lakoff, R.: The logic of politeness; or, minding your P's and Q's. In: Papers from the 9th Regional meeting of the Chicago Linguistic Society, pp. 292–305. Chicago Linguistic Society, Chicago, IL (1973b)
Langkilde, I., Knight, K.: Generation that exploits corpus-based statistical knowledge. In: Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics (ACL), pp.704–710. Montreal (1998)
Langkilde-Geary, I.: An empirical verification of coverage and correctness for a general-purpose sentence generator. In: Proceedings of the International Conference on Natural Language Generation, pp. 17–24. Harriman, NY (2002)
Lavoie, B., Rambow, O.: A fast and portable realizer for text generation systems. In: Proceedings of the 3rd Conference on Applied Natural Language Processing, pp. 265–268. Washington, D.C. (1997)
Lester, J.C., Stone, B., Stelling, G.: Lifelike pedagogical agents for mixed-initiative problem solving in constructivist learning environments. User Model. User-Adapt. Inter. **9**(1–2), 1–44 (1999a)

Lester, J.C., Towns, S.G., Fitzgerald, P.J.: Achieving affective impact: Visual emotive communication in lifelike pedagogical agents. Int. J. Artif. Intell. Edu. **10**(3–4), 278–291 (1999)
Levelt, W.J.M., Kelter, S.: Surface form and memory in question answering. Cogn. Psychol. **14**, 78–106 (1982)
Lin, J.: Using distributional similarity to identify individual verb choice. In: Proceedings of the 4th International Natural Language Generation Conference, pp. 33–40. Sydney (2006)
Litman, D.J., Allen, J.E.: A plan recognition model for subdialogues in conversations. Cogn. Sci. **11**, 163–200 (1987)
Loyall, A.B., Bates, J.: Behavior-based language generation for believable agents. Technical Report CMU-CS-95-139. School of Computer Science, Carnegie Mellon University (1995)
Mairesse, F.: Learning to Adapt in Dialogue Systems: Data-driven Models for Personality Recognition and Generation. PhD thesis, Department of Computer Science, University of Sheffield (2008)
Mairesse, F., Walker, M.A.: PERSONAGE: Personality generation for dialogue. In: Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL), pp. 496–503. Prague (2007)
Mairesse, F., Walker, M.A.: A personality-based framework for utterance generation in dialogue applications. In: Proceedings of the AAAI Spring Symposium on Emotion, Personality, and Social Behavior. Palo Alto, CA (2008a)
Mairesse, F., Walker, M.A.: Trainable generation of Big-Five personality styles through data-driven parameter estimation. In: Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics (ACL), pp. 165–173. Columbus, OH (2008b)
Mairesse, F., Walker, M.A., Mehl, M.R., Moore, R.K.: Using linguistic cues for the automatic recognition of personality in conversation and text. J. Artif. Intell. Res. (JAIR) **30**, 457–500 (2007)
Mann, W.C., Thompson, S.A.: Rhetorical structure theory. Toward a functional theory of text organization. Text **8**(3), 243–281 (1988)
Marcu, D.: Building up rhetorical structure trees. In: Proceedings of the 13th National Conference on Artificial Intelligence (AAAI), vol. 2, pp. 1069–1074. Portland, OR (1996)
Marcus, D.K., Lilienfeld, S.O., Edens, J.F., Poythress, N.G.: Is antisocial personality disorder continuous or categorical? A taxometric analysis. Psychol. Med. **36**(11), 1571–1582 (2006)
Mateas, M., Stern, A.: Façade: An experiment in building a fully-realized interactive drama. In: Proceedings of the Game Developers Conference, Game Design track. San Jose, CA (2003)
McCrae, R.R., Costa, P.T.: Validation of the five-factor model of personality across instruments and observers. J. Pers. Soc. Psychol. **52**, 81–90 (1987)
Mehl, M.R., Gosling, S.D., Pennebaker, J.W.: Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. J. Pers. Soc. Psychol. **90**, 862–877 (2006)
Melčuk, I.A.: Dependency Syntax: Theory and Practice. State University of New York, Albany, NY (1988)
Moore, J.D., Paris, C.L.: Planning text for advisory dialogues: Capturing intentional and rhetorical information. Comput. Linguist. **19**(4), 651–694 (1993)
Moser, M.G., Moore J.D.: Investigating cue selection and placement in tutorial discourse. In: Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics (ACL), pp. 130–137. Cambridge, MA (1995)
Mott, B., Lester, J.: U-director: A decision-theoretic narrative planning architecture for storytelling environments. In: Proceedings of the 5th international joint conference on Autonomous agents and multiagent systems, pp. 977–984. Hakodate, Japan (2006)
Murray, J.H.: Hamlet on the Holodeck: The Future of Narrative in Cyberspace. The Free Press, New York (1997)
Nass, C., Lee, K.: Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. J. Exper. Psychol.: Appl. **7**(3), 171–181 (2001)
Nenkova, A., Gravano, A., Hirschberg, J.: High frequency word entrainment in spoken dialogue. In: Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics (ACL), pp. 169–172. Columbus, OH (2008)
Niederhoffer, K.G., Pennebaker, J.W.: Linguistic style matching in social interaction. J. Lang. Soc. Psychol. **21**, 337–360 (2002)
Norman, W.T.: Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality rating. J. Abnorm. Soc. Psychol. **66**, 574–583 (1963)
Nowson, S.: The Language of Weblogs: A study of genre and individual differences. PhD thesis, School of Informatics, University of Edinburgh (2006)

Oberlander, J., Gill, A.J.: Individual differences and implicit language: Personality, parts-of-speech, and pervasiveness. In: Proceedings of the 26th Annual Conference of the Cognitive Science Society, pp. 1035–1040. Chicago, IL (2004a)

Oberlander, J., Gill, A.J.: Language generation and personality: Two dimensions, two stages, two hemispheres?. In: Proceedings from the AAAI Spring Symposium on Architectures for Modeling Emotion: Cross-Disciplinary Foundations, pp. 104–111. Palo Alto, CA (2004b)

Oberlander, J., Gill, A.J.: Language with character: A stratified corpus comparison of individual differences in e-mail communication. Discourse Proces. **42**, 239–270 (2006)

Oberlander, J., Nowson, S.: Whose thumb is it anyway? Classifying author personality from weblog text. In: Proceedings of the 44th Annual Meeting of the Association for Computational Linguistics (ACL), pp. 627–634. Sydney (2006)

Paiva, D.S., Evans, R.: Empirically-based control of natural language generation. In: Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL), pp. 58–65. Ann Arbor, MI (2005)

Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up? Sentiment classification using machine learning techniques. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 79–86. Philadelphia, PA (2002)

Paris, C., Scott, D.R.: Stylistic variation in multilingual instructions. In: Proceedings of the 7th International Workshop on Natural Language Generation, pp. 45–52. Kennebunkport, MN (1994)

Peabody, D., Goldberg, L.R.: Some determinants of factor structures from personality-trait descriptor. J. Pers. Soc. Psychol. **57**(3), 552–567 (1989)

Pennebaker, J.W., King, L.A.: Linguistic styles: Language use as an individual difference. J. Pers. Soc. Psychol. **7**, 1296–1312 (1999)

Pickering, M., Garrod, S.: Towards a mechanistic theory of dialogue. Behav. Brain Sci. **27**, 169–226 (2004)

Piwek, P.: A flexible pragmatics-driven language generator for animated agents. In: Proceedings of Annual Meeting of the European Chapter of the Association for Computational Linguistics (EACL), pp. 151–154. Budapest (2003)

Plummer, J.T.: How personality makes a difference. J. Advert. Res. **24**, 27–31 (1984)

Porayska-Pomsta, K., Mellish, C.: Modelling politeness in natural language generation. In: Proceedings of the International Conference on Natural Language Generation, pp. 141–150. Sydney (2004)

Power, R.: A Computer Model of Conversation. PhD thesis, Department of Machine Intelligence, University of Edinburgh (1974)

Power, R., Scott, D.R., Bouayad-Agha, N.: Generating texts with style. In: Computational Linguistics and Intelligent Text Processing. Lecture Notes in Computer Science, vol. 2588, pp. 93–105 (2003)

Prince, E., Frader, J., Bosk, C.: On hedging in physician-physician discourse. In: Di Pietro, R.J. (ed.) Linguistics and the Professions, pp. 83–97. Ablex, Norwood, NJ (1982)

Rambow, O., Rogati, M., Walker, M.A.: Evaluating a trainable sentence planner for a spoken dialogue travel system. In: Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics (ACL), pp. 434–441. Toulouse (2001)

Reeves, B., Nass, C.: The Media Equation. University of Chicago Press, Chicago (1996)

Rehm, M., André, E.: From annotated multimodal corpora to simulated human-like behaviors. In: Modeling Communications. Lecture Notes in Computer Science, vol. 4930, pp. 1–17 (2008)

Reiter, E., Dale, R.: Building Natural Language Generation Systems. Cambridge University Press, Cambridge (2000)

Reiter, E., Sripada, S.: Human variation and lexical choice. Comput. Linguist. **28**, 545–553 (2002)

Reitter, D., Keller, F., Moore, J.D.: Computational modelling of structural priming in dialogue. In: Proceedings of the Human Language Technology Conference—North American Chapter of the Association for Computational Linguistics (HLT-NAACL), pp. 121–124. New York City, NY (2006)

Revelle, W.: Personality processes. Annu. Rev. Psychol. **46**, 295–328 (1991)

Riloff, E., Wiebe, J., Phillips, W.: Exploiting subjectivity classification to improve information extraction. In: Proceedings of the National Conference On Artificial Intelligence (AAAI), pp. 1106–1111. Pittsburgh, PA (2005)

Rushton, J.P., Murray, H.G., Erdle, S.: Combining trait consistency and learning specificity approaches to personality, with illustrative data on faculty teaching performance. Pers. Indiv. Diff. **8**, 59–66 (1987)

Ruttkay, Z., Dormann, C., Noot, H.: Embodied conversational agents on a common ground. In: Ruttkay, Z., Pelachaud, C. (eds.) From Brows to Trust: Evaluating Embodied Conversational Agents, chap. 2, pp. 27–66. Kluwer Academic Publishers, Norwell, MA (2004)

Scherer, K.R.: Personality markers in speech. In: Scherer, K.R., Giles, H. Social Markers in Speech, pp. 147–209. Cambridge University Press, Cambridge (1979)
Scherer, K.R.: Vocal indicators of stress. In: Darby, J. (ed.) Speech Evaluation in Psychiatry, pp. 171–187. Grune & Stratton, New York (1981)
Schober, M.F., Brennan, S.E.: Processes of interactive spoken discourse: The role of the partner. In: Graesser, A.C., Gernsbacher, M.A., Goldman, S.R. (eds.) Handbook of Discourse Processes, pp. 123–164. Lawrence Erlbaum Associates, Hillsdale, NJ (2003)
Scott, D.R., de Souza, C.S.: Getting the message across in RST-based text generation. In: Dale, R., Mellish, C., Zock, M. (eds.) Current Research in Natural Language Generation, pp. 47–73. Academic Press, NY (1990)
Siegman, A.W.: The telltale voice: Nonverbal messages of verbal communication. In: Feldstein, S., Siegman, A.W. (eds.) Nonverbal Behavior and Communication, chap. 7, pp. 183–243. Lawrence Erlbaum Associates, Hillsdale, NJ (1978)
Siegman, A.W., Pope, B.: Personality variables associated with productivity and verbal fluency in the initial interview. In: Proceedings of the 73rd Annual Conference of the American Psychological Association, Chicago, IL (1965)
Slater, M., Pertaub, D.-P., Barker, C., Clark, D.: An experimental study on fear of public speaking using a virtual environment. CyberPsychol. Behav. **9**(5), 627–633 (2006)
Stenchikova, S., Stent, A.: Measuring adaptation between dialogs. Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue, pp. 166–173 (2007)
Stent, A., Prasad, R., Walker, M.A.: Trainable sentence planning for complex information presentation in spoken dialog systems. In: Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL), pp.79–86. Barcelona (2004)
Tapus, A., Mataric, M.: Socially assistive robots: The link between personality, empathy, physiological signals, and task performance. In: Proceedings of the AAAI Spring Symposium on Emotion, Personality and Social Behavior, Palo Alto, CA (2008)
Thorne, A.: The press of personality: A study of conversations between introverts and extraverts. J. Pers. Soc. Psychol. **53**, 718–726 (1987)
Traum, D., Roque, A., Georgiou, A.L.P., Gerten, J., Narayanan, B. M.S., Robinson, S., Vaswani, A.: Hassan: A virtual human for tactical questioning. In: Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue, pp. 71–74. Antwerp (2007)
Vogel, K., Vogel, S.: L'interlangue et la personalité de l'apprenant. Int. J. Appl. Linguist. **24**(1), 48–68 (1986)
Walker, M.A., Rambow, O.: Spoken language generation. Comput. Speech. Lang. Special Issue Spoken Lang. Gen. **16**(3–4), 273–281 (2002)
Walker, M.A., Cahn, J.E., Whittaker, S.J.: Improvising linguistic style: Social and affective bases for agent personality. In: Proceedings of the 1st Conference on Autonomous Agents, pp. 96–105. Marina Del Rey, CA (1997)
Walker, M.A., Rambow, O., Rogati, M.: Training a sentence planner for spoken dialogue using boosting. Comput Speech Lang **16**(3–4) (2002)
Walker, M.A., Whittaker, S., Stent, A., Maloor, P., Moore, J.D., Johnston, M., Vasireddy, G.: Generation and evaluation of user tailored responses in multimodal dialogue. Cogn. Sci. **28**(5), 811–840 (2004)
Walker, M.A., Stent, A., Mairesse, F., Prasad, R.: Individual and domain adaptation in sentence planning for dialogue. J. Artif. Intell. Res. (JAIR) **30**, 413–456 (2007)
Wang, N., Johnson, W.L., Mayer, R.E., Rizzo, P., Shaw, E., Collins, H.: The politeness effect: Pedagogical agents and learning gains. Front. Artif. Intell. Appl. **125**, 686–693 (2005)
Watson, D., Clark, L.A.: On traits and temperament: General and specific factors of emotional experience and their relation to the five factor model. J. Pers. **60**(2), 441–476 (1992)
Weaver, J.B: Personality and self-perceptions about communication. In: McCroksey, J.C., Daly, J.A., Martin, M.M., Beatty, M.J. (eds.) Communication and Personality: Trait perspectives, chap. 4, pp. 95–118. Hampton Press, NJ (1998)
Wiebe, J.: Recognizing subjective sentences: A computational investigation of narrative text. PhD thesis, State University of New York, Buffalo, NY (1990)
Wilkie, J., Jacka, M.A., Littlewood, P.J.: System-initiated digressive proposals in automated human-computer telephone dialogues: The use of contrasting politeness strategies. Int. J. Hum.-Comput. Stud. **62**, 41–71 (2005)

Wilson, T., Wiebe, J., Hwa, R.: Just how mad are you? Finding strong and weak opinion clauses. In: Proceedings of the 19th National Conference on Artificial Intelligence (AAAI), pp. 761–769. San Jose, CA (2004)

Zukerman, I., Litman, D.J.: Natural language processing and user modeling: Synergies and limitations. User Model. User-Adap. Inter. **11**(1–2), 129–158 (2001)

Author Biographies

François Mairesse received his Master's degree in Computer Science and Engineering from the Catholic University of Leuven, Belgium, and his Ph.D. in Computer Science from the University of Sheffield, UK. Since 2008 he has been a Research Associate in the Dialogue Systems Group of the Cambridge University Machine Intelligence Laboratory, working on statistical methods for optimizing human-machine dialogue. His work has focused on data-driven approaches to natural language understanding, natural language generation, and expressive text-to-speech synthesis. This paper is based on his thesis work on models for controlling the personality conveyed by a conversational agent.

Marilyn A. Walker is Professor of Computer Science at University of California Santa Cruz, and the founder and director of the Natural Language and Dialogue Systems Lab. Dr. Walker received her B.A. degree in Computer and Information Science from the University of California Santa Cruz and her M.S. and Ph.D. degrees in Computer Science from Stanford University and the University of Pennsylvania. Dr. Walker has worked on many aspects of dialogue interaction, both in algorithms for dialogue management and language generation for dialogue systems, as well as computational analysis of human-human dialogue. She conducted the first experiments using reinforcement learning to adapt a dialogue system to human users. Her work on personalization of dialogue systems involves algorithms for user tailoring, individual adaptation of linguistic style using boosting, and generation using theories of politeness and personality. She has authored over a hundred technical papers and is holder of more than 10 patents. She has edited a book on Centering Theory and special issues of journals on empirical methods in discourse and dialogue and on spoken language generation for dialogue systems.