Definite Description Lexical Choice: taking Speaker's Personality into account

Alex Gwo Jen Lan, Ivandré Paraboni
University of São Paulo, School of Arts, Sciences and Humanities
São Paulo, Brazil
alex.lan@usp.br, ivandre@usp.br

Abstract

In Natural Language Generation (NLG), Referring Expression Generation (REG) lexical choice is the subtask that provides words to express a given input meaning representation. Since lexical choices made in real language use tend to vary greatly across speakers, computational models of lexicalisation have long addressed the issue of human variation in the REG field as well. However, studies of this kind will often rely on large collections of pre-recorded linguistic examples produced by every single speaker of interest, and on every domain under consideration, to obtain meaning-to-text mappings from which the lexicalisation model is built. As a result, speaker-dependent lexicalisation may be impractical when suitable annotated corpora are not available. As an alternative to corpus-based approaches of this kind, this paper argues that differences across human speakers may be at least partially influenced by personality, and presents a personality-dependent lexical choice model for REG that is, to the best of our knowledge, the first of its kind. Preliminary results show that our personality-dependent approach outperforms a standard lexicalisation model (i.e., based on meaning-to-text mappings alone), and that the use of personality information may be a viable alternative to strategies that rely on corpus knowledge.

Keywords: Referring Expressions, Lexical choice, Personality, Big Five

1. Introduction

In Natural Language Generation (NLG), *lexical choice* is understood as the task of selecting words to express an input meaning representation. This paper focuses on the particular subtask of definite descriptions lexical choice, that is, the generation step that follows Referring Expression Generation (REG) content selection (Krahmer and van Deemter, 2012) in a traditional NLG architecture (Reiter and Dale, 2000).

The input to the lexicalisation task is a set of meanings (or properties) represented as attribute-value pairs to be expressed in surface form, and the output is a word string. For simplicity, in what follows we shall focus on the choice of words that realise input properties, and we will leave aside issues of linearisation, agreement and others.

Let us consider the goal of producing a lexicalisation for a possible description of the person illustrated in Figure 1, taken from Face Place[1] images (Righi et al., 2012).

Figure 1: An example image from *Face Place*.

In a context of this kind, an underlying REG algorithm may produce a description as in

{<*gender-male*>,<*hair.style-curly*>}.

The task of the lexicalisation model in this case is to assign words to these properties, which in the present example may result in a word string, as in

'the man with curly hair'.

Existing approaches to definite description lexicalisation will often generate a single, fixed surface realisation from the given input. Human descriptions, on the other hand, show much greater linguistic variation, that is, different speakers will often choose different words to express the same meaning. In the previous example, these may include, for instance, 'the guy with wavy hair', 'the boy with frizzy hair' and many others.

The issue of human variation in definite description lexicalisation has been addressed in a few REG studies. In particular, the work in (Hervás et al., 2013) has extensively analysed the lexicalisation of referring expressions from the TUNA corpus (Gatt et al., 2007), and it has provided a number of insights on how human speakers may be grouped together according to their lexical preferences. Corpus-based studies of this kind, however, will usually rely on a large collection of pre-recorded linguistic examples produced by every single speaker of interest, and on every domain of interest. As a result, corpus-based lexicalisation may not always be a viable solution for practical NLG applications.

As an alternative to corpus-based lexicalisation, we notice that differences across speakers may be at least partially influenced by *personality* traits. Personality models such as the well-known Big Five model (Goldberg, 1990) are largely motivated by linguistic choices made by individuals (e.g., an extrovert may use more words than an introvert etc.) and, in particular, by their use of adjectives (which are ubiquitous in definite descriptions). In addition to that, we notice that personality traits are easily obtainable from a number of sources (e.g., inferable from text on social networks as in (Mairesse et al., 2007)), and this may be usually

[1]Stimulus images courtesy of Michael J. Tarr, Center for the Neural Basis of Cognition and Department of Psychology, Carnegie Mellon Univ. Funding provided by NSF award 0339122.

accomplished at a lower cost than collecting a large REG corpus.
Based on these observations, this paper describes a study on personality-dependent definite description lexical choice that is, to the best of our knowledge, the first of its kind. We propose a machine learning approach to lexical choice that takes as an input, in addition to the intended meaning representations, the personality traits of a target speaker. Results show that our personality-dependent approach outperforms standard lexical choice (i.e., based on meaning representations alone), and suggest that the use of personality information may be a viable alternative to strategies that rely on corpus knowledge.
The rest of this paper is organised as follows. Section 2 describes existing work on personality-based NLG and related fields. Section 3 presents our first experiment, concerning the issue of personality recognition from referring expressions. Section 4 proposes the personality-dependent lexicalisation model, and Section 5 describes its evaluation work. Finally, Section 6 presents a number of conclusions and discusses future work.

2. Background

This section briefly discusses the issue of human variation in the lexical choice task for definite description generation, the Big Five personality model, and the corpus to be taken as our test data.

2.1. Lexical choice and human variation

In Referring Expression Generation, once the generation of a description has been decided (Paraboni and van Deemter, 1999; Paraboni and van Deemter, 2002) and its semantic contents have been determined (Dale and Reiter, 1995), the next and final step consists of performing lexical choice to produce actual text. Given an input meaning representation (i.e., the output of a content selection algorithm) consisting of a set of semantic properties represented as attribute-value pairs, as in *gender-male*, the goal of a lexicalisation model is to generate the surface form of a definite description in a target language (e.g., 'the guy').
Although human variation is a popular research topic in REG content selection (Bohnet, 2008; Fabbrizio et al., 2008; Viethen et al., 2013; Ferreira and Paraboni, 2014), there are few studies focused on the issue of lexical choice of definite descriptions. A remarkable exception is the work in (Hervás et al., 2013; Hervás et al., 2015), which presents a corpus-based approach to lexical choice that attempts to mimic descriptions produced by human speakers in the TUNA domain (Gatt et al., 2007). The study compares a standard baseline model and a proposal that takes individual preferences into account, and results show that the proposal leads to a 40% decrease in similarity error against the reference corpus.

2.2. The Big Five personality model

Studies in Psychology and related fields have devoted great attention to the Big Five model (Goldberg, 1990), which contemplates five fundamental dimensions of human personality: Openness to experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. Big Five personality traits are largely motivated by the linguistic choices made by an individual, and may be estimated by a wide range of methods proposed in the Psychology field, the most common being the use of personality inventories. Among these, the need for a quick assessment tool led to the 44-item *BFI* inventory (John et al., 1991), which consists of brief statements containing personality-related adjectives that capture the most essential aspects of each factor in the Big Five model, such as 'Is depressed, blue'. BFI items are answered in a scale ranging from 1 (totally disagree) to 5 (totally agree), and these responses are combined using positive and negative weights to form as a single scalar value representing each of the five dimensions of human personality and additional facets (Soto and John, 2009).

2.3. Personality-based NLG

Given the close relation between personality and natural language, it is not surprising that the use of the Big Five model has been ubiquitous both in natural language understanding (Golbeck et al., 2011; Farnadi et al., 2013; Plank and Hovy, 2015; Najib et al., 2015) and NLG research (Mairesse and Walker, 2010; Marshall et al., 2015; Lukin et al., 2015). In particular, the work in (Mairesse and Walker, 2011) has addressed a wide range of generation decisions that may be driven by a target personality profile. The work focuses on practical, end-to-end language generation by presenting a configurable NLG system to generate restaurant textual recommendations. The system - called PERSONAGE - is trained on personality-annotated data, and the generated text is shown to be recognisable by human judges as reflecting certain well-defined personality traits.
Lexicalisation in PERSONAGE is performed for each content word in the text (and not only for the realisation of definite descriptions) through three parameters: lexicon frequency, lexicon word length and verb strength (e.g., 'suggest' versus 'recommend'). These parameters make use of knowledge obtained from several online lexical resources (e.g., WordNet and VERBOCEAN), and from corpus frequency counts.

2.4. The b5-ref corpus

As a means to investigate the relation between personality and the lexical choice in referring expressions, we make use of the *b5-ref* corpus of definite descriptions annotated with personality information about the individuals who produced them (Paraboni et al., 2017). The corpus is part of a larger dataset of text and accompanying personality information, the *b5* corpus (Ramos et al., 2018).
The *b5-ref* corpus contains descriptions of human photographs elicited from a set of 12 visual contexts built from Face Place (Righi et al., 2012) and further annotated with their semantic properties. This procedure is similar to standard data collection tasks intended to build referring expression corpora (Gatt et al., 2007; Dale and Viethen, 2009).
The choice for the Face Place domain was motivated by the observation that these images are annotated with affective information (e.g., sad, angry etc.), which may arguably help

to make more explicit the possible (personality) differences across speakers. An example of stimulus context from the *b5-ref* corpus is illustrated in Figure 2.

Figure 2: Example of stimulus image built from *Face Place* for the b5-ref corpus.

Based on situations of reference of this kind, subjects were instructed to complete a sentence in the form 'The person outlined in red is the ... ', which requires a uniquely identifying description of the target object. In this example, uniqueness could be achieved, for instance, by making use of expressions such as 'the guy with curly hair', ' the only man in the scene', etc.
The *b5-ref* corpus contains 1810 descriptions produced by 152 native speakers of Brazilian Portuguese who responded a 44-item BFI personality inventory (John et al., 2008) for this language (de Andrade, 2008). The descriptions were subsequently annotated with the 27 most frequent semantic properties observed in the corpus. Each property is represented as an attribute-value pair as in *hair.style-curly*. The collected descriptions are represented, on average, by four annotated properties each. Further details are provided in (Paraboni et al., 2017). Table 1 summarises the attribute frequencies in this domain.

Attribute	Possible values	Instances	%
gender	{male,female}	1707	23.7%
race	{asian,black,cauc.}	794	11.0%
smile	{yes,no}	784	10.9%
isYoung	{yes}	705	9.8%
hair.colour	{dark,blonde}	633	8.8%
hair.length	{short,long }	434	6.0%
emotion	{pos.,neg.,neutral}	266	3.7%
eye.colour	{light,dark}	191	2.7%
ponytail	{yes,no}	174	2.4%
eyebrows	{other}	156	2.2%
skin	{fair,dark}	150	2.1%
hair.style	{straight,curly}	134	1.9%
nose	{other}	115	1.6%
face	{other}	109	1.5%
facial.hair	{yes}	109	1.5%
lips	{other}	97	1.3%
spots	{yes}	96	1.3%
eyes	{other}	67	0.9%
hair	{other}	66	0.9%
mouth	{shut, open}	64	0.9%
narrow.eyed	{yes,no}	55	0.8%
shape	{other}	55	0.8%
glasses	{yes,no}	52	0.7%
earring	{yes,no}	50	0.7%
fringe	{yes}	49	0.7%
unkempt	{yes}	41	0.6%
ears	{other}	37	0.5%

Table 1: Annotation scheme for the *b5-ref* corpus and attribute frequencies.

3. Pilot study: Personality recognition from input meaning representations

Before addressing our personality-dependent lexical choice model in the next section, we first carried out an analysis to investigate the relation between Big Five personality traits and the annotated meaning representations that we intend to use as the input to our lexical choice model. To this end, an experiment on personality recognition from referential attribute sets was developed. The goal of this experiment was to illustrate to which extent *b5-ref* referring expressions - and, in particular, the underlying annotation scheme - would reflect personality differences across speakers.
The present Big Five recognition task is in principle analogous to personality recognition from text sources such as social networks (Golbeck et al., 2011), blogs (Iacobelli et al., 2011) essays (Mairesse et al., 2007) and others, except that our input consists of sets of semantic properties representing referring expressions. Although this may not have an obvious, real-world application *per se*, learning personality traits from referential attribute sets may provide indirect evidence that our personality-dependent approach (to be discussed in the next section) is feasible. We notice also that, in a related study, a subset of the actual *b5-ref* word strings (as opposed to their semantic annotation considered in the present case) was applied to a number of the personality recognition tasks. Details are provided in (dos Santos et al., 2017).

3.1. Computational models

Personality recognition is presently modelled as a binary classification task to determine whether an individual shows positive or negative tendency towards each trait. To this end, we assign positive/negative class labels based on the average score for each trait, that is: positive instances of the class representing a personality trait t consist of the individuals with an equal or above-average score for the trait t, and negative instances correspond to those individuals with below-average scores for t. As a result, 1656 instances were produced for each of the five personality traits. The distribution of positive and negative instances for each class is illustrated in Table 2.

Trait	positive	negative
Extraversion	828	828
Agreeableness	767	889
Conscientiousness	875	781
Neuroticism	829	827
Openness	802	854

Table 2: Learning instances distribution

As learning features, we consider a set of binary values representing the seven most frequent attributes in the corpus (cf. previous section.) Each value indicates whether that particular attribute appeared in a referring expression or not. Moreover, since the choice of referential attributes may vary considerably across stimuli (e.g., in a scene in which nobody is smiling, the use of the *smile* attribute is far less common than in scenes in which one or more characters are smiling), the learning features also include a context identifier value.

3.2. Results

We performed five independent rule-based decision table classification tasks (Kohavi, 1995) with 10-fold cross-validation over the entire dataset. Precision, Recall and F1-measure results are summarised in Table 3.

	positive class			negative class		
Trait	P	R	F1	P	R	F1
Extraversion	0.55	0.51	0.53	0.55	0.60	0.57
Agreeableness	0.58	0.61	0.59	0.59	0.56	0.57
Conscientiousness	0.57	0.58	0.58	0.57	0.57	0.57
Neuroticism	0.53	0.49	0.51	0.52	0.56	0.54
Openness	0.54	0.52	0.53	0.54	0.55	0.54

Table 3: Personality recognition from attribute sets.

3.3. Discussion

Overall best results were observed for Agreeableness and Conscientiousness, and for the negative instances of Extraversion. On the other hand, results for Neuroticism were considerably lower, suggesting that this particular trait may have less influence over referential attribute choice in the present domain.

4. Lexical choice model

In what follows we will focus on the task of providing lexical choices for definite descriptions alone. This can be viewed as the final generation step that takes the output of a standard REG algorithm (e.g., (Dale and Reiter, 1995)) as its input, and then generates text in a target language - in the present case, Brazilian Portuguese.

Our lexical choice model takes as an input the context id (representing a visual scene in the *b5-ref* corpus) and a concept - hereby represented as a semantic property p - to produce the most likely wording w of p.

Using the entire corpus as test data, all properties with five or more references were mapped onto their lexical forms through manual annotation. For instance, *gender-male* was found to be lexicalised as ‘man’, ‘boy’, ‘guy’ and so on. Given our goal of learning alternative lexicalisations for a given input property, the trivial cases represented by properties with a single possible lexicalisation in the corpus were disregarded.

As a result of the annotation task, a set of 4,345 property-word mappings was obtained. The number of alternative wordings per property ranged from 2 to 9, with an average 4.6 wordings each.

Using the property-word mappings, lexical choice was modelled as a multi-class learning task. The goal of the model is to predict the wording of a given property based on its referential context and on the personality traits of the target speaker.

As learning features, we considered the input property p to be lexicalised, the context id in which p occurs, and five features representing the Big Five personality scores of the speaker as scalar values. The inclusion of the context identifier id is intended to reflect the practical observation that a concept may not have exactly the same meaning (and therefore not necessarily the same wording) in different contexts. For instance, *gender-male* may be lexicalised as ‘boy’ in a scene showing a child, and as ‘young man’ in a scene showing a slightly more mature individual.

5. Evaluation

In this section we discuss the evaluation of our personality-dependent lexical choice model. The model is compared against a baseline alternative in which the five personality-related features are omitted. This, in practice, amounts to a baseline method that chooses the most frequent wording for each input property. The goal of this evaluation is to show that the use of personality information leads to more accurate lexical choices than the baseline method.

Both models - with and without personality information - were built from the entire set of 4,345 lexicalisations discussed in the previous section using decision-tree induction with 10-fold cross-validation.

5.1. Results

Table 4 presents Precision, Recall and F1-measure results for the baseline and personality-dependent models for properties with 20 or more instances in the corpus. The ‘choice’ column shows the number of possible alternative lexicalisations available for each property in the data, and it is indicative of the complexity of each individual task. For brevity, properties for which both models achieved zero F1 scores (mainly due to data sparsity) are not represented.

5.2. Discussion

From the results in the previous section we notice that taking personality information into account generally increases (and never decreases) lexicalisation performance. This offers support to our main research hypothesis.

Personality-dependent lexical choice does seem to make more accurate decisions for most input properties, including even those with a relatively small number of instances. However, a post hoc analysis suggested that the use of personality information is particularly helpful in the lexicalisation of *affective information* (e.g., properties conveying attributes such as *smile*, *emotion* etc.), and only to a lesser extent in the case of more physical features. Although this outcome may seem in principle intuitive, more work is still require to determine why some concepts seem to be more dependent on personality than others.

6. Final remarks

This work has investigated the role of personality traits in the lexical choice in definite description generation. Based on a corpus of definite descriptions annotated with Big Five

			Baseline			Proposal		
Property	instances	choice	P	R	F1	P	R	F1
gender-male	708	7	0.13	0.25	0.17	0.43	0.31	**0.30**
gender-female	494	6	0.61	1.00	0.76	0.61	1.00	0.76
race-asian	363	6	0.14	0.33	0.20	0.49	0.47	**0.45**
smile-yes	347	4	0.32	0.50	0.39	0.51	0.50	**0.41**
smile-no	266	7	0.36	0.50	0.42	0.65	0.79	**0.70**
hair.col-blonde	238	9	0.57	1.00	0.72	0.57	1.00	0.72
race-black	227	2	0.92	0.94	0.93	0.90	0.96	0.93
hair.len-short	223	5	0.81	0.99	0.89	0.81	0.99	0.89
hair.col-dark	153	4	0.26	0.50	0.34	0.55	0.59	**0.55**
race-caucasian	139	3	0.33	0.50	0.40	0.68	0.73	**0.70**
emotion-neg.	75	4	0.15	0.28	0.19	0.30	0.31	**0.30**
hair.len-long	74	4	0.24	0.48	0.32	0.42	0.48	**0.40**
emotion-pos.	45	2	0.32	0.50	0.39	0.88	0.87	**0.88**
emotion-neut.	20	2	0.23	0.50	0.32	0.33	0.43	**0.37**

Table 4: Personality-dependent lexical choice. Best F1 scores for each class are highlighted.

personality information, we have shown that taking personality information into account increases lexical choice accuracy, an insight that may help the design of more realistic (i.e., human-like) models of Natural Language Generation. As future work, we intend to extend our current model to address the task of surface realisation in general, allowing the generation of full text sentences according to a set of target personality traits.

7. Acknowledgements

This work has been supported by grant # 2016/14223-0, São Paulo Research Foundation (FAPESP).

8. Bibliographical References

Bohnet, B. (2008). The fingerprint of human referring expressions and their surface realization with graph transducers. In *5th International Natural Language Generation Conference*, pages 207–210, Salt Fork, Ohio, USA. Association for Computational Linguistics.

Dale, R. and Reiter, E. (1995). Computational interpretations of the Gricean maxims in the generation of referring expressions. *Cognitive Science*, 19.

Dale, R. and Viethen, J. (2009). Referring expression generation through attribute-based heuristics. In *12th European Workshop on Natural Language Generation*, ENLG '09, pages 58–65, Athens, Greece. Association for Computational Linguistics.

de Andrade, J. M. (2008). *Evidências de validade do inventário dos cinco grandes fatores de personalidade para o Brasil.* Ph.D. thesis, Universidade de Brasília.

dos Santos, V. G., Paraboni, I., and Silva, B. B. C. (2017). Big five personality recognition from multiple text genres. In *Text, Speech and Dialogue (TSD-2017) Lecture Notes in Artificial Intelligence vol. 10415*, pages 29–37, Prague, Czech Republic. Springer-Verlag.

Fabbrizio, G. D., Stent, A., and Bangalore, S. (2008). Trainable speaker-based referring expression generation. In *12th Conference on Computational Natural Language Learning*, pages 151–158, Manchester, UK. Association for Computational Linguistics.

Farnadi, G., Zoghbi, S., Moens, M.-F., and de Cock, M. (2013). Recognising personality traits using Facebook status updates. In *Proceedings of WCPR13 in conjunction with ICWSM-13*, Boston, USA. The AAAI Press.

Ferreira, T. C. and Paraboni, I. (2014). Classification-based referring expression generation. In *Computational Linguistics and Intelligent Text Processing, Lecture Notes in Computer Science 8403*, pages 481–491, Kathmandu, Nepal. Springer.

Gatt, A., van der Sluis, I., and van Deemter, K. (2007). Evaluating algorithms for the generation of referring expressions using a balanced corpus. In *Proceedings of ENLG-07*, Schloss Dagstuhl, Germany. Association for Computational Linguistics.

Golbeck, J., Robles, C., and Turner, K. (2011). Predicting personality with social media. In *CHI Í1 Extended Abstracts on Human Factors in Computing Systems*, CHI EA Í1, pages 253–262, Vancouver, BC, Canada. ACM.

Goldberg, L. R. (1990). An alternative description of personality: The Big-Five factor structure. *Journal of Personality and Social Psychology*, 59:1216–1229.

Hervás, R., Francisco, V., and Gervás, P. (2013). Assessing the influence of personal preferences on the choice of vocabulary for natural language generation. *Information Processing & Management*, 49(4):817–832.

Hervás, R., Arroyo, J., Francisco, V., Peinado, F., and Gervás, P. (2015). Influence of personal choices on lexical variability in referring expressions. *Natural Language Engineering*, inpress:1–34, 12.

Iacobelli, F., Gill, A. J., Nowson, S., and Oberlander, J. (2011). Large scale personality classification of bloggers. In Sidney K. D'Mello, et al., editors, *ACII (2)*, volume 6975 of *Lecture Notes in Computer Science*, pages 568–577, Memphis, TN, USA. Springer.

John, O. P., Donahue, E., and Kentle, R. (1991). The Big Five inventory - versions 4a and 54. Technical report, Inst. Personality Social Research, University of California, Berkeley, CA, USA.

John, O. P., Naumann, L. P., and Soto, C. J., (2008).

Paradigm Shift to the Integrative Big-Five Trait Taxonomy: History, Measurement, and Conceptual Issues, pages 114–158. Guilford Press, New York, NY.

Kohavi, R. (1995). The power of decision tables. In *8th European Conference on Machine Learning*, pages 174–189.

Krahmer, E. and van Deemter, K. (2012). Computational generation of referring expressions: A survey. *Computational Linguistics*, 38(1):173–218.

Lukin, S. M., Reed, L., and Walker, M. A. (2015). Generating sentence planning variations for story telling. In *16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 188–197, Prague, Czech Republic. Association for Computational Linguistics.

Mairesse, F. and Walker, M. A. (2010). Towards personality-based user adaptation: psychologically informed stylistic language generation. *User Model. User-Adapt. Interaction*, 20(3):227–278.

Mairesse, F. and Walker, M. A. (2011). Controlling user perceptions of linguistic style: Trainable generation of personality traits. *Computational Linguistics*, 37(3):455–488.

Mairesse, F., Walker, M., Mehl, M., and Moore, R. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research (JAIR)*, 30:457–500.

Marshall, T. C., Lefringhausen, K., and Ferenczi, N. (2015). The Big Five, self-esteem, and narcissism as predictors of the topics people write about in Facebook status updates. *Personality and Individual Differences*, 85:35–40.

Najib, F., Cheema, W. A., and AdeelNawab, R. M. (2015). Author's traits prediction on Twitter data using content based approach. In Linda Cappellato, et al., editors, *CLEF 2015 Working Notes*, volume 1391 of *CEUR Workshop Proceedings*, Toulouse, France. CEUR-WS.org.

Paraboni, I. and van Deemter, K. (1999). Issues for the generation of document deixis. In *Procs. of workshop on Deixis, Demonstration and Deictic Belief in Multimedia Contexts, in association with the 11th European Summers School in Logic, Language and Information (esslli99)*, pages 44–48.

Paraboni, I. and van Deemter, K. (2002). Towards the generation of document-deictic references. In *Information sharing: reference and presupposition in language generation and interpretation*, pages 329–352. CSLI Publications.

Paraboni, I., Monteiro, D. S., and Lan, A. G. J. (2017). Personality-dependent referring expression generation. In *Text, Speech and Dialogue (TSD-2017) Lecture Notes in Artificial Intelligence vol. 10415*, pages 20–28, Prague, Czech Republic. Springer-Verlag.

Plank, B. and Hovy, D. (2015). Personality traits on Twitter - or - how to get 1,500 personality tests in a week. In *Proc. of WASSA-2015*, pages 92–98.

Ramos, R. M. S., Neto, G. B. S., Silva, B. B. C., Monteiro, D. S., Paraboni, I., and Dias, R. F. S. (2018). Building a corpus for personality-dependent natural language understanding and generation. In *11th International Conference on Language Resources and Evaluation (LREC-2018) (to appear)*, Miyasaki, Japan. ELRA.

Reiter, E. and Dale, R. (2000). *Building natural language generation systems*. Cambridge University Press, New York, NY, USA.

Righi, G., Peissig, J. J., and Tarr, M. J. (2012). Recognizing disguised faces. *Visual Cognition*, 20(2):143–169.

Soto, C. J. and John, O. P. (2009). Ten facet scales for the Big Five Inventory: Convergence with NEO PI-R facets, self-peer agreement, and discriminant validity. *Journal of Research in Personality*, 43(1):84–90.

Viethen, J., Mitchell, M., and Krahmer, E. (2013). Graphs and spatial relations in the generation of referring expressions. In *14th European Workshop on Natural Language Generation*, pages 72–81, Sofia, Bulgaria. Association for Computational Linguistics.