

UC Santa Cruz

UC Santa Cruz Electronic Theses and Dissertations

Title

Linguistic and Gestural Adaptation

Permalink

<https://escholarship.org/uc/item/0vb2611x>

Author

Hu, Zhichao

Publication Date

2018

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

LINGUISTIC AND GESTURAL ADAPTATION

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

COMPUTER SCIENCE

by

Zhichao Hu

June 2018

The Dissertation of Zhichao Hu
is approved:

Marilyn Walker, Chair

Jean Fox Tree

Michael Neff

Tyrus Miller
Vice Provost and Dean of Graduate Studies

Copyright © by

Zhichao Hu

2018

Table of Contents

List of Figures	v
List of Tables	vii
Abstract	ix
Dedication	xi
Acknowledgments	xii
1 Introduction	1
1.1 Motivation	7
1.2 Linguistic Adaptation	11
1.3 Gestural Adaptation	13
1.4 Contributions	14
2 Related Work	16
2.1 Partner Adaptation and Theories of Adaptation	16
2.2 Measuring Lexical Adaptation	19
2.3 Adaptation in Natural Language Generation	22
2.4 Theories and Studies of Personality and Gesture	24
2.5 Gesture Adaptation and Generation	27
3 Datasets	31
3.1 ArtWalk Corpus (AWC)	31
3.2 Walking Around Corpus (WAC)	33
3.3 Map Task Corpus (MPT)	33
3.4 Switchboard Corpus (SWBD)	34
3.5 Story Dialog with Gestures Corpus (SDG)	35
3.5.1 Personal Narrative Monologic Corpus	38
3.5.2 Dialog Annotations	41
3.5.3 Gesture Annotations	43

4	Linguistic Adaptation	49
4.1	Implementing and Evaluation Linguistic Adaptation	50
4.1.1	Personage-primed	50
4.1.2	Methodology	58
4.1.3	Experiments and Results	60
4.1.4	Discussion	64
4.2	Measuring Linguistic Adaptation in Dialogs	65
4.2.1	Method and Overview	66
4.2.2	Experimental Setup	70
4.2.3	Experiments on Modeling Adaptation	71
4.2.4	Discussion	79
5	Gestural Adaptation	85
5.1	Stimulus Construction	86
5.2	Experiment Method	89
5.2.1	Experiment 1: Personality Variation.	89
5.2.2	Experiment 2: Gestural Adaptation.	90
5.3	Experimental Results	92
5.3.1	Personality Results	92
5.3.2	Adaption Results	93
5.4	Discussion	95
6	Conclusion and Future Work	97
6.1	Linguistic Adaptation	98
6.2	Gestural Adaptation	100
6.3	Future Work	100

List of Figures

3.1	Sample dialog Excerpt from the ArtWalk Corpus.	32
3.2	Sample dialog excerpt from the Walking Around Corpus.	34
3.3	Dialog excerpt from the Map Task Corpus.	35
3.4	Dialog excerpt from the Switchboard Dialog Act Corpus.	36
3.5	An example story on the topic of Pet.	37
3.6	Manually constructed dialog from the pet story in Figure 3.5.	38
3.7	Overview of the SDG corpus.	39
3.8	An example story on the topic of Protest.	40
3.9	Part of Scheherazade annotations for the protest story in Figure 3.8. . . .	41
3.10	Manually constructed dialog from the protest story in Figure 3.8. . . .	42
3.11	A subset of the 271 gestures in our gesture library that can be used in annotation and produced in the animations.	44
3.12	Prep, stroke, hold and retract phases of gesture “Cup_Horizontal”. . . .	45
3.13	Pet dialog with gesture annotations. Pictures show the first 6 gestures in the dialog.	47
3.14	Protest dialog with gesture annotations. Pictures show the first 6 ges- tures in the dialog.	48
4.1	The architecture of Personage-primed.	51
4.2	Instructions and statements supported in Personage-primed.	52
4.3	Sample Discourse Model Representation.	53
4.4	Example text plan tree.	54

4.5	DSyntS for the instruction <i>turn-DIR-onto-STREET</i> . Relation I: the component is the subject of the parent; relation II: the component is the direct object of the parent; relation ATTR: the component is a modifier(adjective/prepositional phrase) of the parent.	55
4.6	An example question from Experiment 2.	59
4.7	Regression models.	61
4.8	Decision tree model for Experiment 2 naturalness.	63
4.9	Decision tree model for Experiment 2 friendliness.	64
4.10	Plots of average DAS on different window sizes (1 to 5) for original dialogs vs. randomized dialogs, using all feature sets except Personality LIWC.	79
4.11	Plots of average DAS as the dialogs progress, using all LIWC features vs. extraversion LIWC features.	84
5.1	A snapshot of the experimental stimuli.	86
5.2	Part of DANCE gesture sequence scripts for speaker A in the protest dialog in Fig 3.14.	88
5.3	Virtual agent with different gesture expanse and height for the same gesture.	90

List of Tables

1.1	An example dialog exchange with linguistic and gestural adaptation.	2
1.2	Linguistic adaptation example: no adaptation vs. moderate adaptation vs. too much adaptation.	3
1.3	Linguistic adaptation example from ArtWalk Corpus: no adaptation vs. moderate adaptation (human response) vs. too much adaptation.	4
1.4	Gestural adaptation example: no adaptation vs. adapting only on gestural forms vs. adapting on gestural forms and parameters.	5
2.1	The gestural correlates of extraversion.	25
2.2	The gestural correlates of neuroticism.	25
2.3	The gestural correlates of agreeableness.	26
3.1	Distribution of story topics in the SDG corpus.	40
4.1	Possible values for features used in decision tree model.	62
4.2	Example DAS vector learned from the ArtWalk Corpus.	67
4.3	Number of LIWC features for each personality trait and example features.	72
4.4	Number of dialogs in four corpora, and average DAS scores of different feature sets for original and randomized dialogs. Bold numbers indicate statistically significant differences ($p < 0.0001$) between DAS scores for original and randomized dialogs in paired t -tests.	73
4.5	Average DAS scores for each feature set.	74

5.1	Experiment results: participant evaluated extraversion scores (range from 1 - 7, with 1 being the most introverted and 7 being the most extraverted).	92
5.2	Experiment results: number and percentage of subjects who preferred the adapted (A) stimulus and the non-adapted (NA) stimulus. The letters in the story version refer to dialog turns by speaker A or B. For example, ABA means A takes dialog turns 1 and 3 in the stimuli, while B takes dialog turn 2.	93
5.3	Answers to the second survey question (“why” question) classified into categories. Note that one subject could belong to none or multiple categories, so the percentages for each line don’t add up tp 100%.	94

Abstract

Linguistic and Gestural Adaptation

by

Zhichao Hu

Human conversants in dialog adjust their behavior to their conversational partner in many ways. In terms of language, they adapt to their partners both lexically and syntactically, by using the same referring expressions or sentence structure. In terms of gesture, they mimic their partners' gestural form, frequency, expanse and speed. However, adaptation is not about simply copying dialog partners' words, syntax, and gestures. The process of adaptation to the partner takes place under other special constraints, e.g. providing coherent and informative turns in conversation, expressing one's own personality, or achieving other social and interpersonal goals. How do speakers adapt to one another and at the same time achieve their own conversational goals?

In this thesis, we first carry out an exploratory study to show that adapting to different linguistic features results in different style perceptions: adapting to hedges increase perceptions of friendliness, while adapting to syntactic structures increases perceptions of naturalness. On the basis of these results, we propose an adaptation measure that allows us to capture and model adaptation behaviors that may orient to different levels or types of linguistic representations, such as lexical, syntactical, or stylistic variations in the way speakers talk. We build and compare linguistic adaptation models using our measure with four human dialog corpora, and with different feature sets to represent different levels of linguistic representations.

We also explore gestural adaptation on both particular gesture forms and gesture style. We set up our experiment in the form of two virtual gents co-telling a story. We first verify that we can express the personality of virtual agents through varying gesture

parameters such as speed, height, and expanse. We then show that human subjects prefer adaptive to nonadaptive virtual agents, where the adaptive virtual agent adapts the gesture form and personality of their dialog partner.

Dedicated to my loving parents, 石峰 and 胡贤华; my sister, 胡敏; and my husband,
Yan Li.

Acknowledgments

I would like to express my special appreciation and thanks to Professor Marilyn Walker, my dissertation advisor and academic mother, for her guidance and contribution to this work, and for her friendship during my study at the University of California, Santa Cruz. She is not only my mentor in computer science, but also my mentor in life.

I am deeply indebted to Professor Jean Fox Tree, Professor Michael Neff, and Professor Sri Kurniawan for their guidance, support, and encouragement. I could not have finished my research and dissertation without their help. I wish to thank Yali Mu, Yi Yang, Shenshen Liang, Junce Zhang, Kerui Huang and Wei Huang, who help me through the first few quarters when I first set foot upon this land miles away from where I grew up. I also wish to thank every student in the Natural Language and Dialog Systems Lab: Stephanie Lukin, Elahe Rahimtoroghi, Jiaqi Wu, Lena Reed, Amita Misra, Shereen Oraby, Kevin Bowden, Geetanjali Rakshit, Brian Schwarzmann, Vrindavan Harrison, Jurik Juraska, Grace Lin, Rob Abbott, Gabrielle Halberg, Julia Kelly, Jennifer Sawyer, Sam Wing, Elena Rishes, Larissa Munishkina, Reid Swanson, Carolyn Jimenez, Michelle Dick and Chung-Ning Chang. We had a great time together, you guys rock!

My parents, sister, friends, and my other family are a very important part of my life, and through their love and support have contributed significantly to this work. My father, Xian Hua Hu, sparked my curiosity in science and led me to this fantastic world of computer science. My mother, Feng Shi, has always been supportive of my decisions to pursue whatever dreams I have. My sister, Min Hu, helped me grow and be independent. My husband, Yan Li, provided support and encouragement through the many years we have been together. I am very fortunate to have him.

Several members of the Natural Language and Dialog Systems Lab have contributed to this work. This dissertation includes material of the following previously published

papers. It is my pleasure to acknowledge my coauthors of these papers.

- Zhichao Hu, Gabrielle Halberg, Carolynn R. Jimenez, and Marilyn A. Walker. “Entrainment in Pedestrian Direction Giving: How many kinds of entrainment?” *Workshop on Spoken Dialog Systems (IWSDS 2014)*, Napa, CA, USA, Jan. 2014. With permissions from Gabrielle Halberg and Carolynn R. Jimenez. Marilyn Walker directed and supervised the research which formed the basis for this paper. Gabrielle Halberg and Marilyn Walker have done the ideation, design, and implementation of the paper. I have done the evaluation, and drafting of the paper with Carolynn’s help on user studies.
- Zhichao Hu, Marilyn A. Walker, Michael Neff, and Jean E. Fox Tree. “Storytelling Agents with Personality and Adaptivity,” *Intelligent Virtual Agents (IVA 2015)*, Delft, Netherlands, Aug. 2015. Marilyn Walker, Jean Fox Tree, and Michael Neff directed and supervised the research which formed the basis for this paper. I have done the ideation, design, implementation, evaluation, and drafting of the paper.
- Zhichao Hu, Michelle Dick, Chung-Ning Chang, Michael Neff, Jean E. Fox Tree, and Marilyn A. Walker. “A Corpus of Gesture-Annotated Dialogues for Monologue-to-Dialogue Generation from Personal Narratives,” *Language Resources and Evaluation Conference (LREC 2016)*, Portorož, Slovenia, May 2016. With permissions from Michelle Dick and Chung-Ning Chang. Marilyn Walker, Jean Fox Tree, and Michael Neff directed and supervised the research which formed the basis for this paper. Michelle Dick and Chung-Ning Chang have done the annotations of the corpus. I have done the ideation, design, implementation, evaluation, and drafting of the paper.
- Zhichao Hu, Jean E. Fox Tree, and Marilyn A. Walker. “Modeling Linguistic

and Personality Adaptation for Natural Language Generation,” *The 19th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL 2018)*, Melbourne, Australia, July 2018. Marilyn Walker and Jean Fox Tree directed and supervised the research which formed the basis for this paper. I have done the ideation, design, implementation, evaluation, and drafting of the paper.

This dissertation was completed with the aid of the National Science Foundation under awards NSF CISE RI EAGER #IIS-1044693, NSF CISE CreativeIT #IIS-1002921, NSF CHS #IIS-1115742, of Nuance Foundation Grant SC-14-74, and of auxiliary REU supplements.

Chapter 1

Introduction

With the rapid development of artificial intelligence research, Intelligent Virtual Agents (IVA) are making their ways into everyday life. Automatic chatbots in customer service center, virtual personal assistants in smartphones and smart speakers, and embodied virtual agent companions for children and the elderly are all IVAs, and they help us resolve issues, manage our everyday lives, and keep us company. Pew research center reported 46% of Americans used digital voice assistants on smartphones, tablets, or other stand-alone devices in 2017 [147]. NPR and Edison Research confirmed that one-in-six Americans (16%) owned at least one voice-activated smart speaker, up 128% from January 2017, in early 2018 [139].

However, intelligent virtual agents are far from perfect, and one of their most glaring weaknesses is the lack of verbal and non-verbal adaption, because their responses are mostly generated from a hand-crafted set of scripts. For example, when asked “set a timer for half an hour”, most smart home agents will respond “setting a timer for 30 minutes”, using a new phrase “30 minutes” to refer to the timespan, instead of using the same phrase as the user (“ half an hour”). Human conversation is not a matter of following scripts, but a collaborative interaction where speakers adapt to each other’s behaviors in order to reach a common understanding and, eventually, achieve the goal

of the conversation. Even today’s most advanced intelligent virtual agents lack such kind of adaptation, and this is the problem we aim to solve in this thesis.

This thesis tackles both linguistic adaption and gestural adaption. An example dialog exchange demonstrating both linguistic and gestural adaptation is shown in Table 1.1. This dialog exchange is taken from the Story Dialog with Gestures Corpus (see Section 3.5). Speaker B adapts to A linguistically by reusing the same referring expression “teenagers” and adjective phrase “a bit bigger ”. Speaker B also adapts to A gesturally by mimicking specific gesture forms of A’s (“Cup_Horizontal” and “ShyCalmShake”), as well as performing gestures using the same gestural parameters of A’s (high gesture expanse, height, outwardness, speed and scale).

Speaker	Utterance	Adapted Features
A	<i>Well, (Cup_Down_alt) the no-kill shelter also had what they called (Cup_Horizontal) “teenagers”, which were cats around four to six months old... a bit bigger than the (ShyCalmShake) little kitties.</i>	Linguistic: teenagers, a bit bigger than Gesture Forms: Cup_Horizontal, ShyCalmShake
B	<i>Oh yeah, I saw those (Cup_Horizontal) “teenagers”. They (HandToChest_Vibrate) weren’t exactly adults, but they were a bit (ShyCalmShake) bigger than the little kittens.</i>	Gesture parameters: high gesture expanse, height, outwardness, speed and scale

Table 1.1: An example dialog exchange with linguistic and gestural adaptation.

However, linguistic adaption is not about simply copying dialog partners’ words, because too much mimicking could incur negative user interaction experience. For example, Table 1.2 shows a prime utterance and three target utterances. The prime utterance is from the first speaker and contains linguistic features, which the target (response) from the other speaker can adapt to. In this example, prime contains linguistic features such as discourse markers, referring expressions, syntax structures, and so on. Target Type 1 and 2 show sample responses with too little or too much adaptation. Target Type 3 is the actual response to the prime utterance from the corpus with the right

amount of adaptation (to referring expression “teenager” and adjective phrase “a bit bigger”). The utterance conveys new information (that the speaker saw those cats too) and adapt to the other speaker (by repeating some phrases in prime) naturally. While in Target Type 1, very little adaptation makes the utterance rigid and unnatural: the short response to prime’s detailed description of the cats gives an impression that the second speaker is not engaged in the dialog. In Target Type 2, too much adaptation (to all referring expressions in prime) makes the utterance repetitive and verbose.

Prime utterance: *Well, the no-kill shelter also had what they called “teenagers”, which were cats around four to six months old...a bit bigger than the little kitties.*

Linguistic Features:

Discourse markers: well

Referring expressions: no-kill shelter, teenagers, cats around four to six months old, ...

Syntactic structures: NP->DT+JJ+NN, VP->ADJP+PP ...

Adjective phrases: a bit bigger

Target Type	Target Utterance	Adapted Linguistic Features
1. Little adaptation	<i>I saw them too.</i>	None
2. Too much adaptation	<i>Well, I saw what the no-kill shelter called “teenagers”, cats around four to six months old. They weren’t exactly adults, but they were a bit bigger than the little kittens.</i>	well, no-kill shelter, teenagers, cats around four to six months old, a bit bigger
3. Moderate adaptation	<i>Oh yeah, I saw those “teenagers”. They weren’t exactly adults, but they were a bit bigger than the little kittens.</i>	teenagers, a bit bigger

Table 1.2: Linguistic adaptation example: no adaptation vs. moderate adaptation vs. too much adaptation.

Table 1.3 shows another example taken from human dialogs in the ArtWalk Corpus (see Section 3.1). Target Type 3 is the actual human response to the prime utterance from the corpus. We can see that that Target Type 3 has the right amount of adaptation (to discourse marker “okay”) and the utterance is fluent and natural. While in Target Type 1, very little adaptation makes the utterance stiff, and in Target Type 2, too much

adaptation (to all four discourse markers in prime) makes the utterance unnatural.

Prime utterance: *okay alright so yeah I'm looking at 123 Locust right now*

Linguistic Features:

Discourse markers: okay, alright, so, yeah

Referring expressions: 123 Locust

Syntactic structures: VP->VBP+VP, VP->VBG+PP+ADVB ...

Target Type	Target Utterance	Adapted Linguistic Features
1. Little adaptation	<i>it should be somewhere</i>	None
2. Too much adaptation	<i>okay alright so yeah it should be somewhere</i>	okay, alright, so, yeah
3. Human response	<i>okay I mean it should be somewhere</i>	okay

Table 1.3: Linguistic adaptation example from ArtWalk Corpus: no adaptation vs. moderate adaptation (human response) vs. too much adaptation.

Gestural adaptation is even less explored in intelligent virtual agents. The most primitive method of gestural adaptation is copying the dialog partner's gesture forms. However, mimicking dialog partner's gesture forms is only part of the story. A more natural method of gestural adaptation is to determine gesture features, such as rate, speed, scale, and expanse, according to dialog partner's style such as personality. For instance, Table 1.4 shows a prime utterance by an extraverted speaker taken from the Story Dialogs with Gesture Corpus (see Section 3.5). The utterance has high gesture rate (3 gestures performed: Cup_Down_alt, Cup_Horizontal, ShyCalmShake) with extraversion gesture parameters: high values for gesture expanse, height, outwardness, speed, and scale. A target utterance with no adaptation results in Target Type 1, which portrays an introverted agent that uses different gesture forms when referring to the same entities ("teenagers" and "little kittens") with small, low, inward, and slow gestures. A response like Target Type 1 lacks the sense of interaction. Adapting to only gesture forms (Target Type 2) still results in personality mismatch: an outgoing context with a reserved response. Adapting to both gesture form and parameters creates a most suitable response.

Prime Utterance:

Well, (Cup_Down_alt) the no-kill shelter also had what they called (Cup_Horizontal) “teenagers”, which were cats around four to six months old... a bit bigger than the (ShyCalmShake) little kitties.

Gestural Parameters:

Extraversion gestural parameters: high gesture expanse, height, outwardness, speed, and scale

Target Type	Utterance	Gestural Parameters
1. No adaptation	<i>Oh yeah, I saw those (Side-Out_vibrate) “teenagers”. They weren’t exactly adults, but they were a bit (SideOut1) bigger than the little kittens.</i>	Introversion parameters: low gesture expanse, height, outwardness, speed and scale
2. Adapting only to gesture forms	<i>Oh yeah, I saw those (Cup_Horizontal) “teenagers”. They weren’t exactly adults, but they were a bit (ShyCalmShake) bigger than the little kittens.</i>	Introversion parameters: low gesture expanse, height, outwardness, speed and scale
3. Adapting to gesture forms and parameters	<i>Oh yeah, I saw those (Cup_Horizontal) “teenagers”. They (HandToChest_Vibrate) weren’t exactly adults, but they were a bit (ShyCalmShake) bigger than the little kittens.</i>	Extraversion parameters: high gesture expanse, height, outwardness, speed and scale

Table 1.4: Gestural adaptation example: no adaptation vs. adapting only on gestural forms vs. adapting on gestural forms and parameters.

In this thesis, we propose a vector-based adaptation framework that controls adaptation to linguistic and gestural features in both verbal and non-verbal interaction. We aim to build adaptation models in forms of vectors, which can be used to produce the right amount of adaptation in natural language generation and gesture generation. We believe the amount of linguistic adaptation differs across feature sets, such as referring expressions, discourse markers and syntactical structures. To produce a natural adaptation behavior, a language generator needs to satisfy adaptation constraints from different feature sets. Thus, we aim to learn models of adaptation from human dialogs and store adaptation model of each feature set in a vector. The learned vector of adap-

tation model can be used to control adaptation behavior in natural language generation (NLG). We discuss ways of applying learned adaptation models in various NLG architectures. However, the application of such models in NLG is outside the scope of the thesis.

We first set up experiments to explore various linguistic feature sets that can be considered for adaptation, as well as understand the effect of adapting to different linguistic features. Adapting to different linguistic features results in different style perceptions. In an exploratory study, we show that adapting to hedges increases perceptions of friendliness, while adapting to syntactic structure increases perceptions of naturalness. On the basis of those results, we propose an adaptation measure that aims to reflect different adaptation models of feature sets that describe certain linguistic styles, such as personality traits. We then build and compare adaptation models using four human dialog corpora. Our learned models can be integrated into various natural language generation architectures, such as overgenerate and rank, statistical parameterized methods, and natural language generation using neural networks.

As shown in Table 1.4, gestural adaptation should include not only copying of particular gesture forms, but also mimicry of personality, through the expression of gestural parameters, such as gesture rate, expanse, height, outwardness, speed and scale. Following our adaptation framework, we aim to control gestural adaptation through a vector of gestural parameters. However, unlike linguistic adaptation, dialog corpora annotated with co-speech gestures are not widely available. Instead of measuring gestural parameters and identifying gesture forms from human dialogs, we experiment with values from literature and human annotations. We first carry out experiments to verify whether our parameters values communicate desired personality. On the basis of those results, we experiment on gestural adaptation.

We set up our experiment in the form of two virtual agents co-telling a story and

conduct two experiments using four different dialogic stories. We first verify that we can express the personalities of virtual agents on the extraversion scale through gestures, regardless of agent gender and the story they are telling. We then show that human subjects prefer adaptive to nonadaptive virtual agents, where the adaptive virtual agent adapts the gesture form and personality to their dialog partner. Our studies support that determining gesture features according to the dialog partner's personality could play an important role in gestural adaptation, and we hope they can motivate further studies in this area.

1.1 Motivation

There is substantial evidence that human conversants adjust their behavior to their conversational partner, either due to priming [164], or beliefs about their partner's knowledge and understanding [39, 86, 185], or to serve social goals such as communicating liking or to show distance [34, 66]. Clark and Wilkes-Gibbs state that participants in a conversation are mutually responsible for establishing what the speaker meant. One part of that process involves speakers and addressees work together in the making of a definite reference [40]. Chartrand and Bargh argue that the tendency to adopt the behaviors, postures, or mannerisms of interaction partners might have played an important role in human evolution by allowing individuals to maintain harmonious relationships with fellow group members [34].

Conversants lexically adapt or align to particular ways of referring to the same object [22], even when their partner is a computer agent [18, 21]. Stoyanchev and Stent states that users adapt to the system's choice of time form: e.g. four o'clock vs. four in the morning [167]. Parent and Eskenazi shows that users adapt their vocabulary to the dialog system's, in terms of distribution of words [141]. Individuals also adapt to each other's non-content speech features, such as segmental phonology and other dialect fea-

tures, speech rates, pause and utterance lengths [41]. Willemyns et al. shows adaptation to the accent of the interviewer in a job interview setting [187].

Conversants mimic their interaction partners' postures, mannerisms, facial expressions and other behaviors unconsciously [34]. They also mimic each other's co-speech gestures in face-to-face dialogs under a task-oriented setting [81]. Experiments have also shown that the copying of hand gestures is driven by representations at the conceptual level [124].

Studies have shown that adaptation measures are correlated with task success [154], dialog naturalness [132], user satisfaction [149] and learning gains [183], and that social variables such as power effect the directionality of adaptation [45]. There is also evidence that people prefer virtual agents that align with human behavior, such as by mimicking head movements [8, 156] or speech style [125], and human attraction to a virtual agent is increased when the agent adapts its personality to the human over time rather than maintaining a consistently similar personality [125]. When creating character-based interactive systems, researchers put a tremendous amount of effort into making an intelligent virtual agent more human-like. However, a majority of these efforts are put into creating handcrafted virtual agent behaviors, which are inflexible and not scalable. Recently, as the effects of linguistic and gestural adaptation have become clearer, it is obvious that just replaying existing handcrafted behaviors is not enough and being able to dynamically adapt to users is a key part of successful dialogs.

The other missing piece of the story is the constraints of expressing speakers' own linguistic styles. Every person is a unique individual with their unique patterns for verbal and non-verbal expression as well as adaptation. However, when interacting with others, people make inferences that generalize from specific behaviors to explanations for those behaviors in terms of dispositional traits [135]. The Big Five theory of personality is one theory that attempts to explain such inferences. It assumes that the consistent

patterns that color individual behavior, feeling, and thinking across different situations, can be described in terms of trait adjectives, such as sociable, shy, trustworthy, disorganized or imaginative [120, 137].

Previous research suggests that personality traits are useful as a basis for modeling Intelligent Virtual Agents (IVAs) for a wide range of applications [15, 77, 79, 97, 172]. Many findings about how people perceive other humans appear to carry over to their perceptions of IVAs [4, 53, 125, 136, 159]. Human users are more engaged and thus learn more when interacting with characters endowed with personality, and a character's personality, surprisingly, affects users' perceptions of the system's competence [171, 179]. Recent experiments show that the Big Five theory is a useful basis for multi-modal integration of nonverbal and linguistic behavior [104, 116], and that automatically generated variations in personality are perceived as intended [115, 127, 128].

Previous research has attempted to integrate linguistic adaptation with natural language generation. Isard, Brockmann, and Oberlander use n-gram models to generate dialogs between pairs of agents with personality and adaptation [87]. Buschmeier, Bergmann, and Kopp introduces the alignment-capable microplanner SPUD *prime* that models adaptation to lexical and syntactical features using activation functions [28, 29] Dušek and Jurčiček generate responses adapting to users' utterances with recurrent neural network and an n-gram ranker [48]. They prepend the user utterance to the responding dialog act and feed into the encoder. However, these methods have no explicit parameter control, and the outputs are often evaluated by computational method rather than human perception.

Moreover, adaptation is not simple mimicking. The process of adaptation takes place under other special constraints, e.g. providing coherent and informative turns in conversation, expressing one's own personalities, or achieving other social and interper-

sonal goals. When implementing dynamic adaptation in natural language generation, we need to take these constraints into account. A sensible approach is to measure adaptation in human dialogs, and use models produced by these measures to control adaptation behaviors. Recent measures of linguistic adaptation fall into three categories [188]: probabilistic measures, document similarity measures, and repetition decay measures.

- Probabilistic measures: these measures compute the probability of a single linguistic feature appearing in the target after its appearance in the prime. Some measures in this category focus more on comparing adaptation amongst features and do not handle turn by turn adaptation [37, 166]. Moreover, these measures produce scores for individual features, which need aggregation to reflect overall adaptivity [43, 45].
- Document similarity measures: these measures calculate the similarity between prime and target by measuring the number of features that appear in both prime and target, normalized by the size of the two text sets [180]. Both probabilistic measures and document similarity measures require the whole dialog to be complete before they can be calculated.
- Repetition decay measures: these measures observe the decay rate of repetition probability of linguistic features. Previous work has fit the probability of linguistic feature repetition decrease with the distance between prime and target in logarithmic decay models [152, 153, 155], linear decay models [182], and exponential decay models [148].

However, a majority of these measures are descriptive and not useful for controlling adaptation behaviors in natural language generation. Moreover, these measures often focus on single linguistic features. Controlling adaptation behavior on single features

is not enough to satisfy special constraints of adaptation, such as expressing one’s personality. Personality is not expressed through a single linguistic feature. Studies have shown that extraversion is positively correlated with a number of Linguistic Inquiry and Word Count (LIWC) features, such as frequency of swear words and positive feeling words [116]. In summary, we need an adaptation measure that (1) produces models that can be used to control adaptation in natural language generation, and (2) models adaptation on a set of linguistic features.

Finally, much previous work has proven gestural adaptation to be real and useful. Tolins et al. show that in a face-to-face dialog, gestures of two extraverted speakers move together towards a more extraverted personality [177]. Luo, Ng-Thow-Hing, and Neff demonstrate that users have preference for motions similar to their own in computer agents [113]. However, no previous work has implemented nor evaluated gestural adaptation systematically in dialogs.

1.2 Linguistic Adaptation

We first describe our exploratory study on natural language generation with adaptation to features, such as referring expressions, pragmatic markers and syntactic structures. Our system Personage-primed, discussed in Section 4.1.1, has the ability to adapt on all or any of the combinations of the adaptation features. In Section 4.1.2, we present a method of testing human perceptions on (1) adapted utterances with different feature combinations, (2) an utterance without adaptation, and (3) a random human utterance that has the same meaning. Human participants are asked to do surveys, in which, given the context of the utterance, they rank a subset of those utterances on both naturalness and friendliness. The results in Section 4.1.3 show that human judgments of naturalness are distinct from friendliness: adapting on a user’s hedges increase perceptions of friendliness while reducing naturalness, while adapting on user’s referring expressions,

syntactic template selection and tense/modal choices increase perceptions of both naturalness and friendliness.

We then propose an algorithm for adaptive natural language generation for dialog that integrates the predictions of both personality theories and adaptation theories. Natural language generators need to operate as a dialog unfolds on a turn-by-turn basis, thus the requirements for a model of adaptation are different from those for simply measuring adaptation. Another challenge is that dialogs exhibit many different types of linguistic features, any or all of which, in principle, could be adapted. Previous work has often focused on individual features when measuring adaptation, and referring expressions have often been the focus, but the conversants in the dialog in Figure 3.1 from the Art-Walk Corpus appear to be adapting to the discourse marker *okay* in D98 and F98, the hedge *kinda like* in D100 and F100, and to the adjectival phrase *like a vase* in D101.

We present a method to learn an adaptation model from a corpus that can be applied to any segment of a dialog as the dialog unfolds, and which can model any possible subset of linguistic features. We show that, by focusing our adaptation model on features correlated with personality, we can learn how to control an adaptive NLG that expresses personality. We apply our method to multiple corpora to investigate how the dialog situation and speaker roles affect the level and type of adaptation to the other speaker. We show that:

- The models for adaptation depend on the feature set used;
- Different conversational situations can have different adaptation models;
- The level of adaptation varies according to which speaker has the initiative;
- The degree of adaptation varies over the course of a dialog.

DAS scores calculated using human dialogs can be expressed in a vector form where each dimension contains the DAS score for a linguistic feature set. This DAS vector

can be used as an adaptation model in various natural language generation architectures to control the amount of adaptation. In overgenerate and rank, the system can calculate a DAS score for each response, rank all possible responses by the distance between its DAS score and the adaptation model. The best response is the one with the smallest distance to the adaptation model. In statistical parameterized natural language generation, DAS scores can be used as probability based parameters. In natural language generation using neural networks, our adaptation model can be encoded into the context vector.

1.3 Gestural Adaptation

To explore our framework of gesture adaptation, we use a gesture synthesis system built on top of the Dynamic Animation and Control Environment (DANCE) [165] as our simulation platform, and constructed an experiment stimulus, in which two IVAs collaboratively tell a story. Our stories come from a weblog corpus of monologic personal narratives [71] whose content has been regenerated as dialogs to support story co-telling. Example dialogs from the four we use in our experiments are provided in Fig. 3.14 and Fig. 3.13 in Sec. 3.5. These dialogs have a fixed linguistic representation and use oral language, discourse markers, shorter sentences, and repetitions and confirmations between speakers, as well as techniques to make the story sound like the two speakers experienced the event together. Our aim is to mimic the finding that storytelling in the wild is naturally conversational [170, 173, 174], e.g. Thorne’s work shows that the style of oral storytelling among friends varies depending on their personalities [173].

In the stimuli, we vary (1) virtual agent personality through gestural parameters of gesture rate, speed, expanse and form; (2) whether the agents adapt to one another’s gestures in gesture rate, speed, expanse and form and use of specific gestures. We aim to test the effect of, and interaction between, these variations with human perceptual

experiments and report our results. We carry out two experiments. In the personality experiment, we aim to have subjects perceive the two virtual agents as having the designed personality expressed through a vector of gesture parameters. In the gestural adaptation experiment, we aim to determine whether subjects' prefer adaptive vs. non-adaptive agents. Our results show that agents intended to be extraverted or introverted are perceived as such, and that subjects prefer adaptive stories.

1.4 Contributions

This thesis proposes an adaptation framework for linguistic and gestural adaptation. We aim to build adaptation models in forms of vectors, which can be used to produce natural adaptation behavior in virtual agents. Through a series of experiments, we explore how to build such models in linguistic adaptation, and test the effects of our gestural adaptation models. The contributions of this thesis are:

- We build a natural language generator, Personage-primed, with linguistic adaptation capability on any combination of linguistic features. Using Personage-primed, we perform exploratory experiments to evaluate all possible combinations of adaptation in a particular discourse context in order to test whether some types of adaptation are preferred, either because they make the utterance more natural, or because humans perceive the system as more friendly. Our experimental results suggest that human judgments of naturalness are distinct from friendliness: adapting to a user's hedges increases perceptions of friendliness while reducing naturalness, while adapting to user's referring expressions, syntactic template selection, and tense/modal choices increase perceptions of both naturalness and friendliness.
- We propose a measure of linguistic adaptation, the Dialog Adaptation Score (DAS).

DAS can model any possible subset of linguistic features and can be applied on a turn by turn basis to any segment of dialog as the dialog unfolds. We first show that DAS meets criteria for validity. We then apply DAS to four dialog corpora and show that adaptation varies according to corpora and task, speaker, and the set of features used to model it. We also show that we can model adaptation with high level personality features and that we can produce fine-grained models according to the dialog segmentation or the speaker. Adaptation models learned using DAS can be used in both rule-based and machine learning natural language generation with adaptation.

- We explore the expression of personality and adaptivity through the gestures of virtual agents in a storytelling task. We conduct two experiments using four different dialogic stories. We manipulate agent personality on the extraversion scale, whether the agents adapt to one another in their gestural performance and agent gender. Our results show that subjects are able to perceive the intended variation in extraversion between different virtual agents, independently of the story they are telling and the gender of the agent. A second study shows that subjects also prefer adaptive to nonadaptive virtual agents.

In the rest of this thesis, we discuss previous work on linguistic adaptation, gesture adaptation and personality in Chapter 2; Chapter 3 introduces datasets; Chapter 4 presents our work in implementing, evaluating and measuring linguistic adaptation; Chapter 5 presents our work in implementing and evaluating gestural adaptation. Chapter 6 concludes and describes future work.

Chapter 2

Related Work

In this section, we first introduce studies about partner adaptation and theories of adaptation, in which we give evidence for both linguistic and gestural adaptation. Second, we review previous work on measuring linguistic adaptation. Third, we discuss adaptation in current natural language generation systems. Fourth, we summarize theories and studies of personality and the relationship between personality and gesture. Finally, we discuss recent work about gesture adaptation and generation.

2.1 Partner Adaptation and Theories of Adaptation

The *collaborative theory of language use* [22, 38–40, 58–60, 86, 162–164, 186] is largely supported by previous research on human communication. This theory states that speakers adjust their behavior based on their beliefs about their conversation partner’s knowledge and understanding. Research shows that speaker lexically adapt to each other’s ways of referring to things [22, 86, 123], adjust to dialog partner’s linguistic style [134], and use of function words [44, 85]. This phenomenon is also referred to as linguistic alignment, entrainment, adaptation, or accommodation [132].

However, over accommodation happens when conversants produce behaviors based

on stereotypes of their partners, which leads to negative perceptions. For example, using baby talk when conversing with the elderly [160]. On the other hand, although some speech behaviors are predicted to show accommodation, research has suggested otherwise. For example, Heinz shows that the rate and style (overlapping versus non-overlapping) of backchannels, which vary cross-culturally, does not show the convergence pattern expected [78].

Communication accommodation theory [66] proposes that conversants in a dialog mimic their partner's behaviors in order to achieve social goals, such as showing liking. Speakers mimic each other's speech accent, speech rate, pause length, utterance length, and lexical diversity [41, 187]. Jurafsky, Ranganath, and McFarland shows that friendlier and less-awkward participants in a speed-dating setting were more likely to use collaborative techniques, such as the use of questions and laughter [89]. Gueguen, Jacob, and Martin shows that women who mimicked men's verbal expressions and non-verbal behaviors were liked more [76]. A computer animated agent who mimicked the user's head movements increases the persuasiveness of the message and the liking of the agent [9]. In terms of personality adaptation, a widely claimed effect is similarity attraction [30], which suggests that people like others with personalities similar to their own. Similarity-attraction would predict that a virtual agent with personality which matches the user's will be preferred. However, other research has shown that in some situations, adaptation is not necessarily related to liking. Niederhoffer and Pennebaker shows that in a study of spontaneous writing, although participants matched each other's linguistic styles across conversations, they did not necessarily like each other more [134].

The collaborative theory of language use and communication accommodation theory both suggest that at least some communicative behavior will vary based on partner specificity. Although both theories are based on audio data primarily, we believe that both theories will easily extend into the gestural domain. In terms of the collaborative

theory of language use, we propose that speakers adapt to each other's gestures in conversations; in terms of accommodation theory, we propose that conversants in a dialog mimic their partner's gestures in hope to create a positive interaction experience.

Previous research has shown strong evidence that gestural adaptation does occur, and that it has had positive communicative effects. Rizzolatti and Arbib shows that mirror neurons in the brain activate when observers view others performing actions [158]. Although mirror neuron research began with monkeys, experiments with humans have also shown the existence of mirror neurons [1]. Viewing gestures result in brain effects that are predicted to be beneficial on communication [55, 84].

Kimbara carries out experiments where participants collaborated on retelling a story to a listener. Participants produced more similar hand gestures when they could see each other than when they could not [93]. The results of this experiment provide evidence that conversants in a face to face dialog adapt to each other's gestures when expressing similar contents. However, as Kimbara points out, the study only looks at iconic gestures. Thus, it remains unclear whether other types of gestures, such as beat and metaphoric gestures, also show signs of adaptation. Parrill and Kimbara shows that participants who observed conversants using gestural mimicry were more likely to reuse those gestures when retelling video clips, than participants that observed conversants who did not use gestural mimicry [142]. Participants who heard word mimicry were also more likely to reuse mimicked words. However, Parrill and Kimbara shows that mimicked words and mimicked gestures were independent of each other. Researchers believe that adaptation (entrainment, alignment, convergence) occurs beneath conscious awareness [18, 142].

Taken together, research strongly supports the theory of communicative adaptation in both verbal and non-verbal behaviors. A computer agent without the ability to adapt would be behaving in a noticeably non-human way.

2.2 Measuring Lexical Adaptation

Measures of linguistic adaptation fall into three categories: probabilistic measures, repetition decay measures, and document similarity measures [188]. When measuring linguistic adaptation, both methods divide documents into “prime” and “target” part, in which the prime part contains linguistic features that the target part may adapt on.

Probabilistic measures compute the probability of a single linguistic feature appearing in the target after its appearance in the prime. Some measures in this category focus more on comparing adaptation amongst features and do not handle turn by turn adaptation [37, 166]. Moreover, these measures produce scores for individual features, which need aggregation to reflect overall adaptivity [43, 45]. Probabilistic measures aim to find out (1) whether or not adaptation happens, and (2) what type of features are more likely to be adapted. Probabilistic measures require the whole dialog to be complete before they can be calculated.

Church’s [37] method for measuring lexical adaptation in text determines whether the appearance of a lexical feature in the prime portion of a document affects the likelihood of its appearance in the target portion. For each feature in a corpus, this method counts how many of the documents contain the feature: (a) in the priming portion only, (b) in the target portion only, (c) in both portions, and (d) in neither portion. The method computes the probability of positive adaptation as $\frac{c}{a+c}$, compared with a prior probability $\frac{a+c}{a+b+c+d}$. Church applied this method to a corpus of text, taking the first half of a document as the priming portion and the second half as the target portion. Results show that positive lexical adaptation does occur, and content words have stronger adaptation than function words.

Dubey et al. [47] use Church’s method, and evaluate adaptation for some of the syntactic features annotated in the Brown corpus and the Switchboard corpus. Their results show positive adaptation for each of the syntactic structures they test.

Stenchikova and Stent [166] differ partner adaptation (adaptation to conversational partner) and recency adaptation (adaptation to the local dialog content), and measure both adaptation ratio and strength by using the frequency of occurrence of prime and target features rather than merely its presence or absence. They define the adaptation ratio for a certain feature as $+adapt/chance$, in which *chance* is the probability of a feature co-occurring in prime and target by chance, and $+adapt$ is the probability of positive adaptation (the probability that the features occurs in target given it occurs in prime). They define the adaptation strength for a certain feature using a *distance* measure. The *distance* is the difference between a feature's frequency in target and average (mid-point) frequency of prime and baseline frequency. They consider a feature to be adapted in a pair of dialogs if the target frequency is closer to that of prime than baseline. The measuring results on the Maptask Corpus [2] show that for syntactic features, recency adaptation is stronger than partner adaptation.

Repetition decay measures observe the decay rate of the repetition probability of linguistic features. Previous work has fit the probability of linguistic feature repetition decrease with the distance between prime and target in logarithmic decay models [152, 153, 155], linear decay models [182], and exponential decay models [148].

Reitter et al. [153] model priming of syntactic rules in the Maptask corpus (task-oriented dialogs) and Switchboard corpus (spontaneous conversations). Every pair of two equal syntactic rules is considered to be a potential case of adaptation if it is within a predefined maximal distances (15 seconds). They use generalized linear mixed effects regression models to show that priming exists, and to predict the decline of repetition probability with increasing distance between prime and target and depending on other variables. Their results show that speakers are more receptive to priming from their interlocutor in task-oriented dialog than in spontaneous conversation. Low-frequency syntactic rules are more likely to show priming.

Ward and Litman [182] use a corpus of human-human tutoring transcripts [105] to measure adaptation. In these tutoring sessions, a human tutor presents a problem in qualitative physics to a student, who answers it in essay form. The tutor examines this essay, identifies flaws in it, and engages the student in a tutorial dialog to remediate those flaws. They also create a corpus of randomized tutoring dialogs from the previous corpus by leaving tutor utterances in the original positions, while randomizing the order of the student utterances. Similar to Reitter et al., they define the next N student turns as a window, look for lexical priming inside the window, and then count the target distance and frequency. Linear regression is used to determine the relationship between distance from the prime and lexical repetition count. Their measures discriminate randomized from naturally ordered data, and demonstrate both lexical and acoustic/prosodic convergence.

Pietsch et al. [148] also use the Switchboard corpus and MapTask corpus. Similar to Reitter et al., they also measure the temporal distance between the occurrence of adjacent syntactic features. Their null hypothesis is a random distribution, described as a Poisson process, where the distances are exponentially distributed: $p(x) = \lambda_0 e^{-\lambda_0 x}$ (λ_0 is the frequency of the feature within the corpus, $p(x)$ is the expected frequency of seeing the next instance of the feature at exactly distance x). They compare this expected distribution of distances to the actual distribution, where the actual distribution is fitted in an exponential curve with decay parameter λ . They interpret the ratio $r = \frac{\lambda}{\lambda_0}$ as the strength of priming: the more the fitted parameter deviates from the expected one, the more skewed is the distribution. Their results show that λ is larger than its expected value, which is the feature's frequency. The effect appears to be larger for rare features.

Document Similarity measures originate from information retrieval. The measures calculate the similarity between prime and target. Based on Fusaroli et al. [64], Wang, Reitter, and Yen [180] measure the number of features that appear in both prime and

target, normalized by the size of the two text sets. They create two kinds of prime-target pairs using online health forum threads: posts within thread vs. posts in different threads. They observe adaptation at lexical and syntactic levels, as well as decay of adaptation value as the post distance increases. Document similarity measures also require the whole dialog to be complete before they can be calculated.

2.3 Adaptation in Natural Language Generation

As a natural phenomenon in conversations, adaptation is also taken into account in various natural language generation systems to achieve better generation results.

Jong et al. [46] presents an approach that focuses on affective language use for aligning specifically to user's politeness and formality. Brockman et al. [26] illustrates a model in which alignment is simulated using word sequences alone. An extension of this work in Isard et al. [87] simulates both personality and alignment in dialog between pairs of agents with the CrAg-2 language generation system. The system generates a dialog between two agents discussing a film. Its natural language realizer takes as input a logical form, and outputs numerous generated utterances which are ranked using one or more language models. However, their underlying method has no explicit parameter control.

Further extension of Isard, Brockmann, and Oberlander's work by Gill et al. [69] also use the CrAg-2 language generation system [87], which provide a framework for generating dialogs between two computer characters discussing a movie. Within CrAg-2, linguistic personality and alignment are modeled using the OpenNLP CGG Library (OpenCGG) natural language realizer [184]. Language models are trained on a corpus of weblogs from authors of known personality. Alignment is modeled via cache language models (CLMs): for each utterance to be generated, a language model is computed based on the utterance that was generated immediately before it. They exam-

ine how accurately judges can perceive character personality from short, automatically generated dialogs, and how alignment alters judge perceptions of the characters' relationship. They find that the personality perception of the dialogs is consistent with perceptions of human behavior, but the introduction of alignment leads to negative perceptions of the dialogs and the interlocutors' relationship. A follow up evaluation study of the perceptions of different forms of alignment in the dialogs reveals that while similarity at polarity, topic and construction levels is viewed positively, similarity at the word level is regarded negatively.

Buschmeier et al. [28, 29] introduced the alignment-capable microplanner SPUD *prime*. SPUD *prime* is a computational model for language generation in dialog that focuses heavily on relevant psycholinguistic and cognitive aspects of the interactive alignment model. Their system is driven by a method of activating relevant rules in a detailed contextual model according to user behavior during a dialog. They adopt an idealized view, in which priming of linguistic structures results from two basic activation mechanisms: temporary activation (increase abruptly and then decrease slowly over time until it reaches zero again) and permanent activation (increase by a certain quantity and then maintain the new level). Part of the activation functions originates from the repetition decay model from Reitter [152]. However, the parameters are not learned from real data. Repetition decay models do well in statistical parameterized NLG, but are hard to apply to overgenerate and rank NLG. They use an exponential decay to model temporary activation and a modified exponential saturation function to model permanent activation. A combined model is built for a model of alignment. SPUD carries out microplanning tasks including lexical choice, syntactic choice, referring expression generation and aggregation. Although the underlying system seems to be capable of producing both syntactic and lexical alignment, it is evaluated for accurate representation of lexical alignment in a corpus of dialogs from a controlled experiment.

Dušek and Jurčíček [48] present a natural language generation system generator based on recurrent neural networks and the sequence-to-sequence approach. The system is able to generate responses adapting to previous context in a bus information system domain using previous context and response dialog act as input. They use an n-gram match ranker that promotes those outputs that have phrases overlapping the context. The generation results are evaluated using both automatic metrics and human evaluation. However, the BLEU and NIST scores are lower than the baseline (context not used) without the n-gram rankers.

2.4 Theories and Studies of Personality and Gesture

Over the last fifty years the Big Five theory has become a standard in psychology. There is significant evidence that personality traits are both *real* and *useful*. Research has shown that: (1) individual differences in self-reported traits are significantly associated with trait-consistent behavioral trends when behavior is aggregated across situations; (2) traits are powerful predictors of important life outcomes; (3) individual differences in traits show substantial longitudinal consistency; (4) traits are significantly heritable; and (5) traits appear to be complexly linked to specific brain processes (e.g., the amygdala, prefrontal cortex) and to certain neurotransmitters (e.g., dopamine) [119]. From a computational perspective, the correlations that have been systematically documented between a wide range of verbal and nonverbal behaviors and the Big Five traits are incredibly *useful* [63, 80, 120, 137, 145]. Recent work has used these findings to develop rule-based personality models for Big Five traits and apply them on a natural language generator. Results to date suggest that perceptions of agreeableness, neuroticism, and extraversion are easier to model, whereas conscientiousness and openness to experience are more difficult.

In the aspect of nonverbal expression of personality, a summary for the extraversion

Parameters	Introvert Findings	Extravert Findings
Body attitude	backward leaning, turning away	forward leaning
Gesture amplitude	narrow	wide, broad
Gesture direction	inward, self-contact	outward, table-plane and horizontal spreading gesture
Gesture rate	low	high more movements of head, hands and legs
Gesture speed, response time	slow	fast, quick
Gesture connection	low smoothness, rhythm disturbance	smooth, fluent
Body part		head tilt, shoulder erect, chest forward, limbs spread, elbows away from body, hands away from body, legs apart, legs leaning, bouncing, shaking of legs

Table 2.1: The gestural correlates of extraversion.

Parameters	Low-neurotic Findings	Neurotic Findings
Body attitude	more movement and more smooth movement	more posture changes, less relaxed posture, more forward lean with self reported anxiety
Gesture length		shorter gestures
Gesture direction		fewer other directed gestures, also self references, fewer outward gestures
Gesture rate	short unfilled pauses, fewer filled pauses	high activity level, long unfilled pauses, more filled pauses
Gesture speed, response time	more uniform velocity	more variation in gesture velocities, longer pauses before responding
Gesture connection		pauses for longer duration during speech, longer pauses before responding
Body part		head lowering, more gaze aversion with self reported anxiety, less eye contact with interviewer for higher anxiety
Other features		more self touch, greater distance from others

Table 2.2: The gestural correlates of neuroticism.

sion trait is shown in Table 2.1. Postural and gestural styles are linked to personality, attitude and status in relationships [88, 121]. The position of the head and trunk are the most visually salient indicators of status and attitude (**body attitude** in Table 2.1); leaning forward communicates a relatively positive attitude to the interlocutor whereas leaning backward or turning away communicates a more negative attitude. Leaning the torso forward is also positively correlated with extraversion [104]. Frank argues that extraverts amplify a sense of space by moving the upper body (chest and limbs) forward whereas introverts maintain a more vertical orientation [62].

Several studies have shown that gestural expansiveness and range of movement (**gesture amplitude and direction** in Table 2.1) is positively correlated with extraversion [7, 20]. Specifically extraversion is positively correlated to factors like “expan-

Parameters	Low-agreeable Findings	Agreeable Findings
Body attitude	closed posture	open, leaning forward, open posture
Gesture amplitude	open	
Gesture length	shorter gestures	longer gestures
Gesture direction	more sagittal plane gestures, more aggressive gestures at other	more horizontal plane gestures, more positive/supporting gestures at other
Gesture speed, response time	ease-in (accelerate towards end), starts and stops	even (ease-in, ease-out)
Gesture connection	many pauses	few pauses
Gesture form	more palm vertical (hand pointed forward), or palm down; more negative form gestures (e.g. wipes, dismiss)	more palm up, or palm towards subject with hand pointed up, fewer negative form gestures.
Body part	decreased nods	increased nods

Table 2.3: The gestural correlates of agreeableness.

sive”, “broad gestures”, “elbows away from body”, “hands away from body”, and “legs far apart while standing” [96, 104, 157]. Gesture direction is also important. Argyle [7] states that introverts use fewer outward directed gestures and touch themselves more. North [138] indicates that extraverts likely show a significant number of table plane and horizontal spreading gestures. Takala’s analysis [169] supports the hypothesis that introverts use more inward directed movements in the horizontal dimension and extraverts more outward directed movements. Furthermore, movements directed away from the person could be an indication of aggressiveness, while inward directed shifts indicate passiveness.

Extraverts are found to be more energetic or have more physical strength, and higher gesture rates, while the gestures of introverts persist more [7, 20, 67, 96, 104]. A number of studies have examined the temporal properties of gestures (**gesture rate, speed, response time and connection** in Table 2.1) [96, 157]. Extraverts tend to have faster speech, which leads to higher gesture rates due to the correlation between speech and gesture. This has been experimentally demonstrated by Lippa [104]. Brebner [20] also found differences between introverts and extraverts in speed and frequency of movement. Extraverts not only behave in a more rapid manner than introverts, the time to first response, or the response latency, is shorter as well. Results related to the smoothness and rhythm of gesture suggest that introversion is negatively correlated with smoothness

and positively correlated with rhythm disturbance [104, 157, 169].

Other research discusses extraversion and its relation to particular body parts in gesturing (**body part** in Table 2.1). Knapp [96] mentions more leg lean for an ambitious personality. Experiments by Riggio [157] suggest extraverts have more “body emphasis”, defined as more head movements, more parallel gestures of the hands, more movement of the legs (position, bouncing, shaking) and more posture shifts. Besides using broad gestures, Lippa [104] also found that extraverts use most of their body when gesturing, tend to tilt their heads and raise their shoulders. Extraverts are universally believed to maintain more eye contact, and a positive correlation between eye contact, shoulder orientation, leg orientation, and body orientation indicates extraversion [121]. In addition, findings indicate that spatial behavior differs for extraverts [7, 138], e.g. extraverts stand closer to others, either because of greater tolerance for a close interaction distance, or because of high individual confidence, self-esteem, or assertiveness.

A summary for the neuroticism trait is shown in Table 2.2, and agreeableness trait in Table 2.3.

2.5 Gesture Adaptation and Generation

Gestural adaptation, imitation, mimicry, and alignment can be classified into two main groups based on the cognitive mechanisms behind these non-verbal behaviors.

The first group include non-conscious mimicry of behaviors. Recent studies consider this type of alignment as social glue. Because there is a tight linking between perception and behavior, that leads an individual perceiving another’s behavior to automatically behave in a similar way [34, 35]. Behaviors in non-conscious mimicry include foot tapping and face scratching [34], facial expressions [11, 12], also vocal features such as speech rhythms [31], tone [133], and even accents [68]. During interaction, Individuals who mimic others are considered as more likable, and are more

likely to have their actions viewed positively [76]. More mimicry is seen as having a higher sense of affiliation and rapport [35, 99, 100].

The second group relies on the notion of the collaborative nature of language use. There are recent studies focusing on the relationship between mimicry of co-speech gestures and the development of a shared underlying meaning [13, 81, 94, 124]. Gesture adaptation across speakers is considered to be similar to conceptual pact development in which the two interlocutors develop a common understanding of how an object is to be represented verbally (e.g. [23]), in other words, language adaptation. Because words and gestures are thought to represent a common communicative intent, adaptation in one modality is matched by similar behavior in the other. Similarly, gestural imitation does not occur when there is a mismatch between what is said and what is gestured by the original speaker, which indicates gestural adaptation is based on the underlying semantic conceptualization [124]. In certain contexts, gestural mimicry can stand in for the verbal adaptation, allowing speakers to make use of a more vague vocabulary [81].

Mol et al. [124] carry out three experiments aiming to find out the relationship between the copying of co-speech hand gestures' form and (1) their meaning in the linguistic context, as well as (2) interlocutors' representations of this meaning at the conceptual level. For the three experiments, two of them use recorded videos of a person telling a story, the videos have different versions with variations in the person's gestures. The other experiment has two participants engaged in a face-to face direction-giving task. Their analysis of the experiments show that gestures were repeated only if they could be interpreted within the meaningful context of the speech. There is also evidence that the copying of gesture forms is mediated by representations of meaning. In other words, representations of meaning are also converging across interlocutors rather than just representations of gesture form. The conclusion is that adaptation of representational hand gestures may be driven by representations at the conceptual level, and same for lexical

adaptation. That is, adaptation in gesture resembles adaptation in speech, rather than it being an instance of automated motor-mimicry.

Luo et al. [113] examine whether people prefer gestures that are similar to their own. There is evidence for the “chameleon effect” that in conversation, people will tend to adopt the postures, gestures and mannerisms of their interaction partners. Luo’s research look at the mirroring effect in human-agent interaction by having the agent perform gestures similar to the human. Their study explores if people prefer gestures similar to their own over gestures similar to those of other people. Participants were asked to evaluate a series of agent motions, some of which mimic their own gestures, and rate their preference. A second study first showed participants videos of their own gesturing to see if self-awareness would impact their preference. Different scenarios for soliciting gesture behavior were also explored. Evidence suggests people do have some preference for motions similar to their own, but self-awareness has no effect.

Tolins et al. [177] compared the behavior of an extravert-extravert dyad to an extravert-introvert dyad engaged in face-to-face conversations. Their findings show that in the extravert-introvert dyad, the extravert adapted to the introvert’s gesture rate, reduced the expansiveness of his arm positions to match the introvert, and both the introvert and the extravert move towards each other in gesture broadness. The extravert-extravert dyad adapt on gesture rate and gesture outwardness, but two extraverts moved together towards more open arm positions. Their results show the following implications for IVA: (1) agents need to adapt to their speech partner, and (2) this adaptation needs to be modeled based on the personality the agent is trying to adapt to.

Recent work on gesture generation focused largely on iconic gesture generation. For example, Bergmann and Kopp [14] present a model that allows virtual agents to automatically select the content and derive the form of coordinated language and iconic gestures. Their study is based on an empirical study of spatial descriptions of land-

marks in direction-giving. First, two kinds of knowledge representation (propositional and imagistic) are utilized to capture the modality-specific contents and processes of content planning. Second, specific planners are integrated to carry out the formulation of concrete verbal and gestural behavior. A probabilistic approach to gesture formulation is presented that incorporates multiple contextual factors as well as idiosyncratic patterns in the mapping of visual-spatial referent properties onto gesture morphology.

In terms of how gestures are selected, current systems are either text-to-gesture or concept-to-gesture. Text-to-gesture systems such as VHP [136] have a limited number of gestures (only 7 in this case) and limited gesture placement options, but the alignment of speech content and gestures are more accurate. Concept-to-gesture systems such as PPP [5], BEAT [33] and AC [32] defines general rules for gesture insertion based on linguistic components. For example, Iconic gestures are triggered by words with literally spacial or concrete context (e.g. “check”). These kind of systems have more gestures, but the gesture placement largely depends on general rules derived from literature, thus the accuracy is not guaranteed.

Chapter 3

Datasets

In this section, we introduce all datasets used in this thesis. The ArtWalk Corpus (AWC), the Walking Around Corpus (WAC) and the Map Task Corpus (MPT) are task-oriented, direction-giving dialog corpora. Switchboard Corpus (SWBD) is a topic-centric spontaneous dialog corpus. AWC is used as evaluation data in our exploratory study of testing human perceptions on linguistic adaptation in Chapter 4, Section 4.1. All four corpora are used for measuring adaptation in Chapter 4, Section 4.2. We build the Story Dialog with Gestures Corpus for implementing and evaluating gesture adaptation in Chapter 5. The corpus contains selected stories with story structure annotated, the stories are also converted to a two-person dialog annotated with gesture.

3.1 ArtWalk Corpus (AWC)¹

Figure 3.1 provides a sample of the ArtWalk Corpus [108], a collection of 48 mobile-to-Skype conversations between friend and stranger dyads performing a real world-situated task that was designed to elicit adaptation behaviors. Every dialog involves a stationary director on campus, and a follower downtown. The director provided di-

¹<https://nlds.soe.ucsc.edu/artwalk>

Speaker [Utterance #]: Utterance
F97: okay I'm on pacific avenue and plaza
D98: okay so you just take a right once your out of pacific lane you go wait no to late to your left.
F98: okay
D99: and I think. it's right ther- * alright so im walking down pacific* okay so it's right before the object it's right before the mission and pacific avenue intersection *okay* it's like umm almost brown and kinda like tan colored
F99: is it tan
D100: yeah it's like two different colors its like dark brown and orangey kinda like gold color its kinda like um
F100: okay is it kinda like a vase type of a thing
D101: yeah it has yeah like a vase
F101: okay yeah I got it okay one second just take a picture. Alright

Figure 3.1: Sample dialog Excerpt from the ArtWalk Corpus.

rections to help the follower find 10 public art pieces such as sculptures, mosaics, or murals in downtown Santa Cruz. The director had access to Google Earth views of the follower's route and a map with locations and pictures of art pieces. The corpus consists of transcripts of 24 friend and 24 stranger dyads (48 dialogs). In total, it contains approximately 185,000 words and 23,000 turns, from conversations that ranged from 24 to 55 minutes, or 197 to 691 turns. It includes referent negotiation, direction-giving, and small talk (non-task talk). To our knowledge, this corpus collection is the first experiment to show that adaptation actually occurs in the context of a real task, while people are out in the world, navigating a natural terrain. The excerpt of dialog from the ArtWalk corpus in Fig. 3.1 illustrates adaptation to discourse cues with *okay* in D98 and F98, and in referring expression adaptation in D101 *like a vase*.

3.2 Walking Around Corpus (WAC)²

The Walking Around Corpus [24] consists of spontaneous spoken dialogs produced by 36 pairs of people, collected in order to elicit adaptation behaviors, as illustrated by Figure 3.2. In each dialog, a director navigates a follower using a mobile phone to 18 destinations on a medium-sized campus. Directors have access to a digital map marked with target destinations, labels (e.g. “Ship sculpture”), photos and followers’ real time location. Followers carry a cell phone with GPS, and a camera in order to take pictures of the destinations they visit. Each dialog ranges from 175 to 885 turns. The major differences between AWC and WAC are (1) in order to elicit novel referring expressions and possible linguistic adaptation, destinations in AWC do not have provided labels; (2) AWC happens in a more open world setting (downtown) compared to WAC (university campus).

3.3 Map Task Corpus (MPT)³

The Map Task Corpus [3] is a set of 128 cooperative task-oriented dialogs involving two participants. Each dialog ranges from 32 to 438 turns. A director and a follower sit opposite one another. Each has a paper map which the other cannot see (the maps are not identical). The director has a route marked on their map; the follower has no route. The participants’ goal is to reproduce the director’s route on the follower’s map. All maps consist of line drawings landmarks labeled with their names, such as “parked van”, “east lake”, or “white mountain”. Figure 3.3 shows an excerpt from the Map Task Corpus.

²<https://catalog ldc.upenn.edu/ldc2015s08>

³<http://groups.inf.ed.ac.uk/maptask/>

Speaker (Utterance #): Utterance
D137: and. you know on the uh other side of the math building like there's the uh, there's this weird, little concrete, structure that is sticking up out of the bricks, don't make any sense.
F138: uh.
D139: yeah you'll see it when you get over there.
F140: okay.
D141: so just keep going and then uh. when you get around the building make a left. and you should be.
F142: when I get around the Physics building make a left?
D143: yeah yeah when you get around to the end here.
F144: okay.
D145: and uh you know th-these brick structures, and. you got one look like uh cutout. with your would you be standing on top of like a flat area. you should be coming up on it right, okay.
F146: yeah. just give me a second, I'm walking past this big water, thing, it's noisy.
D147: take your time.

Figure 3.2: Sample dialog excerpt from the Walking Around Corpus.

3.4 Switchboard Corpus (SWBD)⁴

Switchboard [70] is a collection of two-speaker telephone conversations from all areas of the United States. An automatic operator handled the calls (giving recorded prompts, selecting and dialing another speaker, introducing discussion topics and recording the dialog). 70 topics were provided, for example: pets, child care, music, and buying a car. Each topic has a corresponding prompt message played to the first speaker, e.g. “find out what kind of pets the other caller has.” A subset of 200K utterances (1126 dialogs) of Switchboard have also been tagged with dialog act tags [90]. Each dialog contains 14 to 373 turns. Figure 3.1 provides an example of dialog act tags, such as *b* - Acknowledge (Backchannel), *sv* - Statement-opinion, *sd* - Statement-non-opinion, and *%* - Uninterpretable. We focus on this subset of the corpus which contains 1126 dialogs.

Dialogs in SWBD have a different style from task-oriented, direction-giving cor-

⁴<https://catalog.ldc.upenn.edu/ldc97s62>

Speaker (Utterance #): Utterance
D7: and below the graveyard below the graveyard but above the carved wooden pole.
F8: oh hang on i don't have a graveyard.
D9: okay. so you don't have a graveyard. do you have a fast flowing river.
F10: fast running creek.
D11: ehm mm don't know yeah it could be could be.
F12: is that to the right that'll be to my right to my right.
D13: to your. right uh-huh.
F14: right. so i continue and go below the fast running creek.
D15: no. go just until you go go below the diamond mine until just before the fast fast flowing river.

Figure 3.3: Dialog excerpt from the Map Task Corpus.

pora. Figure 3.4 illustrates how the SWBD dialogs are often lopsided: from utterance 14 to 18, speaker B states his opinion with verbose dialog turns, whereas speaker A only acknowledges and backchannels; from utterance 19 to 22, speaker A acts as the main speaker, whereas speaker B backchannels. Some theories of discourse define dialog turns as extending over backchannels, and we posit that this would allow us to measure adaptation more faithfully, so we utilize the SWBD dialog act tags to filter turns that only contain backchannels, keeping only dialog turns with tags *sd* (Statement-non-opinion), *sv* (Statement-opinion), and *bf* (Summarize/reformulate). We then merge consecutive dialog turns from the same speaker. The filtering process removes 48.1% original dialog turns, but only 12.6% of the words. Filtered dialogs have 3 to 85 dialog turns each.

3.5 Story Dialog with Gestures Corpus (SDG)⁵

Sharing experiences by story-telling is a fundamental and prevalent aspect of human social behavior [16, 27, 98, 131]. In the wild, stories are told conversationally in social settings, often as a dialog and with accompanying gestures and other nonverbal behav-

⁵<https://nlds.soe.ucsc.edu/sdg>

Speaker (Utterance #): [Tag] Utterance
B14: [b] Yeah. [sv] Well that's pretty good if you can do that. [sd] I know. [sd] I have a daughter who's ten [sd] and we haven't really put much away for her college up to this point [sd] but, uh, we're to the point now where our financial income is enough that we can consider putting some away
A15: [b] Uh-huh.
B16: [sd] for college [sd] so we are going to be starting a regular payroll deduction
A17: [%] Um.
B18: [sd] in the fall [sd] and then the money that I will be making this summer we'll be putting away for the college fund.
A19: [ba] Um. Sounds good. [%] Yeah [sd] I guess we're, we're just at the point, uh [sd] my wife worked until we had a family [sd] and then, you know, now we're just going on the one income [sv] so it's
B20: [b] Uh-huh.
A21: [sv] a lot more interesting trying to, uh [sv] find some extra payroll deductions is probably the only way we will be able to, uh, do it. [sd] You know, kind of enforce the savings.
B22: [b] Uh-huh.

Figure 3.4: Dialog excerpt from the Switchboard Dialog Act Corpus.

ior [175]. Storytelling in the wild serves many different social functions: e.g. stories are used to persuade, share troubles, establish shared values, learn social behaviors, and entertain [75, 146, 161].

Previous research has also shown that conveying information in the form of a dialog is more engaging, effective and persuasive compared to a monologue [6, 42, 102, 168]. Thus, our long term goal is the automatic generation of story co-tellings as animated dialogs. Given a deep representation of the story, many different versions of the story can be generated, both dialogic and monologic [110, 111].

This section presents a new corpus, the Story Dialog with Gestures (SDG) corpus, consisting of 50 personal narratives regenerated as dialogs by human annotators, complete with annotations of gesture placement and accompanying gesture forms. These annotations can be supplemented programmatically to produce changes in the nonverbal dialogic behavior in gesture rate, expanse and speed. We have thus used these to

Pet Story
I have two cats. I always felt like I was a dog person, but I decided to get a kitty because they are more low maintenance than dogs. I went to a no-kill shelter to get our first cat. I wanted a little kitty, but the only baby kitten they had scratched the crap out of me the minute I picked it up. SO, that was a big “NO”. They had what they called “teenagers”. They were cats that were 4-6 months old. Not adults, but a little bigger than the little kittens. One stood out - mostly because she jumped up on a shelf behind my husband and smacked him in the head with her paw. I had a winner! I had no idea how much personality a cat can have. Our first kitty loves to play. She will play until she is out of breath. Then, she looks at you as if to say, “Just give me a minute, I’ll get my breath back and be good to go.” Sometimes I wish I had that much enthusiasm for anything in my life. She loves to chase a string. It’s the best thing ever. Ok, maybe it runs a close second to hair scrunchies. I play fetch with my hair scrunchies. I throw them down the stairs and she runs (top speed) to get them and bring them back. Again, she will do this until she is out of breath. If only I could work out that hard... I’d probably be thinner.

Figure 3.5: An example story on the topic of Pet.

generate tellings that vary the personality of the teller (introverted vs. extraverted) [83]. An example original monologic personal narrative about having cats as pets is shown in Figure 3.5. The human-generated dialog corresponding to Figure 3.5 is shown in Figure 3.6.

The SDG corpus includes 50 dialogs generated by human annotators, gesture annotations on human generated dialogs, videos of story dialogs generated from this representation that vary the introversion and extraversion of the animated agents, video clips of each gesture used in the gesture annotations, and annotations of the original personal narratives with a deep representation of story called a Story Intention Graph or SIG [49, 50, 52]. We expect this corpus to be a useful resource for researchers interested in natural language generation, intelligent virtual agents, generation of nonverbal behavior, and story and narrative representations.

Section 3.5.1 provides an overview of the original corpus of monologic personal narrative blog posts. Section 3.5.2 describes how the human annotators generated dialogs

Pet Story Dialog	
A1:	I have always felt like I was a dog person but our two cats are great. They are much more low maintenance than dogs are.
B1:	Yeah, I'm really glad we got our first one at a no-kill shelter.
A2:	I had wanted a little kitty, but the only baby kitten they had scratched the crap out of me the minute I picked it up so that was a big "NO".
B2:	Well, the no-kill shelter also had what they called "teenagers", which were cats around four to six months old... A bit bigger than the little kitties.
A3:	Oh yeah, I saw those "teenagers". They weren't exactly adults, but they were a bit bigger than the little kittens.
B3:	Yeah one of them really stood out to me then - mostly because she jumped up on a shelf behind us and smacked me in the head with her paw.
A4:	Yeah, we definitely had a winner!
B4:	I had no idea how much personality a cat can have. Our first kitty loves playing. She will play until she is out of breath.
A5:	Yeah, and then after playing for a long time she likes to look at you like she's saying, "Just give me a minute, I'll get my breath back and be good to go."
B5:	Sometimes I wish I had that much enthusiasm for anything in my life.
A6:	Yeah, me too. Man, she has so much enthusiasm for chasing string too! To her it's the best thing ever. Well ok, maybe it runs a close second to hair scrunchies.
B6:	Oh I love playing fetch with her with hair scrunchies!
A7:	Yeah, you can just throw the scrunchies down the stairs and she runs at top speed to fetch them. And she always does this until she's out of breath!
B7:	If only I could work out that hard before I was out of breath... I'd probably be thinner.

Figure 3.6: Manually constructed dialog from the pet story in Figure 3.5.

from the personal narratives. Section 3.5.3 describes the gesture annotation process and provides more details of the gesture library.

3.5.1 Personal Narrative Monologic Corpus

We selected 50 personal stories from a subset of personal narratives extracted from the corpus of blogs included in the ICWSM 2010 dataset challenge [72, 92]. We manually selected stories that are suitable for retelling as dialogs, i.e. where the events that are discussed in the stories could have been experienced by more than one person. The

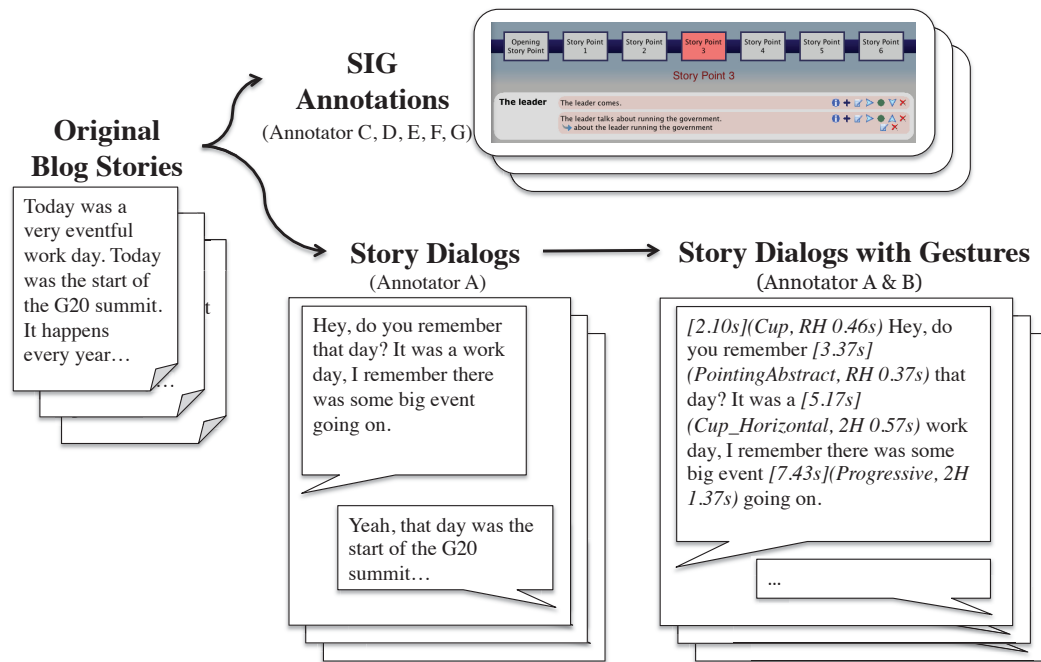


Figure 3.7: Overview of the SDG corpus.

story topics include camping, holidays, gardening, storms, and parties. Table 3.1 shows the number of stories in each topic. Each story ranges from 174 words to 410 words. A sample story about pets is shown in Figure 3.5 and another story about being present at a protest is shown in Figure 3.8.

Although we are not making use of the Story Intention Graphs (SIGs) yet to automatically produce dialogs from their monologic representations, we have annotated each of the 50 stories with their SIG using the freely available annotation tool Scheherazade [51].

More description of the Scheherazade annotation and the resulting SIG representation is provided in our companion paper [109]. Our approach builds on the DramaBank language resource, a collection classic stories that also utilize the SIG representation [49, 50, 52], but the SIG formalism has not previously been used for the purpose of automatically generating animated dialogs, and this is one of the first uses of the SIG

Protest Story
<p>Today was a very eventful work day. Today was the start of the G20 summit. It happens every year and it is where 20 of the leaders of the world come together to talk about how to run their governments effectively and what not. Since there are so many leaders coming together their are going to be a lot of people who have different views on how to run the government they follow so they protest. There was a protest that happened along the street where I work and at first it looked peaceful until a bunch of people started rebelling and creating a riot. Police cars were burned and things were thrown at cops. Police were in full riot gear to alleviate the violence. As things got worse tear gas and bean bag bullets were fired at the rioters while they smash windows of stores. And this all happened right in front of my store which was kind of scary but it was kind of interesting since I've never seen a riot before.</p>

Figure 3.8: An example story on the topic of Protest.

Topic	Number of Stories
Camping	3
Holiday	5
Gardening	7
Party	10
Pet	3
Sports	4
Travel	7
Weather Conditions	3
Other	13

Table 3.1: Distribution of story topics in the SDG corpus.

on personal narratives [110, 111].

DramaBank provides a symbolic annotation tool for stories called Scheherazade that automatically produces the SIG as a result of the annotation. Every annotation involves: (1) identifying key entities that function as characters and props in the story; and (2) modeling events and stative propositions and arranging them in a timeline. Currently our Scheherazade annotations only contain the timeline layers. Figure 3.9 shows a part of the graph structure of Scheherazade annotations for the protest story in Figure 3.8. In timeline layer, the first column shows phrases from the original blog

story; the second column shows events that corresponds to the original phrases, modeled by using pre-defined entities.

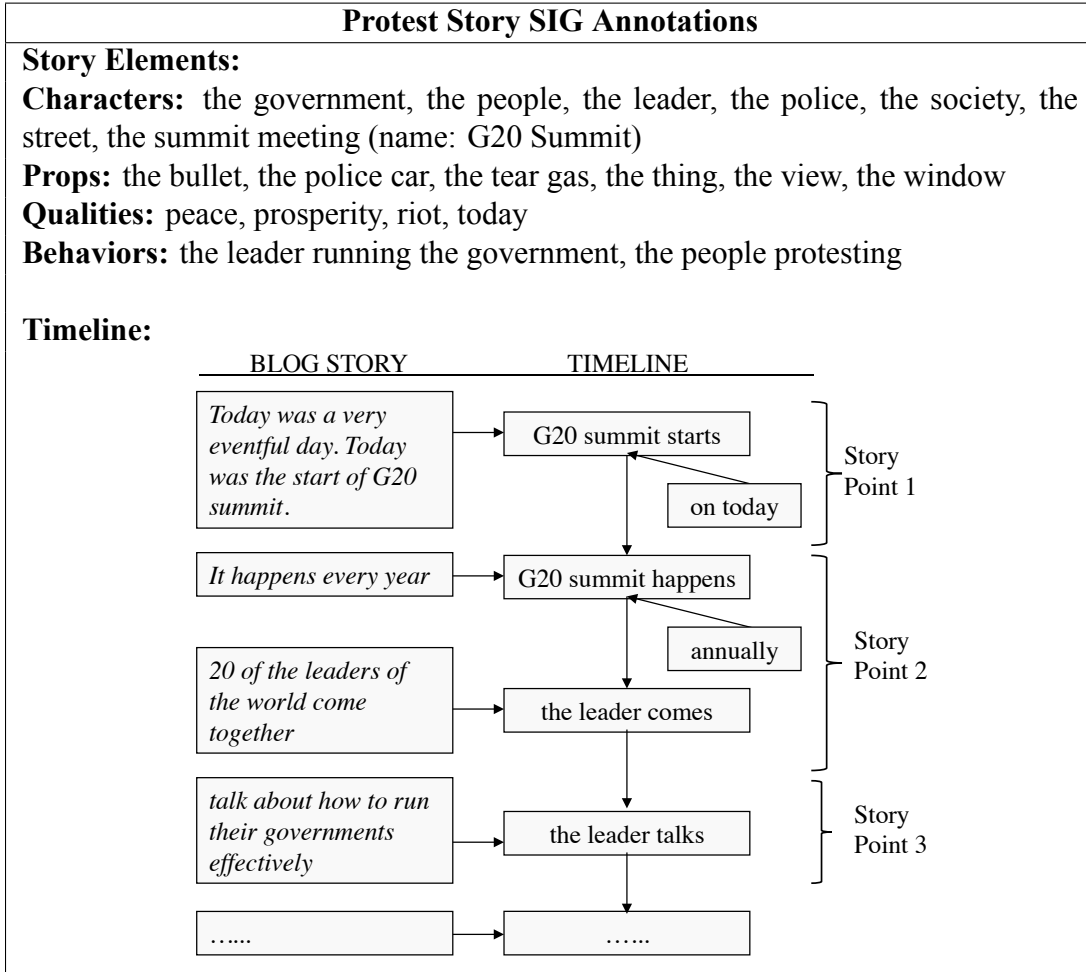


Figure 3.9: Part of Scheherazade annotations for the protest story in Figure 3.8.

3.5.2 Dialog Annotations

In the wild, stories are told conversationally in social settings, and in general research has shown that conveying information in the form of a dialog is more engaging and memorable [6, 102, 126, 168]. For example, Craig et al. show that students demonstrate better recall and ask significantly more questions after hearing a dialog between

a virtual tutor and tutee than a tutor monologue with identical contents [42].

Protest Story Dialog	
A1:	Hey, do you remember that day? It was a work day, I remember there was some big event going on.
B1:	Yeah, that day was the start of the G20 summit. It's an event that happens every year.
A2:	Oh yeah, right, it's that meeting where 20 of the leaders of the world come together. They talk about how to run their governments effectively.
B2:	Yeah, exactly. There were many leaders coming together. They had some pretty different ideas about what's the best way to run a government.
A3:	And the people who follow the governments also have different ideas. Whenever world leaders meet, there will be protesters expressing different opinions. I remember the protest that happened just along the street where we work.
B3:	It looked peaceful at the beginning....
A4:	Right, until a bunch of people started rebelling and creating a riot.
B4:	Oh my gosh, it was such a riot, police cars were burned, and things were thrown at cops.
A5:	Police were in full riot gear to stop the violence.
B5:	Yeah, they were. When things got worse, the protesters smashed the windows of stores.
A6:	Uh huh. And then police fired tear gas and bean bag bullets.
B6:	That's right, tear gas and bean bag bullets... It all happened right in front of our store.
A7:	That's so scary.
B7:	It was kind of scary, but I had never seen a riot before, so it was kind of interesting for me.

Figure 3.10: Manually constructed dialog from the protest story in Figure 3.8.

Our goal with the dialog annotations is to generate a natural dialog from the original monologic text, with the long term aim of using these human-generated dialogs to guide the development of an automatic monologue-to-dialog generation engine. First, one trained annotator writes all the stories into two-person dialogs as illustrated in Figure 3.6 and Figure 3.10. The goal of the annotation process is to create natural dialogs by: (1) adding oral language such as acknowledgements and discourse markers, and breaking longer sentences into shorter ones; (2) adding repetitions and confirmations between

speakers, which are common in human dialog, and can also be used as locations for inserting gesture adaptation; (3) re-using phrases in the original story, but changing or deleting content that doesn't fit the storytelling setting; (4) making the story sound like the two speakers experience the event together.

The audio for each of the agents is then produced by running the human-generated dialog turns through the AT&T Text to Speech engine, one turn at a time. Speaker A uses AT&T's female voice Crystal, Speaker B uses AT&T's male voice Mike. At the beginning of the audio for each story, a two-second blank audio is inserted in order to give the audience time to prepare to listen to the story before it starts. The timeline for the audio is then used to annotate the beginning timestamps for the gestures: as shown in Figures 3.14 and 3.13, in front of every gesture, a time is shown inside a pair of square brackets to indicate the beginning time of this gesture stroke. Thus, any change in the wording of the dialogic telling requires regenerating the audio and relabeling the gesture placement. In our envisioned future Monologue-to-Dialog generator, both gesture placement and gesture form would be automatically determined.

3.5.3 Gesture Annotations

We annotate the dialogs with a gesture tag that specifies the gesture form (e.g. a pointing gesture or a conduit gesture). We also specify gesture start times, but do not specify stylistic variations that can be applied to particular gestures (e.g. gesture expanse, height and speed). Each story has two versions of annotations done by different annotators. The annotators are advised to insert a gesture when the dialog introduces new concepts, and add gesture adaptation (mimicry) when there are repetitions or confirmations in the dialog. The decisions of where to insert a gesture and which gesture to insert are mainly subjective.

We use gestures from a database of 271 motion captured gestures, including metaphoric,

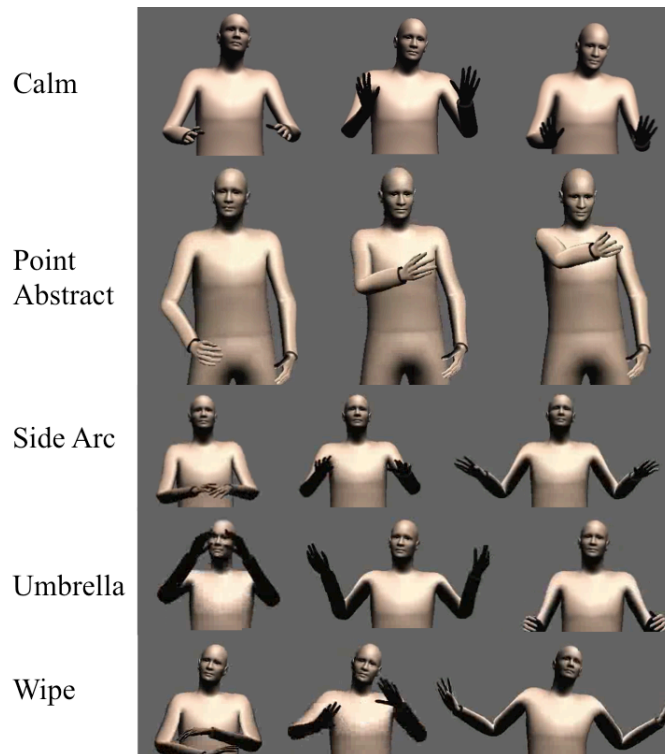


Figure 3.11: A subset of the 271 gestures in our gesture library that can be used in annotation and produced in the animations.

iconic, deictic and beat gestures. The videos of these gestures are included in the corpus. Gesture capture subjects were all native English speakers, but given a database of similarly categorized gestures from a different culture, our corpus could be used to generate culture specific performances of these stories [130, 150].

Figure 3.11 provides samples of the range of gestures in the gesture database, and Figure 3.12 illustrates how every gesture can be generated to include up to 4 phases [91, 95]:

- prep: move arms from default resting position or the end point of the last gesture to the start position of the stroke
- stroke: perform the movement that conveys most of the gesture’s meaning

- hold: remain at the final position in the stroke
- retract: move arms from the previous position to a default resting position

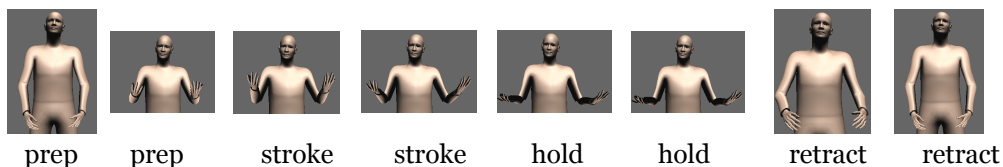


Figure 3.12: Prep, stroke, hold and retract phases of gesture “Cup_Horizontal”.

Figure 3.14 shows the first 5 turns of the protest story annotated with gestures from the library. Figure 3.13 shows the first 6 turns of the pet story annotated with gestures. The timing information of the gestures comes from the TTS audio timeline. Each gesture annotation contains information in the following format: ([gesture stroke begin time]gesture name, hand use[stroke duration]). For example, in the first gesture “([1.90s]Cup, RH [0.46s])”, gesture stroke begins at 1.9 seconds of the dialog audio, it is a “Cup” gesture, uses the right hand, and the gesture stroke lasts 0.46 seconds. Research has shown that people prefer gestures occurring earlier than the accompanying speech [181]. Thus, in this annotation, a gesture stroke is positioned 0.2 seconds before the beginning of the gesture’s following word. For example, the first word after gesture “Cup” is “Hey”, it begins at 2.1 seconds, then the stroke of gesture “Cup” begins at 1.9 seconds. Each story dialog has two versions of gesture annotations from different annotators.


Our gesture annotation does not specify stylistic variations that should be applied to particular gestures. We used custom animation software that can vary the amplitude, direction and speed in order to affect stylistic change. The default gesture annotation frequency is designed for extraverts, with a gesture rate of 1 - 3 gestures per sentence. For an introverted agent, a lower gesture rate can be achieved by removing some of the

gestures. In this way, both speakers' gestural performance can vary from introverted to extraverted using the entire scale of parameter values for every parameter.

In addition, we can also vary gestural adaptation in the annotation. For example, in extravert & extravert gestural adaptation (based on the model and data described in [129, 176, 177]), two extraverts move together towards a more extraverted personality. Gesture rate is increased by adding extra gestures (marked with an asterisk "*"). Specific gestures are copied as part of adaptation, especially when the co-telling involves repetition and confirmation. Gestures in bold indicate copying of gesture form (adaptation), gestures after the slash "/" are non-adapted.

Combined with personality variations for gestures described in the previous paragraph, it is possible to produce combinations of two agents with any level of extraversion engaged in a conversation with or without gestural adaptation.

Pet Story w/Gestures



A1: I ([2.31s]Dismiss1_AntonyTired, RH[0.54s]) have always felt like I was a dog person but our two cats are great. They are ([7.50s]Cup5_ShakesTired2_FingerSkel, RH[0.23s]) much more low maintenance than dogs are.

B1: Yeah, I'm really glad we got ([12.44s]SideOut1, 2H[0.29s]) our first one at a no-kill shelter.

A2: I had wanted a little kitty, ([16.44s]SweepSide1, RH[0.35s]) but the only baby kitten they had ([17.90s]BackHandBeats_Tight, 2H[0.37s]) scratched the crap out of me the minute I ([19.75s]Shovel, 2H[0.17s]) picked it up so that was a big "NO".

B2: Well, ([22.69s]Cup_Down_alt, RH[0.21s]) the no-kill shelter also had what they called ([24.70s]Cup_Horizontal , RH[0.57s]) "teenagers", which were cats around four to six months old... a bit bigger than the ([28.79s]ShyCalmShake, 2H [0.56s]) little kitties.

A3: Oh yeah, I saw those ([31.24s]Cup_Horizontal, RH[0.57s]) / SideOut_vibrate, 2H[0.33s]) "teenagers". They *([32.75s]HandToChest_Vibrate, 2H[0.41s]) weren't exactly adults, but they were a bit ([35.78s]ShyCalmShake, 2H [0.56s]) /SideOut1, LH[0.29s]) bigger than the little kittens.

B3: Yeah ([36.42s]SideOut_vibrate, 2H[0.33s]/ PointingHere, RH[0.36]) one of them really stood out to me then - *([38.92s]Cup11_ShakesTired2_FingerSkel, RH[0.30s]) mostly because she *([40.02s]CupBeats_Small, 2H[0.37s]) jumped up on a shelf behind us and ([42.20s]SideOut1, LH[0.29s]/ GraspSmall, RH[0.18s]) smacked me in the head with her paw.

A4:

Figure 3.13: Pet dialog with gesture annotations. Pictures show the first 6 gestures in the dialog.

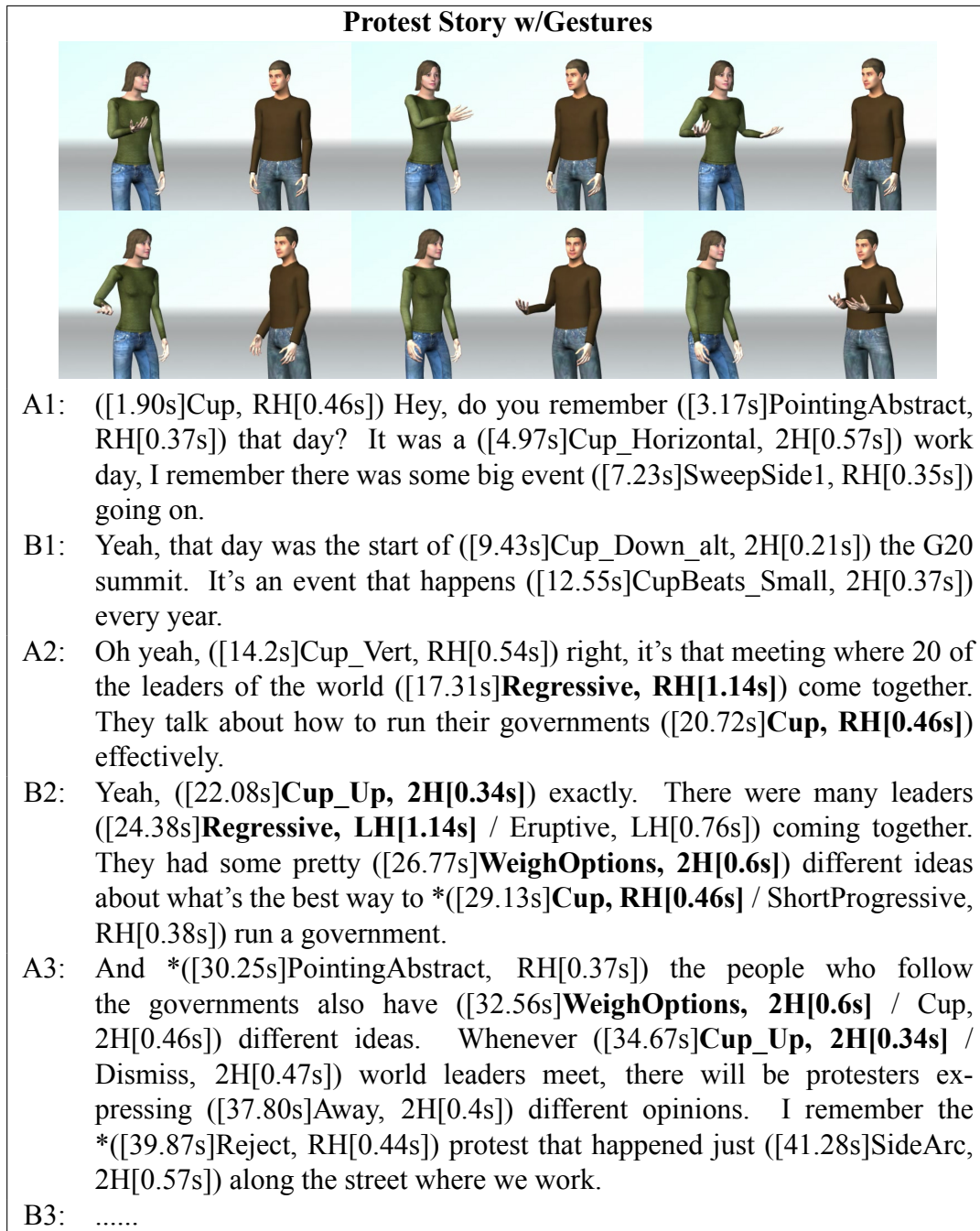


Figure 3.14: Protest dialog with gesture annotations. Pictures show the first 6 gestures in the dialog.

Chapter 4

Linguistic Adaptation

A number of studies have shown that adaptation is highly correlated with task success [154], dialog naturalness [132], user satisfaction [149] and learning gains [183]. To date however, the technical challenges of getting a dialog system to dynamically adapt to the user has made it difficult to test the potential benefits of user adaptation. Moreover, adaptation is not simple mimicking. As shown in Table 1.2 and 1.3 in Chapter 1, too much mimicking could incur negative user perception. The process of adaptation takes place under other special constraints, e.g. providing coherent turns in conversation and expressing one's own personalities. When implementing dynamic adaptation in natural language generation, we need to take these constraints into account. A sensible approach is to measure adaptation in human dialogs, and use models produced by these measures to control adaptation behaviors. In this chapter, we first present our exploratory experiment showing that adapting to different linguistic features results in different perception of friendliness and naturalness. On the basis of these results, we propose an adaptation measure that aims to reflect different adaptation models of feature sets that describe certain linguistic styles such as personality traits. We then build and compare adaptation models using four real human dialog corpora.

4.1 Implementing and Evaluation Linguistic Adaptation

In this section, we describe our exploratory study on testing human perceptions on linguistic adaptation to various feature sets. Section 4.1.1 introduces the architecture of Personage-primed, a natural language generation system that can dynamically adapt to the dialog context. Section 4.1.2 describes the evaluation method for language generation results. The evaluation results in Section 4.1.3 shows that some types of adaptation have a positive effect on the friendliness of system utterances, while other types of adaptation positively effect perceptions of naturalness.

4.1.1 Personage-primed

Personage-primed is a natural language generation system that can dynamically adapt to the dialog context. This work is carried out in the context of the Skipper project whose aim is to study “adaptation in the wild” in pedestrian direction giving dialogs. As part of the Skipper project, the ArtWalk corpus [106] was collected. To our knowledge, this corpus collection is the first experiment to show that adaptation actually occurs in the context of a real task, while people are out in the world, navigating a natural terrain.

Using insights from analysis of ArtWalk, we developed Personage-primed, an extension of the Personage spoken language generator that adapts dynamically to user utterances as represented in the discourse context. The discourse model of Personage-Primed keeps track of user utterance choices in referring expressions, discourse cues, location names, prepositions, and syntactic forms. This allows adaptation to occur at many different stages of the language generation process, and lets Personage-primed produce tens of different possible utterances in a given context.

Personage-primed is an extension of the parameterizable language generator Personage [116]. Personage is capable of producing a wider range of linguistic variation than traditional template-based language generation systems because it dynamically mod-

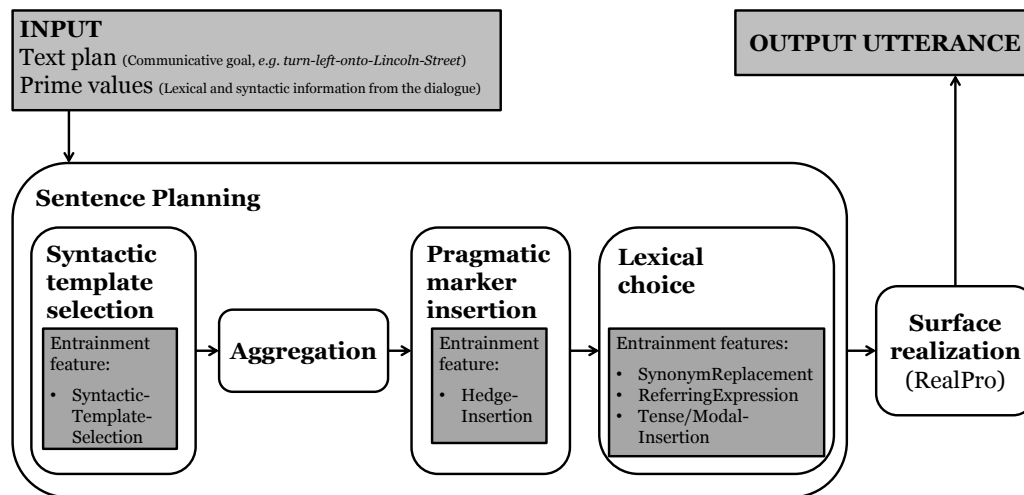


Figure 4.1: The architecture of Personage-primed.

ifies high level representations of the utterances and implements external lexical resources including VerbOcean [36] and WordNet [56]. VerbOcean is a collection of semantic relations between pairs of verbs. For example, *marry* and *divorce* have a relation of Happens-before; *produce* and *create* have a relation of Similarity (similar events). WordNet is an English lexical database that groups nouns, verbs, adjectives and adverbs into sets of cognitive synonyms (synsets), each expressing a distinct concept. For example, $\{communicate, talk, whisper\}$. Synsets are linked by semantic relations, such as hypernym and hyponym relations.

The architecture of Personage-primed is shown in Fig. 4.1. We developed Personage-primed for the pedestrian direction giving domain because our assumption was that “walking around” would be a good context for testing “adaptation in the wild”. Following directions naturally introduces delays between task relevant utterances as the follower navigates an actual landscape. At the same time, pedestrian directions can easily support a range of experimental manipulations. We chose the ArtWalk context, of asking users to find and take pictures of public art, because we assumed that there

would not be known referring expressions for these artworks, and that we should therefore be able to elicit adaptation to referring expressions, as in earlier work on adaptation. However, we also discovered in the corpus that adaptation seems to occur not just to referring expressions, but also to a whole range of lexical and syntactic choices in dialog. Thus, we designed the Personage-primed generator to have the capability of adapting on any one of these generation choices.

Instruction/Statement	Example Utterance
confirm (yes)	<i>That's correct.</i>
turn-DIR	<i>Make a right turn.</i>
turn-DIR-onto-STREET	<i>At Cedar Street, make a right.</i>
continue-on-STREET	<i>Keep going straight down Cedar Street.</i>
continue-on-STREET-for-NUM-blocks	<i>You're going to follow Cedar Street for three more blocks.</i>
go-along-STREET	<i>Head down Cedar Street.</i>
go-NUM-blocks-on-STREET	<i>Walk five blocks along Cedar Street.</i>
go-to-LOC-from-LOC	<i>... you will walk from the bookstore to the coffeeshop.</i>
go-to-LOC-from-STREET-and-STREET	<i>From the corner of Cedar Street and Elm, walk towards Lulu's Coffeeshop.</i>
go-back-to-LOC	<i>Go back to the bookstore.</i>
you-pass-LOC	<i>After you pass the bookstore...</i>
LOC-is-on-the-DIR	<i>The bookshop is on the left-hand side.</i>
arrive-at-LOC	<i>When you get to Lincoln Street...</i>

Figure 4.2: Instructions and statements supported in Personage-primed.

As shown in the architecture diagram in Fig. 4.1, Personage supports parameterization of an output utterance via modules for syntactic template selection, pragmatic marker insertion, and lexical choice. In Personage-primed, the values of the module parameters are controlled by reference to a set of prime values, which represent the content and linguistic information of the dialog context, i.e. the system's output is generated to adapt with the given dialog context. For example, as shown in Fig. 4.1, lexical choice is further refined into parameters and corresponding prime values for referring expressions, synonyms for nouns and verbs, and the tense/modality to be generated in

Follower: Okay, now I'm at the corner of Cedar Street and Elm, so should I head toward the clock tower from here?
Discourse Context Primed Values:
Prepositions: at, toward, from
Noun: I, corner, here
Tense: present
Modals: should
Verbs: am, head
Place names: cedar street, elm, the clock tower
Hedges: Okay, So, Now
Syntax: (VP PP PP)
Director: confirm + go-to-clocktower

Figure 4.3: Sample Discourse Model Representation.

the system utterance.

Fig. 4.3 provides an example of how the context is represented by primed values in the discourse model. The dialog context (prime utterance, or the previous utterance by a different speaker) spoken by the Follower is “Okay, now I’m at the corner of Cedar Street and Elm, so should I head toward the clock tower from here?” The corresponding prime values for this dialog context are listed below. For example, prepositions *at*, *toward*, and *from*. In response to the dialog context, the Director confirms that the Follower should go to the clocktower.

To our knowledge, Personage-primed is the first dialog generator to have the capability of adapting on any of the values shown in the discourse model in Fig. 4.3 and it does so by explicitly manipulating parameters that we have added to Personage-primed. The prime values contain lexical and syntactic information from the dialog to which the generated utterance will be adapted. An utterance can be produced to adapt to **all** of the adaptation prime values or **none** of them, or **any combination**, depending on the adaptation model in effect when the utterance is dynamically generated by the dialog system. Our goal is to explore which combinations have an effect on user perceptions of system naturalness and friendliness.

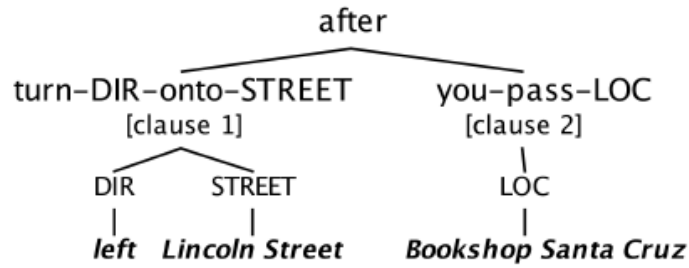


Figure 4.4: Example text plan tree.

Input

The input to Personage-primed consists of a text plan and a set of adaptation target values referred to as the prime values as illustrated in Fig. 4.3. The text plan is a high level semantic representation representing the communicative goal of the desired output utterances. Each text plan contains either a single instruction or a compound instruction. A compound instruction consists of two clauses joined by a temporal relation, such as *after*, *until* or *once*. An example text plan for a compound instruction is shown in Fig. 4.4. Personage-primed currently supports 13 unique instructions and statements for the walking directions domain.

Syntactic Template Selection

While the text plan contains all the information regarding what will be communicated, the sentence planning pipeline controls how that information is conveyed. See Fig. 4.1. Syntactic template selection is the first phase of sentence planning: its goal is to select the most appropriate syntactic form for the instruction(s) in the text plan. Keeping track of user choices in syntactic form is needed in order to produce syntactic adaptation in dialog [17, 19, 151]. If a navigation dialog included the question, *From*

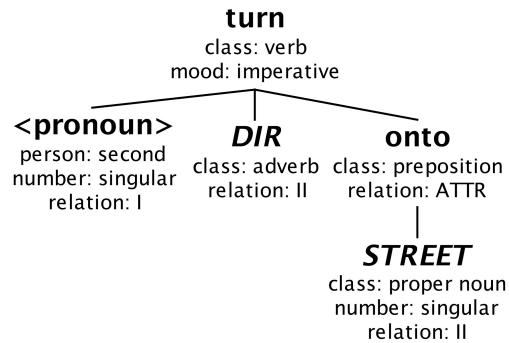


Figure 4.5: DSyntS for the instruction *turn-DIR-onto-STREET*. Relation I: the component is the subject of the parent; relation II: the component is the direct object of the parent; relation ATTR: the component is a modifier(adjective/prepositional phrase) of the parent.

here where should I go to next? a response with syntactic adaptation would be phrased in a similar way, such as *From where you are, walk to Pacific Avenue and then make a left.*

Personage-primed implements the same syntactic dependency tree representation for utterances as used in Personage [114], referred to as a Deep Syntactic Structure (DSyntS) [101, 122]. The DSyntS specifies the dependency relation between the different components of a sentence. An example DSyntS is shown in Fig. 4.5. Each instruction and statement has an associated DSyntS List, which is a collection of semantically equivalent DSyntS with different syntactic structure. In order to produce syntactic adaptation, Personage-primed finds the associated DSyntS List for each instruction in the text plan. It then uses the prime values to select the DSyntS that best matches the lexical and syntactic information. The DSyntS with the highest number of features matching the prime values is designated as the best match. If no best match is found, the default DSyntS is assigned to the instruction.

Aggregation

For compound instructions that contain a temporal relation (such as *after* or *once*), the aggregation component integrates each DSyntS into a larger syntactic structure. For most temporal relations, the clauses can be joined in two ways: e.g. **After** *you pass...*, *turn left onto...* or *Turn left onto...after* *you pass...*. Currently, there is no adaptation for aggregation operations in Personage-primed, however in the future, it would be possible to prime particular rhetorical relations and then control the aggregation component as we do other components.

Pragmatic Marker Insertion

Pragmatic markers, or discourse markers, are elements of spontaneous speech that do not necessarily contribute to the semantic content of a discourse, but serve various pragmatic or social functions. Some common examples include *so*, *okay*, *like*, *umm*, *you know* and *yeah* (not in response to a yes/no question). Research on spontaneous speech has shown that discourse markers not only make a conversation sound more natural but can also serve to highlight or qualify content, help listener's follow a speaker's train of thought, and create a meaningful transition from one utterance to the next [61, 143]. Discourse markers are especially prevalent in task-oriented dialog.

In Personage-primed, sample prime values are shown in Fig. 4.3, e.g. *Okay*, *Now*, *So*. The module for pragmatic marker insertion in Personage-primed will insert up to three of the pragmatic markers found in the prime values.¹ A pragmatic marker is inserted only if one of the insertion points associated with the marker is present in the DSyntS.

¹While use of pragmatic markers varies according to individual personalities, three was chosen to be a maximum value as it reflected an approximation of average use.

Synonym Selection

Synonym selection is a lexical choice operation that checks every verb and preposition in the current utterance and if there exists a synonym in the prime values, the prime synonym replaces the existing verb or preposition. See Fig. 4.1 and the primed context representation in Fig. 4.3. The system does not currently adapt to nouns because most nouns within the walking directions domain are referring expressions, such as *downtown*, *Pacific Avenue*, etc. Adaptation to referring expressions is handled with a separate operation. In addition, many common nouns in the directions domain do not have appropriate synonyms, such as directions like *right* and *left*.

Referring Expression Selection

Referring expression selection is a lexical choice operation that checks every proper noun within the current utterance for a semantic match in the prime values. This operation requires an existing database of referring expressions and their possible variations. For this work we manually created a map from each referring expression to its list of variations. For example, the destination named *Bookshop Santa Cruz* is an entry in the referring expression map with the corresponding list of alternative referring expressions *{bookshop, the bookshop, Santa Cruz bookshop}*.

This operation also accounts for a referring expression form that is commonly found in navigation dialogs, i.e. referencing street names without the street suffix. If one conversant refers to a street as *Pacific* instead of *Pacific Avenue*, it is common for the other participant to do so as well. This step of the referring expression operation checks the prime values for any single instance of this shortened form and modifies all instances of street names in the current utterance to adapt with this stylistic choice.

Tense transformation and modal insertion

Tense transformation and modal insertion are a final set of lexical choice operations that adapt on primed values for tense and modals. If there exists an explicit use of a particular tense or a modal in the prime values, the current utterance is modified to adapt. The most common tenses used for giving directions in the navigation domain are present, future, and simple future. While followers do use past tense to confirm the completion of an action, it is not common for directors to use it. However, the modals *should*, *can* and *might* are commonly found in navigation dialogs. Followers will express uncertainty with questions such as *Should I stay on Pacific Avenue?*. The corresponding director responses sometimes adapt with this lexical addition with confirming responses such as *Yes, you should stay on Pacific Avenue for three more blocks*.

4.1.2 Methodology

In a pilot experiment, we ask naive participants from Mechanical Turk to score three utterances in the same context for naturalness: a generated adapted utterances, generated default (non-adapted) utterance and a human utterance which has the same meaning but which is not from the same context. We hypothesized that the adapted generated utterances would be perceived as more natural than the default generated utterances. But the experimental results (default > adapted > human) did not confirm our hypothesis.

In the pilot, every generated adapted utterance (target) adapts to some of the prime features of its previous (prime) utterance. However, there is very little evidence of what people actually do in human-human conversation, and to our knowledge, no previous work has tested whether mimicking all the linguistic features of a conversational partner is natural or whether some kinds of adaptation are dispreferred. In the experiment, we aimed to systematically explore whether there are clear preferences in types of adaptation by overgenerating possible outputs that adapt on different combinations of prime

5. Please order the utterances based on friendliness.

Group No: 41

Dialogue:

F: Okay I'm at Locust.

D: Okay if you're on Locust Street just go to your right then.

F: To my right, like towards Pacific?

D: [following options] *

Drag items from the left-hand list into the right-hand list to order them.

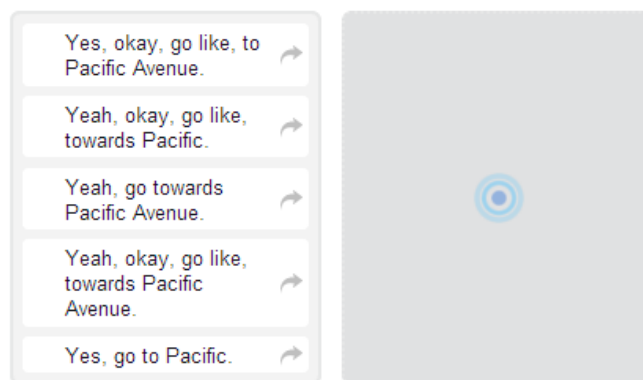


Figure 4.6: An example question from Experiment 2.

features. We sample among all the possibilities for adaptation, and our task becomes simply to find out which adaptation combinations are the best. Our earlier work used a similar overgenerate and rank experimental paradigm for collecting data to train a statistical language generator [116, 178].

Ten dialog excerpts are used as context, in which a director (D) is instructing a follower (F) how to navigate to a destination on foot. The dialog excerpts were taken from the Art Walk corpus [106] and were slightly modified to isolate certain priming values. Following the excerpt, participants were presented with options for what the director could say next. Using overgeneration, together with a generated default utterance and a random human utterance, each director response results in 5 to 22 different variations. Having all 22 utterances in one item and asking participants to rank them all does not seem to be a well-defined experimental task. Therefore, each item of the experiment survey consists of 5 possible utterances in a particular context, selected so that each

possible generated utterance for a particular context appears at least twice across all the survey items. This results in a total of 51 items distributed across 10 surveys. An example item is shown in Fig. 4.6.

In one version of the experiment, participants were asked to rank the possible system utterances based on their naturalness from high to low. In another version, participants were asked to rank the possible system utterances based on their friendliness. This is because default utterances received the highest score for naturalness in the pilot experiment. We hypothesized that one possible explanation of these results was that a director’s utterance is considered “natural” when it is concise and clear, and that people may be accustomed to the type of instructions used in current in-vehicle GPS navigation systems. We hypothesized that perceptions of friendliness might be a better probe for adaptation. We hired three judges trained in linguistics to finish all the surveys over a period of two weeks, doing two surveys per day at most.

4.1.3 Experiments and Results

We use machine learning to evaluate our experiment results. Each generated utterance is represented by the parameters used to generate that utterances. Therefore, each utterance is represented by 7 features: five are adaptation features: `SynonymReplacement`, `ReferringExpression`, `HedgeInsertion`, `SyntacticTemplateSelection`, and `Tense/Modal Insertion` as described above in Sec. 4.1.1; two features represent whether the utterance was a Random Human Utterance or a Default utterance.

To evaluate the effects of the different parameters, we train two types of models for evaluation: multivariate linear regression models and decision tree models. In the regression model, the dependent variable is the average ranking scores across all three judges across all duplicate instances of the utterance (each utterance appears at least twice across all the surveys). In one ranking question, there are 5 utterances. If an

```

Naturalness_score =
    1.6171 * Tense/Modal Insertion +
    1.2486 * ReferringExpression +
    -1.0809 * HedgeInsertion +
    5.5226

Friendliness_score =
    2.9449 * HedgeInsertion +
    2.3239 * RandomHumanUtterance +
    1.4023 * Tense/Modal Insertion +
    0.7223 * SyntacticTemplateSelection +
    3.6261

```

Figure 4.7: Regression models.

utterance is ranked first in a question, the score of the object is 4. If an utterance is ranked last in a question, the score of the object is 0. We use the sum of the scores from all annotators as the label for an utterance. There are 3 annotators in total, so the scores range from 0 to 12. Since an utterance appears at least twice among all surveys, it will have 2 or more scores. We simply take an average of these scores. The features in regression model are 0/1 features, where a value of 1 indicates that the feature positively affected the generation of the utterance, whereas a value of 0 means this feature was not used in generating the utterance.

We use Linear Regression in Weka 3.6.1 with 10-fold cross validation. Fig. 4.7 shows the regression models for naturalness and friendliness. In the naturalness model, the correlation coefficient is 0.31, relative absolute error is 96.33% and root relative squared error is 94.71%. ReferringExpression and Tense/Modal Insertion both have positive weights, which means adapting on these features increases the perception of the naturalness of the utterance. HedgeInsertion has the only negative weight in the model. In contrast, the friendliness model provides a better fit with a correlation coefficient of 0.52, relative absolute error is 81.22% and root relative squared error of 85.76%. Surprisingly, HedgeInsertion has the highest positive weight in the model, suggesting

that more hedging leads to perceptions that the system is more friendly. RandomHumanUtterance has the second highest positive weight.

Feature Value	Meaning
NOMATCH	feature exist in context AND not adapted
MATCHPLUS	feature exist in context AND adapted
MATCHMINUS	feature doesn't exist in context AND not adapted
DEFAULT	generated non-adapted utterance
RAMDOMHUMAN	random human utterance
NULL	for features "Default" and "RandomHumanUtterance", if this feature doesn't exist in the utterance, use NULL

Table 4.1: Possible values for features used in decision tree model.

In the decision tree model, the dependent variables are identical to those used for the regression model. However, here we distinguish more values for each features rather than making them binary features. As shown in Table 4.1, a feature may have any of 6 possibilities. Recall that there are certain features in previous dialog context (10 given dialog excerpts) that the following utterance can adapt to. Since an utterance only adapts to the feature if the feature is present in context, the combination "feature is adapted AND context doesn't have feature" cannot occur.

Decision trees are trained using the REPTree package in Weka 3.6.1. We use the whole evaluation data as both training set and test set, and disable pruning to intentionally force the decision tree to overfit the data. Fig. 4.8 and Fig. 4.9 show the decision trees for naturalness and friendliness. In the leaf nodes, the first number is the predicted score. The numbers in the parentheses are (number of examples in this leaf / number of misclassified examples on average).

In the naturalness model, correlation coefficient is 0.35, relative absolute error is 94.81% and root relative squared error is 93.77%. When previous context provides a prime value for ReferringExpression, and the utterance adapted on the referring expression (MATCHPLUS), we get the highest score with the highest number of examples.

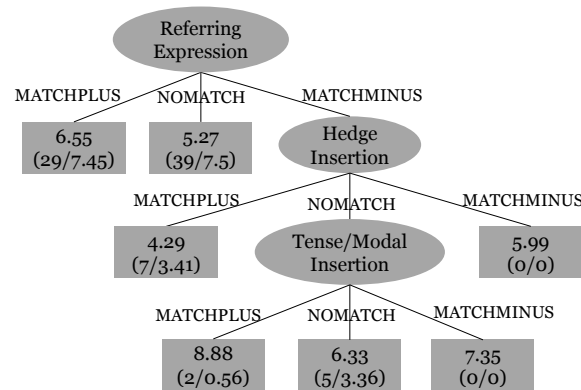


Figure 4.8: Decision tree model for Experiment 2 naturalness.

If the utterance doesn't adapt on referring expressions (NOMATCH), the scores are relatively lower. If the previous context doesn't provide a prime value for Referring-Expression, then HedgeInsertion primes and utterance features are considered by the model. Similar to the regression model in Fig. 4.7, hedging is a negative factor in naturalness. Generally, utterances that adapt on HedgeInsertion (MATCHPLUS) have a lower naturalness score than utterances that don't (NOMATCH).

In the friendliness model, correlation coefficient is 0.63, relative absolute error is 74.95% and root relative squared error is 77.50%. These results also indicate that hedging affects perceptions of friendliness as in the regression model shown in Fig. 4.7. When the dialog context provides a prime value for HedgeInsertion, and the utterance adapts for hedging (MATCHPLUS), the resulting friendliness score is the highest with the highest number of examples. If the utterance did not adapt on hedging (NOMATCH) even though a prime value was available, then SyntacticTemplateSelection is considered by the model. Generally, utterances that adapt on SyntacticTemplate Selection have higher scores.

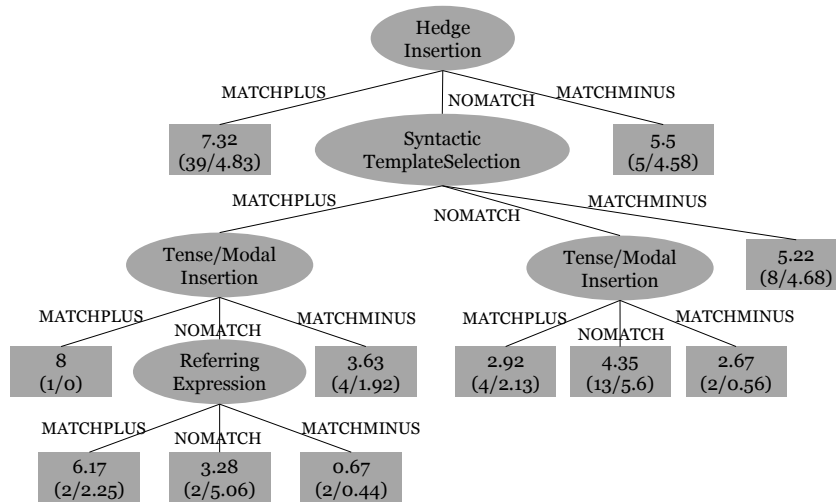


Figure 4.9: Decision tree model for Experiment 2 friendliness.

4.1.4 Discussion

This section presents an experiment based on Personage-primed, an extended version of Personage that can dynamically adapt to the dialogue context. We show that some types of adaptation have a positive effect on the friendliness of system utterances, while other types of adaptation positively effect perceptions of naturalness.

Previous work testing the benefits of adaptation have been measured in different contexts, such as whether adaptation in human-human dialogue predicts success [154]. Much of the previous work on human-computer dialogue has examined whether the human adapted to the computer rather than vice versa. Our work contributes to the limited amount of previous work on adaptive generation using different computational methods for generation. De Jong, Theune, and Hofs presents an approach that focuses on affective language use for aligning specifically to user’s politeness and formality [46]. Brockmann et al. illustrates a model in which alignment is simulated using word sequences alone [26]. An extension of this work in Isard, Brockmann, and Oberlander simulates both individuality and alignment in dialogue between pairs of agents with the

CrAg-2 system [87]. This system uses an over-generation and ranking approach that yields interesting results, but the underlying method has no explicit parameter control and the output has yet to be evaluated.

Perhaps most similar to our goals is the alignment-capable microplanner SPUD *prime* presented by Buschmeier, Bergmann, and Kopp [28]. SPUD *prime* is a computational model for language generation in dialogue that focuses heavily on relevant psycholinguistic and cognitive aspects of the interactive alignment model. Their system is driven by a method of activating relevant rules in a detailed contextual model according to user behavior during a dialogue. Although the underlying system seems to be capable of producing both syntactic and lexical alignment, it was evaluated only for accurate representation of lexical alignment in a corpus of dialogues from a controlled experiment.

In a field study conducted with the Let's Go system [141] however, user utterance behavior was batched to produce new system behaviors in a non-dynamic version of the system, but which however produced behaviors adapted to user behavior in the corpus collected earlier. This study suggested that system adaptation to the user was helpful in some situations, although the switch in system behavior may have confused some users. In contrast, we test a system that is capable of dynamic adaptation, but we test it in the lab with user perceptions. While this is the first study to our knowledge to be based on a generator that can produce utterances dynamically adapted to the context, in future work, we hope to be able to test dynamically produced adaptation in the field.

4.2 Measuring Linguistic Adaptation in Dialogs

In reality, adaptation doesn't always happen whenever it might be possible. As shown in Table 1.2 and 1.3 in Chapter 1, too much mimicking in a dialog can lead to negative impressions. An effective way to control adaptation is using learned adaptation

models from real data in natural language generation. Our previous exploratory experiment also shows that adapting to different linguistic features results in different style perceptions. On the basis of these results, we propose an adaptation measure that aims to reflect different adaptation models of feature sets that describe certain linguistic styles, such as personality traits. We then build and compare adaptation models using four real human dialog corpora. Section 4.2.1 presents our overall method and approach. Section 4.2.2 illustrates feature sets and feature extraction methods. We present our results in Section 4.2.3.

4.2.1 Method and Overview

Our goal is an algorithm for adaptive natural language generation (NLG) that controls the system output at each step of the dialog. Our first aim therefore is a measure of dialog adaptation that can be applied on a turn by turn basis as a dialog unfolds. For this purpose, previous measures of dialog adaptation [43, 166] have two limitations: (1) their calculation require the complete dialog, and (2) they focus on single features and do not provide a model to control the interaction of multiple parameters in a single output, while our method measures adaptation with respect to any set of features.

Measures of adaptation focus on prime-target pairs: (p, t) , in which the prime contains linguistic features that the target may adapt to. While linguistic adaptation occur beyond the next turn, we simplify the calculation by using a window size of 1 for most experiments: for every utterance in the dialog (prime), we consider the next utterance by a different speaker as the target, if any. We show the decay of adaptation with increasing window size in a separate experiment. When generating (p, t) pairs, it is possible to consider only speaker A adapting to speaker B (target=A), only speaker B adapting to speaker A (target=B), or both at the same time (target=Both). In the following definition, $FC_i(p)$ is the count of features in prime p of the i -th (p, t) pair, n is the total

number of prime-target pairs in which $FC_i(p) \neq 0$, similarly, $FC_i(p \wedge t)$ is the count of features in both prime p and target t . We define Dialog Adaptation Score (DAS) as:

$$DAS = \frac{1}{n} \sum_{i=1}^n \frac{FC_i(p \wedge t)}{FC_i(p)}$$

Within a feature set, DAS reflects the average probability that features in prime are adapted in target across all prime-target pairs in a dialog. Thus, our Dialog Adaptation Score (DAS) models adaptation with respect to feature sets, providing a whole-dialog adaptation model or a turn-by-turn adaptation model. The strength of DAS is the ability to model different classes of features related to individual differences such as personalities or social variables of interest such as status.

DAS scores measured using various feature sets can be used as a vector model to control adaptation in Natural Language Generation (NLG). For example, an adaptation model can be learned through processing human-human dialog corpora and stored in a vector. Table 4.2 shows an adaptation model obtained from the ArtWalk Corpus represented as a DAS vector. We will further describe how the vector is obtained in Section 4.2.3. From left to right, the numbers in the vector represent average DAS scores for all prime-target pairs within different linguistic feature sets: lemma, bigram, syntax, referring expression, etc.

Feature	Lemma	Bigram	Syntax	Refer	Hedge	LIWC	Extra	Emot	Agree	Consc	Open
DAS	0.14	0.04	0.17	0.03	0.17	0.48	0.40	0.48	0.47	0.38	0.44

Table 4.2: Example DAS vector learned from the ArtWalk Corpus.

Although we leave the application of DAS to NLG to future work, here we describe how we expect to use it. We consider the use of DAS with three NLG architectures:

- Overgeneration and Rank
- Statistical Parameterized NLG

- Neural NLG

Overgenerate and Rank. The most general architecture for a conversational assistant and its NLG make use of an “overgenerate and rank” approach, where different modules propose a possibly large set of next utterances in parallel, which are then fed to a (trained) ranker that outputs the top-ranked utterance. Previous work on adaptation/alignment in NLG has made use of this architecture [25, 29]. Rankers can be trained using any features of the proposed utterances and any features available in the discourse context: the DAS adaptation models can be used to provide one or more features to the ranker.

When ranking generated responses, we can choose the best response based on learned adaptation model vector. We first calculate a DAS vector for every response (target) to the discourse context (prime). We then rank responses based on the distance between DAS vector of response and DAS vector of the adaptation model. The response with the smallest distance is presumably the response with the best amount of adaptation. We can also emphasize the importance of specific feature sets by giving weights to different dimensions of the vector and calculating weighted distance. For instance, in order to adapt more to personality, one could prioritize related LIWC features, and adapt by using words from the same LIWC categories, which could avoid too much lexical and syntactical mimicking.

Statistical Parameterized NLG. Some NLG engines provide a list of parameters that can be controlled at generation time [103, 140]. Personage [116] utilizes continuous parameters to control linguistic style variations in its generation results. For example, a Verbosity value of 1 maximize the verbosity of output utterance. In addition, some generation decisions can be non-deterministic, as a result, the values of those parameters are generation decision probabilities. For example, the value of parameter Conjunction is the probability that the aggregation operation combines two propositions with

the conjunction *and*. Similarly, DAS scores can also be used as generation decision probabilities. A LIWC adaptation value of 0.48 indicates that the probability of adapting to LIWC features in discourse context (prime) is 0.48. By mapping DAS scores to generation parameters, the generator could be directly controlled to exhibit the correct amount of adaptation for any feature set.

Neural NLG. Recent work in Neural NLG (NNLG) has started to explore controlling stylistic variation in outputs using a vector to encode style parameters, possibly in combination with the use of a context vector to represent the dialog context [57]. The vector based probabilities that are represented in the DAS adaptation model could be encoded into the context vector in NNLG. No known other adaptation measures could be used in this way. To train the NNLG model, we need a dataset of responses, annotated with a value assignment to each of the adaptation and the content parameters. Adaptation parameters encode the amount of adaptation to the previous dialog turn (prime) within each feature set. Content parameters encode the content of the response.

We hypothesize that different conversational contexts may lead to more or less adaptive behavior, so we apply DAS on the four human-human dialog corpora in Chapter 3: two task-oriented dialog corpora that were designed to elicit adaptation (ArtWalk and Walking Around), one topic-centric spontaneous dialog corpus (Switchboard), and MapTask Corpus used in much previous work. We obtain linguistic features of these corpora using fully automatic annotation tools. We learn models of adaptations from these real human dialogs on various feature sets. We first validate the DAS measure by showing that DAS distinguishes original dialogs from dialogs where the orders of the turns have been randomized, as in previous work [182]. We then show how DAS varies as a function of the feature sets used and the dialog corpora. We also show how DAS can be used for fine-grained adaptation by applying DAS to individual dialog segments, and individual speakers, and illustrating the differences in adaptation as a function of

these variables. Finally, we show how DAS scores decrease as the adaptation window size increases.

4.2.2 Experimental Setup

For AWC and WAC, we remove annotations such as speech overlap, noises (laugh, cough) and indicators for short pauses, leaving only clean text. If more than one consecutive dialog turn has the same speaker, we merge them into one dialog turn.

We consider the following feature sets: unigram, bigram, referring expressions, hedges/discourse markers, and Linguistic Inquiry and Word Count (LIWC) features. Previous work on measuring linguistic adaptation have largely focused on lexical and syntactical features, which are included as baselines. Referring expressions and discourse markers are key features that are commonly studied for adaptation behaviors in task-oriented dialogs, which are often hand annotated. Here we automatically extract these features by rules. To model adaptation on the personality level, we draw features that correlate significantly with personality ratings from LIWC features. We hypothesize that our feature sets described below will demonstrate different adaptation models.

We lemmatize, POS tag and derive constituency structures using Stanford CoreNLP [118]. We then extract the following linguistic features from annotations and raw text. The following example features are based on D137 in Figure 3.2: “and. you know on the uh other side of the math building like there’s the uh, there’s this weird, little concrete, structure that is sticking up out of the bricks, don’t make any sense.”

- Unigram Lemma/POS: we use lemma combined with POS tags to distinguish word senses. E.g. `LEMMAPOS_BUILDING/NN` and `LEMMAPOS_BRICK/NNS` in D137.
- Bigram Lemma: e.g. `BIGRAM_THE-BRICK` and `BIGRAM_SIDE-OF` in D137.
- Syntactic Structure: following [155], we take all the subtrees from a constituency

parse tree (excluding the leaf nodes that contain words) as features. E.g. `SYNTAX_VP->VBP+PP` and `SYNTAX_ADJP->DT+JJ` in D137. The difference is that we use Stanford Parser rather than hand annotations.

- Referring Expression: referring expressions are usually noun phrases. We start by taking all constituency subtrees with root NP, then map the subtrees to their actual phrases in the text and remove all articles from the phrase, e.g., `REFEREXP_LITTLE-CONCRETE` and `REFEREXP_MATH-BUILDING` in D137.
- Hedge/Discourse Marker: hedges are mitigating words used to lessen the impact of an utterance, such as “actually” and “somewhat”. Discourse markers are words or phrases that manage the flow and structure of discourse, such as “you know” and “I mean”. We construct a dictionary of hedges and discourse markers, and use string matching to extract features, e.g., `HEDGE_YOU-KNOW` and `HEDGE_LIKE` in D137.
- LIWC: Linguistic Inquiry and Word Count [144] is a text analysis program that counts words in over 80 linguistic (e.g., pronouns, conjunctions), psychological (e.g., anger, positive emotion), and topical (e.g., leisure, money) categories. E.g., `LIWC_SECOND-PERSON` and `LIWC_INFORMAL` in D137. Because DAS features are binary, features such as Word Count and Number of New Lines are excluded.
- Personality LIWC: previous work reports for each LIWC feature whether it is significantly correlated with each Big Five trait [117] on conversational data [120]. For each trait, we create feature sets consisting of such features. See Table 4.3.

4.2.3 Experiments on Modeling Adaptation

In this section, we apply DAS on four human-human dialog corpora introduced in Chapter 3: two task-oriented dialog corpora that were designed to elicit adaptation (Art-

Personality	#	Example Features
Extraversion	15	Positive Emotion, Swear Words
Emotional Stability	14	Anger, Articles
Agreeable	16	Assent, Insight
Conscientious	17	Fillers, Nonfluencies
Open to Experience	12	Discrepancy, Tentative

Table 4.3: Number of LIWC features for each personality trait and example features.

Walk and Walking Around), one topic-centric spontaneous dialog corpus (Switchboard), and MapTask Corpus used in much previous work.

Experiment 1: Validity Test: Original vs. Randomized Dialogs

We first establish that our novel DAS measure is valid by testing whether it can distinguish dialogs in their original order vs. dialogs with randomly scrambled turns (the order of dialog turns are randomized within speakers), inspired by similar approaches in previous work [10, 65, 182]. We calculate DAS scores for original dialogs and randomized dialogs using target=Both (Sec. 4.2.1) to obtain overall adaptation scores for both speakers.

We first test on lexical features (unigram and bigram) as in previous work. Then we add additional linguistic features (syntactic structure, referring expression, and discourse marker). These five features (see Section 4.2.2) are referred to as “all but LIWC”. Finally, we test DAS validity using the higher level LIWC features.

We perform paired t -tests on DAS scores for original dialogs and DAS scores for randomized dialogs, pairing every original dialog with its randomized dialog. Table 4.4 shows the number of dialogs in each corpus, the average DAS scores of all dialogs within the corpus and p -values of corresponding t -tests. Although the differences between the average scores are relatively small, the differences in almost all paired t -tests are extremely statistically significant (cells in bold, $p < 0.0001$). The paired t -test on

	#	Feature Sets	Original	Random
AWC	48	Unigram + Bigram	0.10	0.07
		All but LIWC	0.13	0.10
		LIWC	0.48	0.46
WAC	36	Unigram + Bigram	0.22	0.19
		All but LIWC	0.18	0.16
		LIWC	0.55	0.54
MPT	128	Unigram + Bigram	0.27	0.24
		All but LIWC	0.20	0.18
		LIWC	0.54	0.54
SWBD	1126	Unigram + Bigram	0.18	0.17
		All but LIWC	0.20	0.19
		LIWC	0.67	0.66

Table 4.4: Number of dialogs in four corpora, and average DAS scores of different feature sets for original and randomized dialogs. Bold numbers indicate statistically significant differences ($p < 0.0001$) between DAS scores for original and randomized dialogs in paired t -tests.

MPT using LIWC features shows a significant difference between the two test groups ($p < 0.05$). The original dialog corpora achieve higher average DAS scores than the randomized corpora for all 12 original-random pairs. Results show that DAS measure is sensitive to dialog turn order, as it should be if it is measuring dialog coherence and adaptation.

Experiment 2: Adaptation Across Corpora and Across Features

This experiment aims to broadly examine the differences in adaptation across different corpora and feature sets. We first compute DAS on the whole dialog level for each feature set from Section 4.2.2, and then calculate the average DAS for each feature set across the corpus. For example, for every dialog in the AWC, we calculate its DAS score of the Lemma/POS feature set, we then take an average of the DAS scores across the corpus and obtain an average DAS score of 0.14. We repeat the same process for the

remaining feature sets. We use target=Both (Sec 4.2.1) to obtain an overall measure of adaptation and leave calculating fine-grained DAS measures to Section 4.2.3. Table 4.5 provides results. We will refer to features in row 1 to 6 as “linguistic features” and row 7 to 11 as “personality features”.

Row	Feature Sets	AWC	WAC	MPT	SWBD
1	Lemma/POS	0.14	0.15	0.29	0.28
2	Bigram	0.04	0.04	0.01	0.07
3	Syntax	0.17	0.14	0.11	0.28
4	ReferExp	0.03	0.03	0.01	0.01
5	Hedge	0.17	0.19	0.18	0.25
6	LIWC	0.48	0.55	0.53	0.71
7	Extra	0.40	0.46	0.30	0.58
8	Emot	0.48	0.50	0.38	0.72
9	Agree	0.47	0.51	0.44	0.71
10	Consc	0.38	0.44	0.20	0.55
11	Open	0.44	0.44	0.31	0.73

Table 4.5: Average DAS scores for each feature set.

Comparing columns, we first examine the DAS scores across different corpora. All p -values reported below are from paired t -tests. The two most similar corpora, the AWC and WAC, show no significant difference on linguistic features ($p = 0.43$). At the same time, the AWC and WAC do differ from the other two corpora. This demonstrates that the DAS reflects real similarities and differences across corpora. MPT shows lower DAS scores on all linguistic features except for lemma (word repetition), where it achieves the highest DAS score. With respect to personality features, WAC has significantly higher DAS scores than AWC ($p < 0.05$), possibly because of the different experiment settings: college student participants are more comfortable around their own campus than in downtown. MPT shows significantly lower DAS scores on personality features than AWC and WAC ($p < 0.05$). This may be because the MPT setting is the most constrained of the four corpora: being fixed in topic and location means dialogs are less likely to be influenced by environmental factors or to contain social chit chat.

SWBD has the highest DAS scores in all feature sets except for referring expression. The higher DAS in non-referring features could be because the social chit chat allows more adaptation to occur. In addition, the dialogs we measure in SWBD are backchannel-filtered. The lower referring expression (relative to other SWBD scores) could be because SWBD does not require the referring expressions necessary for the other three task-related corpora. Compared to AWC, WAC, and MPT, SWBD has longer dialog turns (and even longer after filtering backchannels and merging adjacent dialog turns with the same speaker). Filtered SWBD dialogs have 3 to 85 dialog turns each, while AWC dialogs have 197 to 691 turns, WAC dialogs have 175 to 885 turns, and MPT dialogs have 32 to 438 turns. It is possible that utterances in SWBD are more story like with more narrative structures, which elicits more adaptation behaviors. In addition, SWBD also has shorter conversation length (5 minutes on average) compared to the other three task-oriented corpora. For example, dialogs in AWC range from 24 to 55 minutes. It is possible that the physical time course of the conversation has an effect on the amount of adaptation: shorter dialogs result in more adaptation behavior. It would be interesting to carry out further experiments comparing the DAS scores of the first 5 minutes of AWC, WAC, and MPT to SWBD.

We posit that the DAS adaptation models we present can be used in existing NLG architectures, described in Sec. 4.2.1. The AWC column in Table 4.5 shows adaptation model in the form of a DAS vector obtained from the ArtWalk Corpus, which is further demonstrated in Table 4.2. In order to control adaptation in natural language generation with DAS vector, it is crucial to determine which adaptation model should be used. As we can see from previous results, different task settings yield different adaptation models. Intuitively, we should choose an adaptation model obtained from a dialog corpus with similar task to the desired natural language generation task. For example, a task-oriented computer agent in smart speakers might benefit more from adaptation

models obtained using AWC, WAC or MPT; while a social companion agent might benefit more from adaptation models obtained using SWBD. However, further experiments are needed to determine the most suitable model for a natural language generation task.

Comparing rows, we then examine DAS scores among different features sets. LIWC has the highest DAS score among linguistic features, ranging from 0.48 to 0.71. While other linguistic features are largely content-specific, LIWC consists of higher level features that cover broader categories, thus its high DAS scores are expected. For example, “great” and “wonderful” are both positive emotion words in LIWC categories. If “great” is in prime and “wonderful” is in target, when calculating DAS score using LIWC features, this is considered feature overlap (adaptation). However, when calculating DAS score using lemmas as features, this is not considered feature overlap, thus its DAS score is lower compared to using LIWC features. The DAS scores for the lemma feature range from 0.14 to 0.29, followed by Syntactic Structure (0.11 to 0.28), Hedge (0.17 to 0.25) and Bigram (0.01 to 0.07). Referring Expression has the lowest DAS score (0.01 to 0.03), possibly because our automatic extraction of referring expressions creates numerous subsets of one referring expression. Among personality features, Emotion Stability, Agreeableness, and Openness to Experience traits are adapted more than Extraversion and Conscientiousness. We leave to future work the question of why these traits have higher DAS scores.

Experiment 3: Adaptation by Dialog Segment and Speaker

Our primary goal is to model adaptation at a fine-grained level in order to provide fine-grained control of an NLG engine. To that end, we report results for adaptation models on a per dialog-segment and per-speaker basis.

Reliable discourse segmentation is notoriously difficult [143], thus we heuristically divide each task-oriented dialog into segments based on number of destinations on the

map: this effectively divides the dialog into subtasks. Since each dialog in SWBD only has one topic, we divide SWBD into 5 segments.² We compute DAS for each segment, and take an average across all dialogs in the corpus for each segment.

We compare all LIWC features vs. extraversion LIWC features because they provide high DAS scores across corpora. We also aim to explore the dynamics between two conversants on the extraversion scale. Figure 4.11 illustrates how DAS varies as a function of speaker and dialog segment. In AWC, scores for all LIWC features slightly decrease as dialogs progress (Fig. 4.11a), while extraversion features show a distinct increasing trend with correlation coefficients ranging from 0.7 to 0.86 (Fig. 4.11b), despite being a subset of all LIWC features.³ Average DAS displays the same decreasing trend in all and extraversion LIWC features for SWBD (Fig. 4.11g and 4.11h). We speculate that this might be due to the setup of SWBD: as the dialogs progress, conversants have less to talk about the topic and are less interested. We also calculate per segment adaptation in WAC and MPT, but their DAS scores do not show overall trends across the length of the dialog (Fig. 4.11c to 4.11f).

We also explore whether speaker role and initiative affects adaptation. We use `target=Both`, `target=D`, and `target=F` to calculate DAS for each target. In task-oriented dialogs, D stands for Director, F for Follower. In SWBD, D stands for the speaker initiating the call. A prompt message is played to the initiating speaker in SWBD. For example, “find out what kind of pets the other caller has.” We hypothesize that directors and followers adapt differently in task-oriented dialogs. In all task-oriented corpora (AWC, WAC, and MPT), we observe generally higher DAS scores with `target=D`, indicating that in order to drive the dialogs, directors adapt more to followers. In SWBD, the speaker initiating the call (who brings up the discussion topic and may therefore

²To ensure two way adaptation exists in every segment (both speaker A adapting to B, and B adapting to A), the minimum length (number of turns) of each segment is 3. Thus, we only work with dialogs longer than 15 turns in SWBD.

³Using Simple Linear Regression in Weka 3.8.1.

drive the conversation) generally exhibits more adaptation.

Experiment 4: Adaptation on Different Window Sizes

This experiment aims to examine the trend of DAS scores as the window size increases. We begin with a window size of 1 and gradually increase it to 5. For a window size of n , the target utterance t is paired with the n -th utterance from a different speaker preceding t , if any. For example, in Figure 3.1, when window size is 3, target D100 is paired with prime F97; target D99 does not have any prime, thus no pair is formed.

Similar to Sec. 4.2.3, we compare DAS scores between dialogs in their original order vs. dialogs with randomly scrambled turns. We hypothesize that similar to the results of repetition decay measures [148, 153, 182], the DAS scores of original dialogs would decrease as the window size increases. We use target=both to obtain overall adaptation scores involving both speakers, and calculate DAS with all but the Personality LIWC feature sets introduced in Sec. 5.2. We first compute DAS on the whole dialog level for each window size, and then calculate the average DAS for each window size across the corpus.

Results show that DAS scores for the original dialogs in all corpora decrease as window size increases, while DAS scores for the randomized dialogs stay relatively stable. Figure 4.10 shows plots of average DAS scores on different window sizes for original and randomized dialogs. Plots of the AWC and WAC show similar trends. Experiments with larger window sizes show that the original and random scores meet at window size 6 - 7 (with different versions of randomized dialogs). In MapTask, the original and random scores meet at window size 3 - 4. In SWBD, original and random scores meet at window size 2.

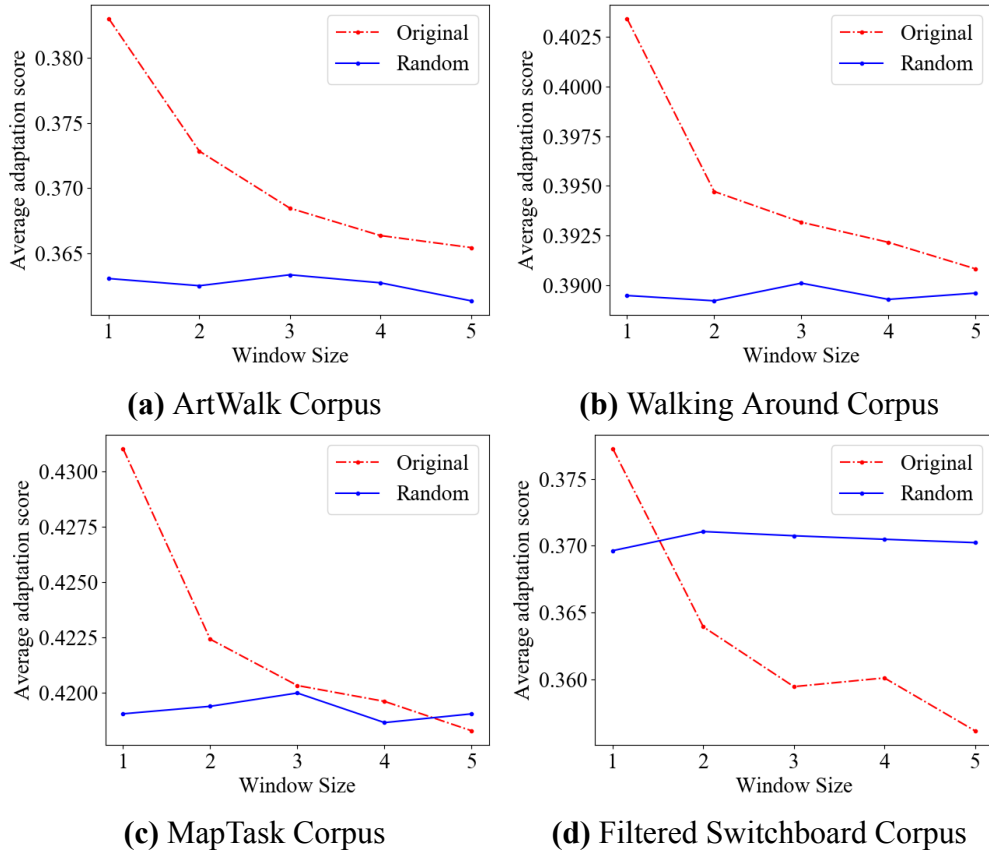


Figure 4.10: Plots of average DAS on different window sizes (1 to 5) for original dialogs vs. randomized dialogs, using all feature sets except Personality LIWC.

4.2.4 Discussion

To obtain models of linguistic adaptation, most measures could only measure an individual feature at a time, and need the whole dialog to calculate the measure [37, 45, 148, 155, 166, 182]. This paper proposes the Dialog Adaptation Score (DAS) measure, which can be applied to NLG because it can be calculated on any segment of a dialog, and for any feature set.

We first validate our measure by showing that the average DAS of original dialogs is significantly higher than randomized dialogs, indicating that it is sensitive to dialog priming as intended. We then use DAS to show that feature sets such as LIWC, Syntactic Structure, and Hedge/Discourse Marker are adapted more than Bigram and Referring

Expressions. We also demonstrate how we can use DAS to develop fine-grained models of adaptation: e.g. DAS applied to model adaptation in extraversion displays a distinct trend compared to all LIWC features in the task-oriented dialog corpus AWC. Finally, we show that the degree of adaptation decreases as the window size increases. We leave to future work the implementation and evaluation of DAS adaptation models in natural language generation systems.

Comparison with Previous Work

Recent measures of linguistic adaptation fall into three categories: probabilistic measures, repetition decay measures, and document similarity measures [188]. Probabilistic measures compute the probability of a single linguistic feature appearing in the target after its appearance in the prime. Some measures in this category focus more on comparing adaptation amongst features and do not handle turn by turn adaptation [37, 166]. Moreover, these measures produce scores for individual features, which need aggregation to reflect overall adaptivity [43, 45]. Document similarity measures calculate the similarity between prime and target by measuring the number of features that appear in both prime and target, normalized by the size of the two text sets [180]. Both probabilistic measures and document similarity measures require the whole dialog to be complete before they can be calculated.

Repetition decay measures observe the decay rate of repetition probability of linguistic features. Previous work has fit the probability of linguistic feature repetition decrease with the distance between prime and target in logarithmic decay models [152, 153, 155], linear decay models [182], and exponential decay models [148].

Previous work on linguistic adaptation in natural language generation has attempted to use adaptation models learned from real human conversations. The alignment-capable microplanner SPUD *prime* [28, 29] use repetition decay model from Reitter as part of

the activation functions for linguistic structures [152]. However, the parameters are not learned from real data. Repetition decay models do well in statistical parameterized NLG, but is hard to apply to overgenerate and rank NLG. Isard, Brockmann, and Oberlander apply a pre-trained n-grams adaptation model to generate conversations [87]. Dušek and Jurčíček use a seq2seq model in order to generate responses adapting to previous context. They utilize an n-gram match ranker that promotes outputs with phrase overlap with context [48]. Our learned adaptation models could also act like such rankers. In addition to n-grams, DAS could produce models with any combinations of feature sets, providing a more versatile adaptation behavior.

Implementation of DAS in Personage-primed

Adaptation models learned using DAS can be stored in a vector and used to control the amount of adaptation in natural language generation (NLG). Section 4.2.1 sketched out how DAS vector models can be implemented in three NLG architectures: Overgeneration and Rank, Statistical Parameterized NLG, and Neural NLG. Here we describe how DAS vector models can be implemented in Personage-primed, the natural language generator used in the exploratory study in Section 4.1.

With Personage-primed, we carried out experiments testing perceptions of adaptation with overgeneration and rank: we generate multiple target utterances that adapt to different combinations of linguistic feature types of the prime utterance. For example, if the prime utterance have syntactic features and referring expressions, then the over-generated utterances will have three combinations of adaptation: adapting to syntactic features only, adapting to referring expressions only, and adapting to both syntactic features and referring expressions.

Building on top of this scheme, the most straightforward implementation of DAS adaptation models is to use DAS vector to rank overgenerated utterances. As discussed

in Section 4.2.1, to rank generated responses, the system chooses the best response based on learned adaptation model. A DAS vector needs to be calculated for every generated response (target) to the discourse context (prime). The responses are then ranked on the distance between their own DAS vectors and the adaptation model. The response with the smallest distance is presumably the response with the best amount of adaptation.

Using this implementation, it is easy to emphasize the importance of specific feature sets by giving weights to different dimensions of the vector and calculating weighted distance. For instance, in order to adapt more to personality, the system could prioritize related LIWC features, and adapt by using words from the same LIWC categories, which could avoid too much lexical and syntactical mimicking. However, this approach requires the calculation of DAS vectors for every overgenerated utterance, which in turn requires annotation (for example, part of speech tagging, lemmatization, parsing, and LIWC) of all generated utterances. This is time-consuming when the system presents a large number of overgenerated candidates.

Personage [115], the underlying natural language generator of Personage-primed, is a parameterized natural language generator. Thus, we propose to implement DAS scores for various linguistic feature sets directly as natural language generator parameters in Personage-primed. Because DAS reflects the average probability that features in prime are adapted in target, we can use probability based parameters to control the amount of adaptation.

For example, parameter HedgeAdaptation is set to the DAS score calculated using Hedge/Discourse Marker feature set. Suppose the value of HedgeAdaptation is 0.2 in Personage-primed, and the system has access to a set of all possible hedges and discourse markers, there are multiple ways to generate utterances that adapt to hedges and discourse markers based on the model provided: (1) adapt to each hedge/discourse marker present in the prime independently, each with a 20% chance; (2) suppose there

are n hedge/discourse marker present in prime, randomly adapt to $\lfloor 0.2 * n \rfloor$ hedge/discourse marker; (3) use method (2), and keep track of the average DAS score between prime and target during generation process: if the DAS score becomes lower than the parameter, use $\lceil 0.2 * n \rceil$ instead. Further experiments are needed to decide which method is optimal.

Another example is parameter SyntacticStructure, which is set to the DAS score calculated using Syntactic Structure feature set. Suppose the value of SyntacticStructure is 0.14, and the system has a set of syntactic variations of the utterance to be generated. Personage-primed can then calculate the DAS score for each syntactic variation based on the prime, and chooses the syntactic structure with DAS score closest to the parameter.

However, this parameterized approach is hard to implement on some feature sets, for example, Unigram Lemma/POS and Bigram Lemma. It is hard to control the percentage of general words that are adapted in the target during generation time. To amend the disadvantage, we can divide the more general linguistic feature sets into specific feature sets, for example, nouns, verbs, and adjectives, and obtain DAS models for all feature sets. Synonym sets of words in each feature set can be constructed using resources such as WordNet and VerbNet to provide variations in the cases of non-adaptation.

In addition to the direction giving domain in current work, we also plan to expand Personage-primed to more domains in future work. For example, the original Personage generator was implemented in restaurant recommendation domain, whose relevant resources such as syntactic structures and synonym sets are largely available. Using the Story Dialogs with Gestures Corpus introduced in Section 3.5, we can also extend Personage-primed to storytelling domain. Following previous work on the Story Intention Graph (SIG) annotations, sentence planning variations can be automatically generated [110].

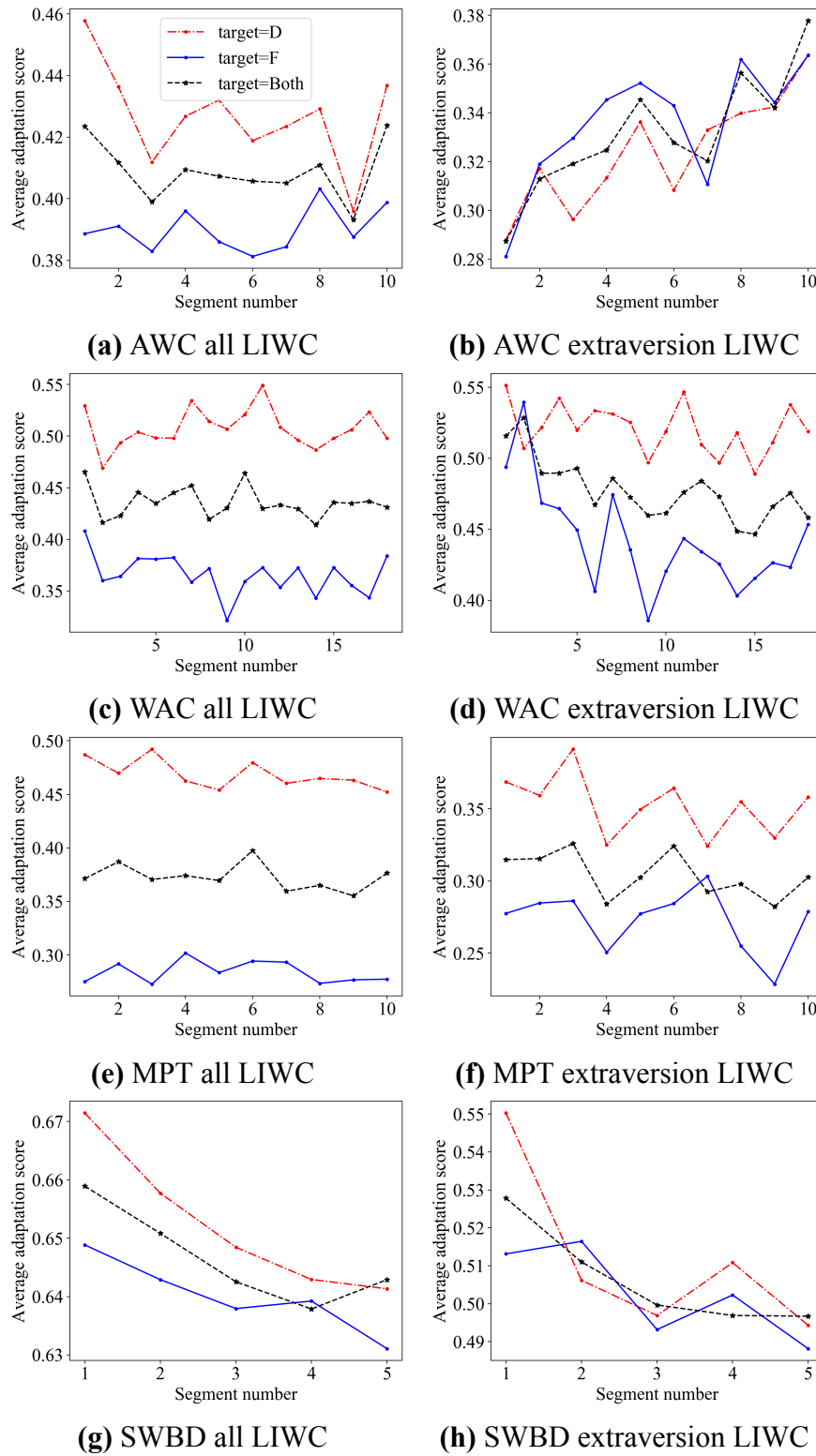


Figure 4.11: Plots of average DAS as the dialogs progress, using all LIWC features vs. extraversion LIWC features.

Chapter 5

Gestural Adaptation

The most primitive method of gestural adaptation is copying the dialog partner's gesture forms. However, as illustrated in Table 1.4, Chapter 1, simple mimicking of dialog partner's gestures is far from providing a positive user experiences. A more natural method is to adapt to gesture parameters, such as rate, speed, scale, and expanse, according to the dialog partner's personality. In this chapter, we not only experiment with adaptation of particular gesture forms, but also go beyond and study adaptation of gestural style involving personality. We carry out two experiments. In the personality experiment, we elicit subjects' perceptions of two virtual agents designed to have different personalities. In the gestural adaptation experiment, we ask whether subjects prefer adaptive vs. nonadaptive agents. Our results show that agents intended to be extraverted or introverted are perceived as such, and that subjects prefer adaptive stories. We describe stimulus construction in Sec. 5.1. Sec. 5.2 and 5.3 present our experimental design and results.



Figure 5.1: A snapshot of the experimental stimuli.

5.1 Stimulus Construction

We construct our experiment stimuli using the Story Dialog with Gestures Corpus described in Sec. 3.5. In this experiment, we use four stories with different subject matter: protest, pet, storm and gardening, as illustrated in Fig. 3.14 and Fig. 3.13. Fig. 5.1 shows a screenshot of the stimuli. We use our own animation software to generate the stimuli based on the specified gesture script. This software uses motion captured data for the wrist path, hand shape and hand orientation for each gesture stroke, motion captured data for body movement, and spline based interpolation for preparation and retractions. It also uses simplified physical simulation to add nuance to the motion. A gesture contains up to 4 phases: prep, stroke, hold and retract: we insert a hold and connecting prep between two strokes if they are less than 2.5 seconds away from each other. Otherwise, we insert a retraction.

- If a gesture is the last gesture of this utterance (dialog turn), then it should retract in the end.
- If a gesture is not the last gesture of this utterance, and its stroke ending time is

at most 2.5 seconds before the next gesture's stroke time, then hold until the next gesture happens.

- If a gesture is not the last gesture of this utterance, and its stroke ending time is larger than 2.5 seconds before the next gesture's stroke time, then retract.

We use custom gesture synthesis algorithms built on top of DANCE (Dynamic Animation and Control Environment) [165] as our simulation platform. DANCE is an open and extensible simulation framework and rapid prototyping environment for computer animation. DANCE mainly focuses on the development of physically-based controllers for articulated figures. In this experiment, we generate gesture sequences based on manually created scripts, while our proposed work aim to generate such scripts procedurally. These scripts are phase, phrase and syn files.

- phase file: exact begin and end time of gesture phases: prep, stroke, hold, retract (introduced in Section 3.5).
- phrase file: exact begin and end time of a gesture, gesture name and hand used.
- syn file: edits applied to gestures (swivel angle of arm, arm distance from body, etc.), body movements, etc.

Figure 5.2 shows an example of the scripts. In the phase file, the first gesture has three components: prep, stroke and hold. From 1.60 second to 1.90 second is the prep phase of this gesture, the agent moves his arm from the default resting position to the stroke starting position. From 1.90 second to 2.36 second is the stroke phase, the agent performs the gesture. From 2.36 second to 2.87 second is the hold phase, the agent holds his arm until the next phase happens. The next phase is the prep phase of the next gesture, so the agent moves his arm directly from the previous hold position, to the stroke starting position of the next gesture. In the phrase file, the TimeOffset decides

from what time the gesture sequence begins. Then a list of headers is defined. Finally, the exact begin and end time of a gesture, gesture name and hand use information is specified for every gesture in the phase file. For example, “1.60 2.87 Cup RH” means that from 1.60 second to 2.87 second, the gesture “Cup” happens, using the right hand. The start and end time 1.60 second and 2.87 second corresponds to the time specified in the phase file. In the syn file, a list of gesture edits are specified. For example, “spatialwarp ALL trans stroke 0.15 -0.10 0.00” translates the strokes of all gestures by moving it 15cm in the x direction (15 cm wider), -10cm to the y direction (10 cm lower). To make the interaction more natural, we add body movements by using “copyData” in the syn file, which reproduces postural movements captured previously via motion capture.

phase file	phrase file	syn file
1.60 1.90 prep 1.90 2.36 stroke 2.36 2.87 hold	TimeOffset 0.0	spatialwarp ALL scale stroke 1.40 1.20 1.10 spatialwarp ALL trans stroke 0.15 -0.10 0.00 timing ALL scaleDuration stroke L 0.80
2.87 3.17 prep 3.17 3.54 stroke 3.54 4.67 hold	start end lexeme handedness path handshape hand-height-1 hand-body-dist-1 hand-radial-orient-1 elbow-inclination-1 hand-height-2 hand-body-dist-2 hand-radial-orient-2 elbow-inclination-2 2H-distance-start 2H-distance-shoulders dss_lex_affil dse_lex_affil lex_affil dss_cooc dse_cooc cooc	swivelArms NONE useDefault swivelArms NONE shift -12.00 offset NONE defaultTorso LCollarZ ADD 0.18 offset NONE defaultTorso RCollarZ ADD -0.18 offset NONE defaultTorso zerspines
4.67 4.97 prep 4.97 5.54 stroke 5.54 6.93 hold	1.60 2.87 Cup RH 2.87 4.67 PointingAbstract RH 4.67 6.93 Cup_Horizontal 2H 6.93 8.90 Progressive 2H 13.90 14.87 PointingAbstract RH	copyData NONE copyLBody data/motionDB/ Tired_CONVERT_wToes2_longClip.dat all 10 offset NONE paramData COM_X REL 0.55 offset NONE paramData COM_X ADD 0.03 offset NONE paramData COM_Z REL 0.55 offset NONE paramData COM_Z ADD 0.03
6.93 7.23 prep 7.23 8.60 stroke 8.60 8.90 retract
13.90 14.20 prep 14.20 14.57 stroke 14.57 14.87 retract
.....

Figure 5.2: Part of DANCE gesture sequence scripts for speaker A in the protest dialog in Fig 3.14.

The animation platform takes the scripts as input and produces a bvh file of the animation, reflecting the detailed timing and movement information of every gesture. We then import this bvh file into Maya for rendering. In the video, two IVAs stand almost face-to-face, but each has an 55° angle towards the audience.

Head movements are inserted using the following rules: (a) the agent looks at the audience when he is talking, (b) the agent looks at the other agent when he's ready to give his turn to the other agent, and (c) the agent keeps looking at the other agent when the other agent is talking.

5.2 Experiment Method

We conduct two separate experiments, one on personality variation during co-telling a story, and the second using the same personalities but with and without adaptation.

5.2.1 Experiment 1: Personality Variation.

We prepared two versions of the video of the story co-telling for each of the four stories, one where the female is extraverted (higher values for gesture rate, gesture expanse, height, outwardness, speed and scale) and the male is introverted (lower values for those gesture features) and one where only the genders (virtual agent model and voice) of the agents are switched. The dialog scripts and corresponding gesture forms do not vary from one co-telling to another. This results in 8 video stimuli for four stories.

We conducted a between-subjects experiment on Mechanical Turk where we first ask Turkers to answer the TIPI [73] personality survey for themselves, and then answer it for only one of the agents in the video, after watching the video as many times as they like. Thus for each video stimulus, there are two surveys. We ran our 16 surveys as 16 HITs (Human Intelligence Tasks) on Mechanical Turk, requesting 20 subjects per HIT (each worker can only do one of the tasks), which results in 320 judgments. The average completion time for the 8 HITs on Mechanical Turk was 5 minutes 15 seconds. The average stimulus length was 1 minute 32 seconds. Since the survey is hosted outside Mechanical Turk, sometimes we get more than 20 subjects for each HIT.



Figure 5.3: Virtual agent with different gesture expanse and height for the same gesture.

5.2.2 Experiment 2: Gestural Adaptation.

For the adaptive experiment, both agents are designed to be extraverted. We chose to use two extraverted agents because we have foundations from previous work showing the adaptation model between two extraverted speakers [177] (where both agents become more extraverted). We use only a part of each story for one experimental task. The stimuli for one task has two variations: adapted and non-adapted. Both stimuli use the same audio, contain 2 to 4 dialog turns with the same gestures as an introduction to the story (which we refer to as context), and the next (and last) dialog turn with gesture adaptation or without gesture adaptation (which we refer to as response). Adaptation only occurs in the last dialog turn. In this way, subjects can get to know the story through the context, and compare the responses to decide whether they like the adapted or non-adapted version.

- Non-adapted: In the last dialog turn, the extraverted agent maintains his or her gesture rate (1 - 2 gestures per sentence), expanse, height, outwardness, speed and scale. There is no copying of specific gestures.
- Adapted: In the last dialog turn, the extraverted agent increases the gesture rate (1 - 3 gestures per sentence), expanse (18 cm further from center), height (10 cm higher), outwardness (10 cm more outward), speed (1.25 times faster) and scale (1.5 times larger). Fig. 5.3 shows the same gesture with different expanses and

heights. In the adapted version, specific gestures are copied (e.g. gestures in bold font in Fig.3.14).

Thus every story has two versions. One version ends with the female agent's response, another ends with the male agent's response. For example, Garden ABA has three turns, ending with the female agent adapting to the male, and Garden ABAB has four turns, ending with the male agent adapting to the female. Every version consists of two conditions (adapted and non-adapted versions) and a short survey. The order of the two conditions is random for every participant. But there is a letter mark assigned to every video for easy reference (see Fig. 5.1).

Subjects are asked to watch the two stimuli first, and then finish the survey. Subjects are told that the audio of the two videos is the same, but only the last few gestures of the female/male agent are different. Subjects are also advised to watch the video as many times as they want. The survey has two questions: (1) Which video is a better story co-telling based on the gestures? (2) Please explain the reason behind your choice to the previous question (which we refer to as the "why" question). Our primary aim is to determine whether people perceive the adaptation and whether it makes a better story.

We ran our 8 tasks for 4 stories as 8 HITs on Mechanical Turk, requesting 25 subjects per task. The average completion time for the 8 tasks on Mechanical Turk was 2 minutes 53 seconds. The average stimulus length was 35.3 seconds. This means that, on average, a subject spent 1 minute 43 seconds answering the questions. We removed subjects who failed to state their reasons of preference in the "why" question.

Story	Intro-Agent	Extra-Agent
Garden	4.2	5.4
Pet	4.7	5.0
Protest	4.2	5.3
Storm	3.7	5.7

Table 5.1: Experiment results: participant evaluated extraversion scores (range from 1 - 7, with 1 being the most introverted and 7 being the most extraverted).

5.3 Experimental Results

5.3.1 Personality Results

We conducted a three-way ANOVA with agent intended personality, agent gender and story as independent variables and perceived agent personality as the dependent variable. See Table 5.1. Results show that subjects clearly perceive the intended extraverted or intended introverted personality of the two agents ($F = 67.1, p < .001$). There is no main effect for story (as intended in our design), but there is an interaction effect between story and intended personality, with the introverted agent in the storm story being seen as much more introverted than in the other stories ($F = 7.5, p < .001$). There is no significant variation by agent gender ($F = 2.3, p = .14$).

Since previous work suggests that personality is perceived for an agent along all Big Five dimensions whether it is designed to be manifest or not [107, 116], we also conducted a two-way ANOVA by story and agent intended personality for the other 4 traits. There are no significant differences for Conscientiousness, or Openness. However Introverted agents are seen as more agreeable ($p = .008$) and more emotionally stable ($p = .016$). There were no significant differences by story except that both agents in the Storm story were seen as less open, presumably because the content of the story is about how scary the storm is.

5.3.2 Adaption Results

Story Version	#A	#NA	%A	%NA
Garden ABA	11	9	55%	45%
Garden ABAB	20	2	91%	9%
Pet ABABA	10	13	43%	57%
Pet ABABAB	19	5	79%	21%
Protest ABAB	8	11	42%	58%
Protest ABABA	11	11	50%	50%
Storm ABABA	16	4	80%	20%
Storm ABABAB	14	5	74%	26%
Total	109	60	64%	36%

Table 5.2: Experiment results: number and percentage of subjects who preferred the adapted (A) stimulus and the non-adapted (NA) stimulus. The letters in the story version refer to dialog turns by speaker A or B. For example, ABA means A takes dialog turns 1 and 3 in the stimuli, while B takes dialog turn 2.

The results in Table 5.2 show that across all the videos, the mean percentage of people who preferred the adapted version was 64% (19% standard deviation), which is marginally better than a predicted preference of 50%, $t(7) = 2.15, p = .07$. Analysis of participants' descriptions of why they preferred one video over another shows 4 distinct categories of reasons of why people made their choices (see Table 5.3).

Subjects who preferred the adapted versions said that the gestures fit the dialog better (“adapted good gestures” in Table 5.3): the subjects stated that the adapted versions had gestures that “flowed better with the words”, were “more natural”, “more appropriate to what he said”, and “relevant to the dialog”, and that they “could imagine a friend making various hand gestures similar” to the ones in the story. Another reason was that gestures were “more animated” (“adapted animated”): the adapted version had “more hand gestures”, and the agent “used his arms more”, “gestured more”, and “was much more alive”. In contrast, in the non-adapted version, the agent “seemed very bored” and “wanted to end the conversation”. This indicates that the subjects preferred agents with

a higher gesture rate. Ten subjects commented on the expanse, height, scale and speed of the gestures: they chose the adapted version because the agent “gestured higher in the air”, “making wider, grander gestures” that were “more expansive” and “bigger”. And in the non-adapted version, the gestures were “too slow”. However, there was no comment about the copying of gestures, possibly because copying was less obvious when the expanse and height of the gestures changed in the adapted version.

Among those who preferred the non-adapted versions of the stories, one reason was that the gestures fit the dialog better (“non-adapted good gestures” in Table 5.3): the subjects stated that the gestures in the non-adapted version “went a lot better with what she was saying” and were “more appropriate”. Another reason is that the gestures were “more realistic” (“non-adapted realistic”) : subjects didn’t like the gestures being “too animated”, or “too busy”, nor did they like the agents “showing way too much emotions” or “looking like she is exercising”. That is, too much animation can be seen as unrealistic.

Story Version	%A good gest	%NA good gest	%A animated	%NA realistic
Garden ABA	30%	30%	20%	30%
Garden ABAB	41%	9%	59%	0%
Pet ABABA	22%	43%	13%	9%
Pet ABABAB	54%	13%	33%	0%
Protest ABAB	21%	32%	26%	0%
Protest ABABA	27%	32%	23%	9%
Storm ABABA	20%	15%	45%	0%
Storm ABABAB	32%	21%	47%	0%
Total	31%	24%	33%	6%

Table 5.3: Answers to the second survey question (“why” question) classified into categories. Note that one subject could belong to none or multiple categories, so the percentages for each line don’t add up tp 100%.

The percentages of the subjects that had comments related to those 4 categories are in Table 5.3. In 7 out of 8 tasks, there were more subjects who preferred the adapted version

because it was animated at the right level (e.g. animated enough, but not too animated). If we only consider the “animated” factor in deciding which is a better stimulus, 84% of the subjects preferred the adapted version.

5.4 Discussion

To our knowledge this is the first time that it has been shown that subjects perceive differences in agent personality during a storytelling task, and that adaptive gestural behavior during storytelling is positively perceived. We re-use natural personal narratives that are rendered dialogically, so that two IVAs co-tell the story.

It is obvious that being able to adapt is a key part of being more human-like. There are attempts to integrate language adaptation within natural language generation [28] and research has shown that human bystanders perceive linguistic adaptation positively [82]. However, this is the first experiment to demonstrate a positive effect for gestural adaptation.

Recent work on gesture generation has focused largely on iconic gesture generation. For example, Bergmann and Kopp present a model that allows virtual agents to automatically select the content and derive the form of coordinated language and iconic gestures [14]. Luo, Kipp, and Neff also presents an effective algorithm for adding full body postural movement to animation sequences of arm gestures [112]. More generally, current systems generally select gestures using either a text-to-gesture or concept-to-gesture mapping. Text-to-gesture systems, such as VHP [136], may have a limited number of gestures (only 7 in this case) and limited gesture placement options, but the alignment of speech content and gestures are more accurate. Concept-to-gesture systems such as PPP [5], AC and BEAT [33] defines general rules for gesture insertion based on linguistic components. For example, iconic gestures are triggered by words with spatial or concrete context (e.g. “check”). This kind of systems have more gestures,

but the gesture placement largely depends on general rules derived from literature, thus the accuracy is not guaranteed. An alternative approach learns a personalized statistical model that predicts a gesture given the text to be spoken and a model that captures an individual's gesturing preferences [130]. None of these models adequately address the production of gesture for dialogues, where a process of co-adaptation will modulate both the type of gesture chosen and the specific form of that gesture (e.g. its size). This current work aims to provide a basis for developing such models.

Gratch et al. investigate creating rapport with virtual agents using gesture adaptation mainly focused on head gestures and posture shifts (while ours focused on hand gestures), and use real human movements as control [74]. Our adaptation stimuli are more similar to Endrass et al. [54]. To investigate culture-related aspects of behavior for virtual characters, they chose prototypical body postures from corpora for German and Japanese cultural background, embodied those postures in a two-agent dialogs, and asked subjects from German and Japanese cultural background to evaluate the dialogs.

Chapter 6

Conclusion and Future Work

Intelligent virtual agents are making their ways to our daily lives. Google Assistant, Siri and Alexa help us with daily tasks; virtual assistants in customer service websites provide us with information; virtual agents in learning programs help children gain knowledge. However, most verbal and non-verbal behaviors of these agents are generated from a hand-crafted set of scripts, limiting their ability to dynamically adapt to human users. Adaptation is natural to human dialogs and correlates highly with task success. A step towards the adaptivity of intelligent virtual agents is a step towards more enjoyable interaction.

This thesis aims at making intelligent virtual agents more human-like by adding the ability to adapt to users. We propose a vector-based adaptation framework for both linguistic and gestural adaptation. Our goal is to control adaptation behaviors in virtual agents with a vector-based adaptation model. In real human dialogs, speakers try to optimize different goals: produce new and coherent contents, express their own personality or style, and adapt to the other speaker. To enable adaptation in virtual agents, we also need to satisfy all these constraints. Adaptation is not blind mimicking, as shown in the examples of Chapter 1, incorrect adaptation behaviors could lead to negative user experience.

In linguistic adaptation, we first explore various linguistic features and user perceptions of adapting to different combinations of linguistic features. Then we propose a measure of adaptation, Dialog Adaptation Score (DAS), which aims to reflect adaptation models of features sets that describe certain linguistic style. We use DAS to model adaptation in four dialog corpora. We obtain DAS vectors which can be applied in natural language generation to control adaptation. In gestural adaptation, the lack of annotated gesture data makes it impossible to measure gestural adaptation in real situations. Thus, we first explore how to produce the desired personality through adjusting gesture parameters, such as gesture rate, speed, and expanse. On the basis of those results, we experiment with adaptation of particular gesture forms, as well as adaptation of gestural parameters involving personality.

6.1 Linguistic Adaptation

In order to test adaptation effects of different linguistic feature sets, this thesis first presented an exploratory study based on Personage-primed, a natural language generator that can dynamically adapt to the linguistic features in previous dialog context, such as syntactic structures, hedges, and referring expressions. We aimed to test user perceptions of adapting to different combinations of linguistic features. In a user study, human participants are asked to do surveys, in which, given the previous context of an utterance, participants rank several generated utterances on both naturalness and friendliness. These utterances include: (1) adapted utterances with different features combinations, (2) an utterance without adaptation, and (3) a random human utterance with the same meaning but out of context. We use linear regression models and decision tree models to learn the correlation between types of linguistic features and user perceptions. Our results show that human perceptions of naturalness are distinct from friendliness: adapting on hedges increase perceptions of friendliness while reducing naturalness, while adapting

on referring expressions, syntactic template selection and tense/modal choices increase perceptions of both naturalness and friendliness.

To obtain models of linguistic adaptation, most measures could only measure an individual feature at a time, and need the whole dialog to calculate the measure [37, 45, 148, 155, 166, 182]. In this thesis, we proposed the Dialog Adaptation Score (DAS) measure, which can be applied to NLG because it can be calculated on any segment of a dialog, and for any feature set. We first validate our measure by showing that the average DAS of original dialogs is significantly higher than randomized dialogs, indicating that it is sensitive to dialog priming as intended. We then use DAS to show that feature sets such as LIWC, Syntactic Structure, and Hedge/Discourse Marker are adapted more than Bigram and Referring Expressions. We also demonstrate how we can use DAS to develop fine-grained models of adaptation: e.g. DAS applied to model adaptation in extraversion displays a distinct trend compared to all LIWC features in task-oriented dialog corpora AWC.

DAS scores calculated using human dialogs can be expressed in a vector form where each dimension contains the DAS score for a linguistic feature set. This DAS vector can be used as an adaptation model in various natural language generation architectures to control the amount of adaptation. In overgenerate and rank, the system can calculate a DAS score for each response, rank all possible responses by the distance between its DAS score and the adaptation model. The best response is the one with the smallest distance to the adaptation model. In statistical parameterized natural language generation, DAS scores can be used as probability based parameters. In natural language generation using neural networks, our adaptation model can be encoded into the context vector.

6.2 Gestural Adaptation

To our knowledge this is the first time that it has been shown that subjects perceive differences in agent personality during a storytelling task, and that adaptive gestural behavior during storytelling is positively perceived. We use personal narratives from web blogs that are rendered dialogically, so that two IVAs co-tell the story. We not only experiment with adaptation of particular gesture forms, but also go beyond and study adaptation of gestural parameters involving personality.

In order to show that personalities can be expressed through gesture parameters, we first perform an experiment where we present story co-telling stimuli between an introverted agent and an extraverted agent to human participants. Extraverted agents have higher gesture rate, speed and expanse compared to introverted agents. Our results show that personalities of agents are perceived as intended. In a second experiment, we vary whether the agents adapt to their dialog partners' gestures in gesture rate, speed, expanse, and use of specific gestures. Our results show that participants prefer storytellings with adaptive agents. We hope our results can provide inspirations for implementation of gestural adaptation in virtual agents.

6.3 Future Work

Our linguistic adaptation experiment is the first to our knowledge to be based on a generator that can produce utterances dynamically adapted to the context. For future work, several aspects of our generator can be improved to enable more experiments on adaptation: (1) automatic extraction of prime/target features can provide us with more annotated utterances; (2) a wider range of linguistic features, such as LIWC features, can enable adaptation beyond words, phrases, and syntactic structures; (3) integration of learned adaptation models, such as ones learned by DAS, can enable decisions of

which features to adapt and how much, thus allowing us to test dynamically produced adaptation in the field. In addition, our exploratory experiment was carried out with three expert participants from linguistic background, which might introduce bias in experimental results. In future work, carrying out the exploratory experiment with a much larger and diverse participant pool (e.g. using Mechanical Turk) is preferred to further verify the friendliness vs. naturalness results.

Adaptation is the phenomenon that human conversants in dialogs adjust their behaviors to their conversational partners. For example, adopting a certain referring expression or reusing a certain syntactic structure. Theoretically, in order for adaptation to happen, there must be a non-adapting way of saying the same thing. That is to say, not all word-copying and syntax-mimicking are adaptation (for example, reusing the word “the”). However, deciding whether the repetition of a certain linguistic feature is adaptation or not is hard without human annotations. As a result, in computation approaches of measuring linguistic adaptation, previous work has largely used feature repetition as adaptation. This thesis adopts the same notion, which is one of the limitations of this work. In future work, we plan to develop methods of recognizing adaptation behaviors by using rules to exclude words without synonyms (for example, “the” and “you”) and by machine learning methods to cluster different ways of expressing the same meaning.

Our linguistic adaptation measure Dialog Adaptation Score (DAS) only utilize linguistic features as binary features. It reflects the probability of linguistic features in prime adapted in target. Thus only the existence of such features are considered. While most primitive linguistic features have a frequency of 1 in natural dialogs (e.g. unigrams and bigrams), many higher level features that reflect personal linguistic style, such as LIWC features, tend to have higher frequencies. In future work, we plan to take into consideration the frequency of linguistic features.

In addition, the feature set Hedge/Discourse Marker used in measuring adaptation

is a mixture of hedges and discourse markers, which are fundamentally different in their functions. Hedges are mitigating words used to lessen the impact of an utterance, such as “actually” and “somewhat”. Discourse markers are words or phrases that manage the flow and structure of discourse, such as “you know” and “I mean”. In future work, we plan to distinguish hedges and discourse markers in modeling adaptation. We also want to consider backchannels, which are listener responses in a primarily one-way communication, such as “yes” and “uh-huh”.

In Section 4.2.3 Experiment 3, we compare adaptation behaviors between different speaker roles. In addition to director/follower and initiator/non-initiator, we plan to investigate more personal traits. For example, do extraverts adapt more? Do agreeable people adapt more? Which personality is more likely to adapt? The ArtWalk Corpus contains personality information of participants, which could potentially help us answer these questions.

As discussed in Section 4.2.4, adaptation models expressed in DAS vectors can be integrated in various natural language generation architectures, such as overgenerate and rank, statistical parameterized language generation, and neural networks. In future work, we hope to apply models we learn using DAS to natural language generation systems, such as our own Personage-primed. We leave to future work the question of which model is the right one for a particular new conversational situation.

Although it might appear obvious that too much adaptation can lead to negative user impression, more evidence is needed to back up such claim. This can be verified with the application of DAS vectors in natural language generation systems, where we compare the generation results with DAS models learned from a human dialog corpora and an artificial DAS model with high scores. We plan to evaluate using human perceptions with surveys that compare two versions of generated utterances in terms of naturalness and friendliness.

In our gestural adaptation experiments, we only tested one aspect of the Big-Five personality traits: extraversion. In future work, we aim to test the expression of personality and adaptivity with different personality combinations, such as agreeableness and neuroticism. Our ultimate goal is to test dynamically produced gestural adaptation to human users in virtual agents. In order to achieve that, we need to combine the latest technologies such as user personality recognition, gesture generation, and models of gesture adaptation. Experimental exploration, such as undertaken here, is crucial for formulating models of gesture generation that correctly incorporate personality and adaptation.

Another limitation of the gestural adaptation experiment is that the results simply show that participants prefer the adapted version of storytelling, where the extraverted speaker become more extraverted (with more gestures, bigger gesture expanse, and faster gesture speed). However, it is not clear whether participants prefer virtual agents who are more extraverted, or prefer virtual agents with adaptation. Future experiments are needed to test adaptation with more virtual agent personalities in order to verify the effect of adaptation.

Bibliography

- [1] Salvatore M Aglioti and Mariella Pazzaglia. “Representing actions through their sound”. In: *Experimental brain research* 206.2 (2010), pp. 141–151.
- [2] Anne H Anderson, Miles Bader, Ellen Gurman Bard, Elizabeth Boyle, Gwyneth Doherty, Simon Garrod, Stephen Isard, Jacqueline Kowtko, Jan McAllister, Jim Miller, et al. “The HCRC map task corpus”. In: *Language and speech* 34.4 (1991), pp. 351–366.
- [3] Anne H Anderson, Miles Bader, Ellen Gurman Bard, Elizabeth Boyle, Gwyneth Doherty, Simon Garrod, Stephen Isard, Jacqueline Kowtko, Jan McAllister, Jim Miller, et al. “The HCRC map task corpus”. In: *Language and speech* 34.4 (1991), pp. 351–366.
- [4] E. André, M. Klesen, P. Gebhard, S. Allen, and T. Rist. “Integrating Models of Personality and Emotions into Lifelike Characters”. In: *Proceedings of the Workshop on Affect in Interactions - Towards a new Generation of Interfaces*. 1999, pp. 136–149.
- [5] Elisabeth André, Jochen Müller, and Thomas Rist. “WIP/PPP: Automatic generation of personalized multimedia presentations”. In: *Proc. of the fourth ACM international conference on Multimedia*. 1997, pp. 407–408.
- [6] Elisabeth André, Thomas Rist, Susanne van Mulken, Martin Klesen, and Stephan Baldes. “The automated design of believable dialogues for animated presentation teams”. In: *Embodied conversational agents*. Ed. by J. Sullivan S. Prevost J. Cassell and E. Churchill. Cambridge, MA: MIT Press, 2000, pp. 220–255.
- [7] Michael Argyle. “Bodily communication .” In: (1988).
- [8] Jeremy N. Bailenson and Nick Yee. “Digital chameleons: Automatic assimilation of nonverbal gestures in immersive virtual environments”. In: *Psychological Science* 16.10 (2005), pp. 814–819.
- [9] Jeremy N Bailenson and Nick Yee. “Digital chameleons automatic assimilation of nonverbal gestures in immersive virtual environments”. In: *Psychological science* 16.10 (2005), pp. 814–819.

- [10] Regina Barzilay and Mirella Lapata. “Collective content selection for concept-to-text generation”. In: *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*. Association for Computational Linguistics. 2005, pp. 331–338.
- [11] Janet B Bavelas, Alex Black, Charles R Lemery, and Jennifer Mullett. ““ I show how you feel”: Motor mimicry as a communicative act.” In: *Journal of Personality and Social Psychology* 50.2 (1986), p. 322.
- [12] Janet Beavin Bavelas, Alex Black, Charles R Lemery, and Jennifer Mullett. “14 Motor mimicry as primitive empathy”. In: *Empathy and its development* (1990), p. 317.
- [13] Kirsten Bergmann and Stefan Kopp. “Gestural alignment in natural dialogue”. In: *Proc. of the 34th Annual Conference of the Cognitive Science Society (CogSci 2012)*. 2012.
- [14] Kirsten Bergmann and Stefan Kopp. “Increasing the expressiveness of virtual agents: autonomous generation of speech and gesture for spatial description tasks”. In: *Proc. of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*. International Foundation for Autonomous Agents and Multiagent Systems. 2009, pp. 361–368.
- [15] T. Bickmore and D. Schulman. “The comforting presence of relational agents”. In: *CHI’06 extended abstracts on Human factors in computing systems*. 2006, pp. 550–555. isbn: 1595932984.
- [16] Jennifer G Bohanek, Kelly A Marin, Robyn Fivush, and Marshall P Duke. “Family narrative interaction and children’s sense of self”. In: *Family process* 45.1 (2006), pp. 39–54.
- [17] Holly P Branigan, Martin J Pickering, and Alexandra A Cleland. “Syntactic coordination in dialogue”. In: *Cognition* 75.2 (2000), B13–B25.
- [18] H.P. Branigan, M.J. Pickering, J. Pearson, and J.F. McLean. “Linguistic alignment between people and computers”. In: *Journal of Pragmatics* 42.9 (2010), pp. 2355–2368. issn: 0378-2166.
- [19] H.P. Branigan, M.J. Pickering, J. Pearson, J.F. McLean, and C.I. Nass. “Syntactic alignment between computers and people: the role of belief about mental states”. In: *Proceedings of the 25th Annual Conference of the Cognitive Science Society*. 2003, pp. 186–191.
- [20] John Brebner. “Personality theory and movement”. In: *Individual differences in movement*. Springer, 1985, pp. 27–41.
- [21] Susan E. Brennan. “Lexical entrainment in spontaneous dialog”. In: *Proceedings of the International Symposium on Spoken Dialogue*. 1996, pp. 41–44.

- [22] Susan E Brennan and Herbert H Clark. “Conceptual pacts and lexical choice in conversation.” In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22.6 (1996), p. 1482.
- [23] Susan E Brennan and Herbert H Clark. “Conceptual pacts and lexical choice in conversation”. In: *Journal of Experimental Psychology-Learning Memory and Cognition* 22.6 (1996), pp. 1482–1493.
- [24] Susan E Brennan, Katharina S Schuhmann, and Karla M Batres. “Entrainment on the move and in the lab: The Walking Around Corpus”. In: *Proc. of the 35th Annual Conference of the Cognitive Science Society. Austin, TX: Cognitive Science Society. 2013.*
- [25] Carsten Brockmann. “Personality and Alignment Processes in Dialogue: Towards a Lexically-Based Unified Model”. PhD thesis. University of Edinburgh, School of Informatics, 2009.
- [26] Carsten Brockmann, Amy Isard, Jon Oberlander, and Michael White. “Modelling alignment for affective dialogue”. In: *Workshop on Adapting the Interaction Style to Affective Factors at the 10th International Conference on User Modeling (UM-05)*. 2005.
- [27] Jerome Bruner. “The narrative construction of reality”. In: *Critical inquiry* (1991), pp. 1–21.
- [28] Hendrik Buschmeier, Kirsten Bergmann, and Stefan Kopp. “An alignment-capable microplanner for natural language generation”. In: *Proceedings of the 12th European Workshop on Natural Language Generation*. Association for Computational Linguistics. 2009, pp. 82–89.
- [29] Hendrik Buschmeier, Kirsten Bergmann, and Stefan Kopp. “Modelling and Evaluation of Lexical and Syntactic Alignment with a Priming-Based Microplanner.” In: *Empirical methods in natural language generation* 5980 (2010).
- [30] Donn Byrne and Don Nelson. “Attraction as a linear function of proportion of positive reinforcements.” In: *Journal of personality and social psychology* 1.6 (1965), p. 659.
- [31] Joseph N Cappella and Sally Planalp. “Talk and silence sequences in informal conversations III: Interspeaker influence”. In: *Human Communication Research* 7.2 (1981), pp. 117–132.
- [32] Justine Cassell, Catherine Pelachaud, Norman Badler, Mark Steedman, Brett Achorn, Tripp Becket, Brett Douville, Scott Prevost, and Matthew Stone. “Animated conversation: rule-based generation of facial expression, gesture & spoken intonation for multiple conversational agents”. In: *Proc. of the 21st annual conference on Computer graphics and interactive techniques*. 1994, pp. 413–420.

- [33] Justine Cassell, Hannes Högni Vilhjálmsson, and Timothy Bickmore. “BEAT: the behavior expression animation toolkit”. In: *Life-Like Characters*. Springer, 2004, pp. 163–185.
- [34] Tanya L Chartrand and John A Bargh. “The chameleon effect: The perception–behavior link and social interaction.” In: *Journal of personality and social psychology* 76.6 (1999), p. 893.
- [35] Tanya L Chartrand, William W Maddux, and Jessica L Lakin. “Beyond the perception-behavior link: The ubiquitous utility and motivational moderators of nonconscious mimicry”. In: *The new unconscious* (2005), pp. 334–361.
- [36] Timothy Chklovski and Patrick Pantel. “Verbocean: Mining the web for fine-grained semantic verb relations”. In: *Proceedings of EMNLP*. Vol. 4. 2004, pp. 33–40.
- [37] Kenneth W Church. “Empirical estimates of adaptation: the chance of two norie-gas is closer to $p/2$ than p^2 ”. In: *Proc. of the 18th conference on Computational linguistics-Volume 1*. 2000, pp. 180–186.
- [38] Herbert H. Clark. *Using Language*. Cambridge University Press, 1996.
- [39] Herbert H. Clark and Susan E. Brennan. “Grounding in communication”. In: *Perspectives on socially shared cognition*. Ed. by L. B. Resnick, J. Levine, and S. D. Bahrend. APA, 1991.
- [40] Herbert H. Clark and Deanna Wilkes-Gibbs. “Referring as a collaborative process”. In: *Cognition* 22 (1986), pp. 1–39.
- [41] N. Coupland, J. Coupland, H. Giles, and K. Henwood. “Accommodating the elderly: Invoking and extending a theory”. In: *Language in Society* 17.1 (1988), pp. 1–41.
- [42] S Craig, B Gholson, M Ventura, A Graesser, and the Tutoring Research Group. “Overhearing Dialogues and Monologues in Virtual Tutoring Sessions: Effects on Questioning and Vicarious Learning”. In: *International Journal of Artificial Intelligence in Education* 11 (2000), pp. 242–253.
- [43] Cristian Danescu-Niculescu-Mizil, Michael Gamon, and Susan Dumais. “Mark my words!: linguistic style accommodation in social media”. In: *Proceedings of the 20th international conference on World wide web*. ACM. 2011, pp. 745–754.
- [44] Cristian Danescu-Niculescu-Mizil and Lillian Lee. “Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs”. In: *Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics*. Association for Computational Linguistics. 2011, pp. 76–87.

- [45] Cristian Danescu-Niculescu-Mizil, Lillian Lee, Bo Pang, and Jon Kleinberg. “Echoes of power: Language effects and power differences in social interaction”. In: *Proceedings of the 21st international conference on World Wide Web*. ACM. 2012, pp. 699–708.
- [46] Markus De Jong, Mariët Theune, and Dennis Hofstede. “Politeness and alignment in dialogues with a virtual guide”. In: *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 1*. International Foundation for Autonomous Agents and Multiagent Systems. 2008, pp. 207–214.
- [47] Amit Dubey, Patrick Sturt, and Frank Keller. “Parallelism in coordination as an instance of syntactic priming: Evidence from corpus-based modeling”. In: *Proc. of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*. 2005, pp. 827–834.
- [48] Ondřej Dušek and Filip Jurčiček. “A context-aware natural language generator for dialogue systems”. In: *Proceedings of the SIGDIAL 2016 Conference*. Association for Computational Linguistics. 2016, pp. 185–190.
- [49] David Elson. “DramaBank: Annotating Agency in Narrative Discourse.” In: *LREC*. 2012, pp. 2813–2819.
- [50] David K. Elson. “DramaBank: Annotating Agency in Narrative Discourse”. In: *Proc. of the Eighth International Conference on Language Resources and Evaluation (LREC 2012)*. 2012.
- [51] David K Elson and Kathleen R McKeown. “A tool for deep semantic encoding of narrative texts”. In: *Proceedings of the ACL-IJCNLP 2009 Software Demonstrations*. Association for Computational Linguistics. 2009, pp. 9–12.
- [52] D.K. Elson and K.R. McKeown. “Building a bank of semantically encoded narratives”. In: *Proc. of the Seventh International Conference on Language Resources and Evaluation (LREC 2010), Malta*. 2010.
- [53] Birgit Endrass, Elisabeth André, Matthias Rehm, and Yukiko Nakano. “Investigating culture-related aspects of behavior for virtual characters”. In: *Autonomous Agents and Multi-Agent Systems 27.2* (2013), pp. 277–304.
- [54] Birgit Endraß, Elisabeth André, Matthias Rehm, and Yukiko I. Nakano. “Investigating culture-related aspects of behavior for virtual characters”. In: *Autonomous Agents and Multi-Agent Systems 27.2* (2013), pp. 277–304. doi: 10.1007/s10458-012-9218-5. url: <http://dx.doi.org/10.1007/s10458-012-9218-5>.
- [55] Peter G. Enticott, Patrick J. Johnston, Sally E. Herringa, Kate E. Hoya, and Paul B. Fitzgerald. “Mirror neuron activation is associated with facial emotion processing”. In: *Neuropsychologia* 46 (2008), 2851–2854.
- [56] Christiane Fellbaum. *WordNet*. Springer, 2010.

- [57] Jessica Fidler and Yoav Goldberg. “Controlling linguistic style aspects in neural language generation”. In: *arXiv preprint arXiv:1707.02633* (2017).
- [58] J. E. Fox Tree. “Listening in on monologues and dialogues”. In: *Discourse Processes* 27 (|1999|), pp. 35–53.
- [59] J. E. Fox Tree and S. A. Mayer. “Overhearing single and multiple perspectives”. In: *Discourse Processes* 45.160-179 (|2008|).
- [60] J.E. Fox Tree and J.C. Schrock. “Basic meanings of you know and I mean”. In: *Journal of Pragmatics* 34.6 (2002), pp. 727–747. issn: 0378-2166.
- [61] J.E. Fox Tree and J.C. Schrock. “Discourse Markers in Spontaneous Speech: Oh What a Difference an Oh Makes”. In: *Journal of Memory and Language* 40.2 (1999), pp. 280–295. issn: 0749-596X.
- [62] Kevin Frank. “Posture & Perception in the Context of the Tonic Function Model of Structural Integration: an Introduction”. In: *IASI Yearbook 2007* (2007), pp. 27–35.
- [63] Adrian Furnham. “Language and Personality”. In: *Handbook of Language and Social Psychology*. Ed. by H. Giles and W. Robinson. Winley, 1990.
- [64] Riccardo Fusaroli, Bahador Bahrami, Karsten Olsen, Andreas Roepstorff, Geraint Rees, Chris Frith, and Kristian Tylén. “Coming to terms: quantifying the benefits of linguistic coordination”. In: *Psychological science* 23.8 (2012), pp. 931–939.
- [65] Sudeep Gandhe and David Traum. “An evaluation understudy for dialogue coherence models”. In: *Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue*. Association for Computational Linguistics. 2008, pp. 172–181.
- [66] H. Giles and N. Coupland. “Ethnicity and intergroup communication”. In: (|1991|). Ed. by H. Giles and N. Coupland, pp. 21–42.
- [67] H. Giles and R. Streets. “Communicator characteristics and behavior”. In: *Handbook of interpersonal communication* 2 (1994), pp. 103–161.
- [68] Howard Giles and Peter F Powesland. *Speech style and social evaluation*. Academic Press, 1975.
- [69] Alastair J Gill, Carsten Brockmann, and Jon Oberlander. “Perceptions of alignment and personality in generated dialogue”. In: *Proc. of the Seventh International Natural Language Generation Conference*. 2012, pp. 40–48.
- [70] John J Godfrey, Edward C Holliman, and Jane McDaniel. “SWITCHBOARD: Telephone speech corpus for research and development”. In: *Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on*. Vol. 1. IEEE. 1992, pp. 517–520.

- [71] Andrew Gordon and Reid Swanson. “Identifying personal stories in millions of weblog entries”. In: *Third International Conference on Weblogs and Social Media, Data Challenge Workshop, San Jose, CA*. 2009.
- [72] Andrew Gordon and Reid Swanson. “Identifying personal stories in millions of weblog entries”. In: *Third International Conference on Weblogs and Social Media, Data Challenge Workshop, San Jose, CA*. 2009.
- [73] S. D. Gosling, P. J. Rentfrow, and W. B. Swann. “A Very Brief Measure of the Big Five Personality Domains”. In: *Journal of Research in Personality* 37 (2003), pp. 504–528.
- [74] Jonathan Gratch, Ning Wang, Jillian Gerten, Edward Fast, and Robin Duffy. “Creating rapport with virtual agents”. In: *International Workshop on Intelligent Virtual Agents*. Springer. 2007, pp. 125–138.
- [75] Jonathan Gratch, Louis-Philippe Morency, Stefan Scherer, Giota Stratou, Jill Boberg, Sebastian Koenig, Todd Adamson, Albert A Rizzo, et al. “User-state sensing for virtual health agents and telehealth applications.” In: *MMVR*. 2013, pp. 151–157.
- [76] Nicolas Gueguen, Celine Jacob, and Angeliqne Martin. “Mimicry in social interaction: Its effect on human judgment and behavior”. In: *European Journal of Social Sciences* 8.2 (2009), pp. 253–259.
- [77] Björn Hartmann, Maurizio Mancini, and Catherine Pelachaud. “Implementing expressive gesture synthesis for embodied conversational agents”. In: *International Gesture Workshop*. Springer. 2005, pp. 188–199.
- [78] Bettina Heinz. “Backchannel responses as strategic responses in bilingual speakers’ conversations”. In: *Journal of Pragmatics* 35.7 (2003), pp. 1113–1142.
- [79] Alexis Heloir and Michael Kipp. “EMBR: A realtime animation engine for interactive embodied agents”. In: *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*. IEEE. 2009, pp. 1–2.
- [80] Francis Heylighen and Jean-Marc Dewaele. “Variation in the contextuality of language: an empirical measure”. In: *Context in Context, Special issue of Foundations of Science* 7.3 (2002), pp. 293–340.
- [81] Judith Holler and Katie Wilkin. “Co-speech gesture mimicry in the process of collaborative referring during face-to-face dialogue”. In: *Journal of Nonverbal Behavior* 35.2 (2011), pp. 133–153.
- [82] Zhichao Hu, Gabrielle Halberg,Carolynn Jimenez, and Marilyn Walker. “Entrainment in Pedestrian Direction Giving: How many kinds of entrainment?” In: *Workshop on Spoken Dialog Systems (IWSDS’14)*. 2014.

- [83] Zhichao Hu, Marilyn A Walker, Michael Neff, and Jean E Fox Tree. “Story-telling agents with personality and adaptivity”. In: *International Conference on Intelligent Virtual Agents*. Springer. 2015, pp. 181–193.
- [84] A. L. Hubbard, S. M. Wilson, D. E. Callan, and M. Dapretto. “Giving speech a hand: Gesture modulates activity in auditory cortex during speech perception”. In: *Human Brain Mapping* 30 (|2009|), pp. 1028–1037.
- [85] M.E. Ireland, R.B. Slatcher, P.W. Eastwick, L.E. Scissors, E.J. Finkel, and J.W. Pennebaker. “Language Style Matching Predicts Relationship Initiation and Stability”. In: *Psychological Science* 22.1 (2011), p. 39. issn: 0956-7976.
- [86] E. A. Isaacs and H.H. Clark. “References in conversations between experts and novices”. In: *Journal of Experimental Psychology* 116.1 (|1987|), pp. 26–37.
- [87] Amy Isard, Carsten Brockmann, and Jon Oberlander. “Individuality and alignment in generated dialogues”. In: *Proceedings of the Fourth International Natural Language Generation Conference*. Association for Computational Linguistics. 2006, pp. 25–32.
- [88] Katherine Isbister and Clifford Nass. “Consistency of personality in interactive characters: verbal cues, non-verbal cues, and user characteristics”. In: *International journal of human-computer studies* 53.2 (2000), pp. 251–267.
- [89] Dan Jurafsky, Rajesh Ranganath, and Dan McFarland. “Extracting social meaning: Identifying interactional style in spoken conversation”. In: *Proc. of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. 2009, pp. 638–646.
- [90] Dan Jurafsky, Elizabeth Shriberg, and Debra Biasca. “Switchboard SWBD-DAMSL shallow-discourse-function annotation coders manual”. In: *Institute of Cognitive Science Technical Report* (1997), pp. 97–102.
- [91] A. Kendon. “Gesticulation and speech: Two aspects of the process of utterance”. In: *The relationship of verbal and nonverbal communication* 25 (1980), pp. 207–227.
- [92] Akshay Java Kevin Burton and Ian Soboroff. “The icwsm 2009 spinn3r dataset”. In: (2009).
- [93] I. Kimbara. “Gesture form convergence in joint description”. In: *Journal of Nonverbal Behavior* 32 (|2008|), pp. 123–131.
- [94] Irene Kimbara. “On gestural mimicry”. In: *Gesture* 6.1 (2006), pp. 39–61.
- [95] S. Kita, I. Van Gijn, and H. Van Der Hulst. “Movement phase in signs and co-speech gestures, and their transcriptions by human coders”. In: *Proceedings of the International Gesture Workshop on Gesture and Sign Language in Human-Computer Interaction*. Springer-Verlag, 1998, pp. 23–35.
- [96] Mark L Knapp. “Nonverbal communication in human interaction”. In: (1978).

- [97] Stefan Kopp and Ipke Wachsmuth. “Synthesizing multimodal utterances for conversational agents”. In: *Computer animation and virtual worlds* 15.1 (2004), pp. 39–52.
- [98] William Labov and Joshua Waletzky. “Narrative analysis: Oral versions of personal experience.” In: (1997).
- [99] Marianne LaFrance. “Posture mirroring and rapport”. In: *Interaction rhythms: Periodicity in communicative behavior* (1982), pp. 279–298.
- [100] Jessica L Lakin and Tanya L Chartrand. “Using nonconscious behavioral mimicry to create affiliation and rapport”. In: *Psychological science* 14.4 (2003), pp. 334–339.
- [101] Benoit Lavoie and Owen Rambow. “A Framework for Customizable Generation of Multi-Modal Presentations”. In: *COLING-ACL98*. ACL. Montréal, Canada, 1998.
- [102] John Lee, Finbar Dineen, and Jean McKendree. “Supporting student discussions: it isn’t just talk”. In: *Education and Information Technologies* 3.3-4 (1998), pp. 217–229.
- [103] Grace Lin and Marilyn Walker. “Stylistic Variation in Television Dialogue for Natural Language Generation”. In: *EMNLP Workshop on Stylistic Variation*. 2017.
- [104] Richard Lippa. “The nonverbal display and judgment of extraversion, masculinity, femininity, and gender diagnosticity: A lens model analysis”. In: *Journal of Research in Personality* 32.1 (1998), pp. 80–107.
- [105] Diane J Litman, Carolyn P Rosé, Kate Forbes-Riley, Kurt VanLehn, Dumisizwe Bhembe, and Scott Silliman. “Spoken versus typed human and computer dialogue tutoring”. In: *International Journal of Artificial Intelligence in Education* 16.2 (2006), pp. 145–170.
- [106] Kris Liu, Natalia Blackwell, Jean E. Fox Tree, and Marilyn A. Walker. “21st annual meeting of the Society for Text and Discourse”. In: *A Hula Hoop almost Hit Me!: Running a Map Task in the Wild to Study Conversational Alignment*.
- [107] Kris Liu, Jackson Tolins, Jean E Fox Tree, Marilyn Walker, and Michael Neff. “Judging IVA personality using an open-ended question”. In: *International Workshop on Intelligent Virtual Agents*. Springer. 2013, pp. 396–405.
- [108] Kris Liu, Jean E Fox Tree, and Marilyn A Walker. “Coordinating Communication in the Wild: The Artwalk Dialogue Corpus of Pedestrian Navigation and Mobile Referential Communication.” In: *LREC*. 2016.
- [109] Stephanie M. Lukin, Kevin Bowden, Casey Barackman, and Marilyn A Walker. “A Corpus of Personal Narratives and Their Story Intention Graphs”. In: *Language Resources and Evaluation Conference (LREC)*. 2016.

- [110] Stephanie M Lukin, Lena I Reed, and Marilyn Walker. “Generating Sentence Planning Variations for Story Telling”. In: *16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. 2015, p. 188.
- [111] Stephanie Lukin and Marilyn Walker. “Narrative Variations in a Virtual Storyteller”. In: *Intelligent Virtual Agents* (2015), p. 30.
- [112] Pengcheng Luo, Michael Kipp, and Michael Neff. “Augmenting gesture animation with motion capture data to provide full-body engagement”. In: *Intelligent Virtual Agents*. Springer. 2009, pp. 405–417.
- [113] Pengcheng Luo, Victor Ng-Thow-Hing, and Michael Neff. “An Examination of Whether People Prefer Agents Whose Gestures Mimic Their Own”. In: *Intelligent Virtual Agents*. Springer. 2013, pp. 229–238.
- [114] F. Mairesse and M.A. Walker. “Towards personality-based user adaptation: psychologically informed stylistic language generation”. In: *User Modeling and User-Adapted Interaction* (2010), pp. 1–52. issn: 0924-1868.
- [115] Francois Mairesse and Marilyn Walker. “Personage: Personality generation for dialogue”. In: *In Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL)*. 2007, pp. 496–503.
- [116] Francois Mairesse and Marilyn A. Walker. “Controlling User Perceptions of Linguistic Style: Trainable Generation of Personality Traits”. In: *Computational Linguistics* (2011).
- [117] François Mairesse, Marilyn A. Walker, Matthias R. Mehl, and Roger K. Moore. “Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text”. In: *Journal of Artificial Intelligence Research (JAIR)* 30 (2007), pp. 457–500.
- [118] Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. “The Stanford CoreNLP Natural Language Processing Toolkit”. In: *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. 2014, pp. 55–60. url: <http://www.aclweb.org/anthology/P/P14/P14-5010>.
- [119] D.P. McAdams and J.L. Pals. “A new Big Five: Fundamental principles for an integrative science of personality.” In: *American Psychologist* 61.3 (2006), p. 204. issn: 1935-990X.
- [120] Matthias R. Mehl, Samuel D. Gosling, and James W. Pennebaker. “Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life”. In: *Journal of Personality and Social Psychology* 90 (2006), pp. 862–877.
- [121] Albert Mehrabian. “Significance of posture and position in the communication of attitude and status relationships.” In: *Psychological Bulletin* 71.5 (1969), p. 359.

- [122] Igor A. Melčuk. *Dependency Syntax: Theory and Practice*. Albany, New York: SUNY, 1988.
- [123] Charles Metzging and Susan E Brennan. “When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions”. In: *Journal of Memory and Language* 49.2 (2003), pp. 201–213.
- [124] Lisette Mol, Emiel Krahmer, Alfons Maes, and Marc Swerts. “Converging Hands or Converging Minds?” In: ().
- [125] Y. Moon and C. Nass. “How ”real” are computer personalities?: Psychological responses to personality types in human-computer interaction”. In: *Communication Research* 23 ([1996]), pp. 651–674.
- [126] Roxana Moreno, Richard Mayer, and James Lester. “Life-like pedagogical agents in constructivist multimedia environments: Cognitive consequences of their interaction”. In: *World Conference on Educational Media and Technology*. 1. 2000, pp. 776–781.
- [127] M. Neff, N. Toothman, R. Bowmani, J.E. Fox Tree, and Walker M. A. “Don’t Scratch! Self-adaptors Reflect Emotional Stability”. In: *Intelligent Virtual Agents*. Vol. 6895. Springer. 2011.
- [128] M. Neff, Y. Wang, R. Abbott, and M. Walker. “Evaluating the effect of gesture and language on personality perception in conversational agents”. In: *Intelligent Virtual Agents*. Springer. 2010, pp. 222–235.
- [129] M. Neff, Y. Wang, R. Abbott, and M. Walker. “Evaluating the effect of gesture and language on personality perception in conversational agents”. In: *Intelligent Virtual Agents*. Springer. 2010, pp. 222–235.
- [130] Michael Neff, Michael Kipp, Irene Albrecht, and Hans-Peter Seidel. “Gesture modeling and animation based on a probabilistic re-creation of speaker style”. In: *ACM Transactions on Graphics (TOG)* 27.1 (2008), p. 5.
- [131] Katherine Nelson. “Narrative and self, myth and memory: Emergence of the cultural self.” In: (2003).
- [132] Ani Nenkova, Agustín Gravano, and Julia Hirschberg. “High frequency word entrainment in spoken dialogue”. In: *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers*. Association for Computational Linguistics. 2008, pp. 169–172.
- [133] Roland Neumann and Fritz Strack. “” Mood contagion”: the automatic transfer of mood between persons.” In: *Journal of personality and social psychology* 79.2 (2000), p. 211.
- [134] Kate G Niederhoffer and James W Pennebaker. “Linguistic style matching in social interaction”. In: *Journal of Language and Social Psychology* 21.4 (2002), pp. 337–360.

- [135] R.E. Nisbett. “The trait construct in lay and professional psychology”. In: *Retrospections on social psychology* (1980), pp. 109–130.
- [136] Tsukasa Noma, Norman I Badler, and Liwei Zhao. “Design of a virtual human presenter”. In: *Center for Human Modeling and Simulation* (2000), p. 75.
- [137] W. T. Norman. “Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality rating”. In: *Journal of Abnormal and Social Psychology* 66 (1963), pp. 574–583.
- [138] M. North. “Personality assessment through movement”. In: (1972).
- [139] NPR and Edison Research. *The Smart Audio Report*. <http://nationalpublicmedia.com/smart-audio-report-fall-winter-2017/>. Jan. 2018.
- [140] Daniel S. Paiva and Roger Evans. “A Framework for Stylistically Controlled Generation”. In: *Natural Language Generation, Third International Conference, INLG 2004*. Ed. by Anja Belz, Roger Evans, and Paul Piwek. LNAI 3123. Springer, July 2004, pp. 120–129.
- [141] Gabriel Parent and Maxine Eskenazi. “Lexical entrainment of real users in the let’s go spoken dialog system”. In: *Proceedings Interspeech*. 2010, pp. 3018–3021.
- [142] Fey Parrill and Irene Kimbara. “Seeing and hearing double: The influence of mimicry in speech and gesture on observers”. In: *Journal of Nonverbal Behavior* 30.4 (2006), pp. 157–166.
- [143] Rebecca J. Passonneau and Diane Litman. “Empirical Analysis of Three Dimensions of Spoken Discourse: Segmentation, Coherence and Linguistic Devices”. In: *Computational and Conversational Discourse: Burning Issues - An Interdisciplinary Account*. Ed. by Donia Scott and Eduard Hovy. Springer-Verlag, Heidelberg, Germany, 1996, pp. 161–194.
- [144] J. W. Pennebaker, M. E. Francis, and R. J. Booth. *Inquiry and Word Count: LIWC 2001*. Mahwah, NJ: Lawrence Erlbaum, 2001.
- [145] J. W. Pennebaker and L. A. King. “Linguistic styles: Language use as an individual difference”. In: *Journal of Personality and Social Psychology* 77 (1999), pp. 1296–1312.
- [146] James W Pennebaker and Janel D Seagal. “Forming a story: The health benefits of narrative”. In: *Journal of clinical psychology* 55.10 (1999), pp. 1243–1254.
- [147] Pew Research Center. *Nearly half of Americans use digital voice assistants, mostly on their smartphones*. <http://www.pewresearch.org/fact-tank/2017/12/12/nearly-half-of-americans-use-digital-voice-assistants-mostly-on-their-smartphones/>. Dec. 2017.

- [148] Christian Pietsch, Armin Buch, Stefan Kopp, and Jan de Ruiter. “Measuring Syntactic Priming in Dialogue Corpora”. In: *Empirical Approaches to Linguistic Theory: Studies in Meaning and Structure* 111 (2012), p. 29.
- [149] Robert Porzel. “How computers (should) talk to humans”. In: *How People Talk to Computers, Robots, and Other Artificial Communication Partners* (2006), p. 7.
- [150] Matthias Rehm, Yukiko Nakano, Elisabeth André, Toyoaki Nishida, Nikolaus Bee, Birgit Endrass, Michael Wissner, Afia Akhter Lipi, and Hung-Hsuan Huang. “From observation to simulation: generating culture-specific behavior for interactive systems”. In: *AI & society* 24.3 (2009), pp. 267–280.
- [151] D. Reitter, F. Keller, and J.D. Moore. “Computational modelling of structural priming in dialogue”. In: *Proc. Human Language Technology conference-North American chapter of the Association for Computational Linguistics annual mtg* (2006).
- [152] David Reitter. “Context effects in language production: models of syntactic priming in dialogue corpora”. PhD thesis. University of Edinburgh, 2008. url: <http://www.david-reitter.com/pub/reitter2008phd.pdf>.
- [153] David Reitter, Frank Keller, and Johanna D Moore. “Computational modelling of structural priming in dialogue”. In: *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers*. Association for Computational Linguistics. 2006, pp. 121–124.
- [154] David Reitter and Johanna D Moore. “Predicting success in dialogue”. In: *Annual Meeting-Association for Computational Linguistics*. Vol. 45. 1. 2007, p. 808.
- [155] David Reitter, Johanna D. Moore, and Frank Keller. “Priming of syntactic rules in task-oriented dialogue and spontaneous conversation”. In: *Proceedings of the 28th Annual Conference of the Cognitive Science Society*. Vancouver, Canada: Cognitive Science Society, 2006, pp. 685–690. url: <http://www.david-reitter.com/pub/reitter2006priming.pdf>.
- [156] Laurel D Riek, Philip C Paul, and Peter Robinson. “When my robot smiles at me: Enabling human-robot rapport via real-time head gesture mimicry”. In: *Journal on Multimodal User Interfaces* 3.1-2 (2010), pp. 99–108.
- [157] Ronald E Riggio and Howard S Friedman. “Impression formation: the role of expressive behavior.” In: *Journal of Personality and Social Psychology* 50.2 (1986), p. 421.
- [158] Giacomo Rizzolatti and Michael A Arbib. “Language within our grasp”. In: *Trends in neurosciences* 21.5 (1998), pp. 188–194.

- [159] Zsófia Ruttkay, Claire Dormann, and Han Noot. “Embodied conversational agents on a common ground”. In: *From brows to trust: evaluating embodied conversational agents*. Ed. by Zsófia Ruttkay and Catherine Pelachaud. Norwell, MA: Kluwer Academic Publishers, 2004. Chap. 2, pp. 27–66.
- [160] Ellen B Ryan, Howard Giles, Giampiero Bartolucci, and Karen Henwood. “Psycholinguistic and social psychological components of communication by and with the elderly”. In: *Language & Communication* 6.1 (1986), pp. 1–24.
- [161] Kimiko Ryokai, Cati Vaucelle, and Justine Cassell. “Virtual peers as partners in storytelling and literacy learning”. In: *Journal of computer assisted learning* 19.2 (2003), pp. 195–208.
- [162] M. F. Schober. “Spatial perspective-taking in conversation”. In: *Cognition* 47 ([1993]), pp. 1–24.
- [163] M. F. Schober and H. H. Clark. “Understanding by addressees and overhearers”. In: *Cognitive Psychology* 21.2 ([1989]), pp. 211–232.
- [164] Michael F. Schober. “Different kinds of conversational perspective-taking”. In: *Social and cognitive psychological approaches to interpersonal communication*. Ed. by S. R. Fussell and R. J. Kreuz. Lawrence Erlbaum, 1998, pp. 145–174.
- [165] Ari Shapiro, Petros Faloutsos, and Victor Ng-Thow-Hing. “Dynamic animation and control environment”. In: *Proc. of graphics interface 2005*. Canadian Human-Computer Communications Society. 2005, pp. 61–70.
- [166] Svetlana Stenchikova and Amanda Stent. “Measuring adaptation between dialogs”. In: *Proc. of the 8th SIGdial Workshop on Discourse and Dialogue*. 2007.
- [167] S. Stoyanchev and A. Stent. “Concept form adaptation in human-computer dialog”. In: *Proc. of the SIGDIAL 2009 Conference: The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. Association for Computational Linguistics. 2009, pp. 144–147.
- [168] Satoshi V Suzuki and Seiji Yamada. “Persuasion through overheard communication by life-like agents”. In: *Intelligent Agent Technology, 2004.(IAT 2004). Proc. . IEEE/WIC/ACM International Conference on*. IEEE. 2004, pp. 225–231.
- [169] M. Takala. “Studies of Psychomotor Personality Tests, 1-”. In: (1953).
- [170] Deborah Tannen. *Talking voices : repetition, dialogue, and imagery in conversational discourse*. CUP, 1989.
- [171] A. Tapus, C. Tapus, and M.J. Mataric. “User robot personality matching and assistive robot behavior adaptation for post-stroke rehabilitation therapy”. In: *Intelligent Service Robotics* 1.2 (2008). this is really not a very good paper, pp. 169–183.

- [172] Marcus Thiebaux, Stacy Marsella, Andrew N Marshall, and Marcelo Kallmann. “Smartbody: Behavior realization for embodied conversational agents”. In: *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 1*. International Foundation for Autonomous Agents and Multiagent Systems. 2008, pp. 151–158.
- [173] Avril Thorne, Neill Korobov, and Elizabeth M Morgan. “Channeling identity: A study of storytelling in conversations between introverted and extraverted friends”. In: *Journal of research in personality* 41.5 (2007), pp. 1008–1031.
- [174] Avril Thorne and V. Nam. “The storied construction of personality.” In: *The Cambridge Handbook of Personality Psychology*. Ed. by Kitayama S. and Cohen D. 2009, pp. 491–505.
- [175] Jackson Tolins, Kris Liu, Michael Neff, Marilyn A Walker, and Jean E Fox Tree. “A Verbal and Gestural Corpus of Story Retellings to an Expressive Virtual Character”. In: *Language Resources and Evaluation Conference (LREC)*. 2016.
- [176] Jackson Tolins, Kris Liu, Yingying Wang, Jean E Fox Tree, Marilyn A Walker, and Michael Neff. “A Multimodal Corpus of Matched and Mismatched Extravert-Introvert Conversational Pairs”. In: *Language Resources and Evaluation Conference (LREC)*. 2016.
- [177] Jackson Tolins, Kris Liu, Yingying Wang, Jean E Fox Tree, Marilyn Walker, and Michael Neff. “Gestural Adaptation in Extravert-Introvert Pairs and Implications for IVAs”. In: *Intelligent Virtual Agents*. Springer. 2013, p. 484.
- [178] Marilyn A. Walker, Amanda Stent, François Mairesse, and Rashmi Prasad. “Individual and Domain Adaptation in Sentence Planning for Dialogue”. In: *Journal of Artificial Intelligence Research (JAIR)* 30 (2007), pp. 413–456.
- [179] Ning Wang, W. Lewis Johnson, Richard E. Mayer, Paola Rizzo, Erin Shaw, and Heather Collins. “The Politeness Effect: Pedagogical Agents and Learning Gains”. In: *Frontiers in Artificial Intelligence and Applications* 125 (2005), pp. 686–693.
- [180] Yafei Wang, David Reitter, and John Yen. “Linguistic adaptation in online conversation threads: analyzing alignment in online health communities”. In: *Proceedings of the Fifth Workshop on Cognitive Modeling and Computational Linguistics (at ACL)*. Baltimore, Maryland, USA, 2014, pp. 55–62. url: <http://www.david-reitter.com/pub/yafei2014cmcl.pdf>.
- [181] Yingying Wang and Michael Neff. “The Influence of Prosody on the Requirements for Gesture-Text Alignment”. In: *Intelligent Virtual Agents*. Springer. 2013, pp. 180–188.
- [182] Arthur Ward and Diane Litman. “Automatically measuring lexical and acoustic/prosodic convergence in tutorial dialog corpora”. In: *Proc. of the SLATE Workshop on Speech and Language Technology in Education*. 2007.

- [183] Arthur Ward and Diane Litman. “Dialog convergence and learning”. In: *Frontiers in Artificial Intelligence and Applications* 158 (2007), p. 262.
- [184] Michael White. “Efficient realization of coordinate structures in Combinatory Categorical Grammar”. In: *Research on Language and Computation* 4.1 (2006), pp. 39–75.
- [185] D. Wilkes-Gibbs and H. H. Clark. “Coordinating beliefs in conversation”. In: *Journal of Memory and Language* 31 ([1992]), pp. 183–194.
- [186] Deanna Wilkes-Gibbs and Herbert H Clark. “Coordinating beliefs in conversation”. In: *Journal of memory and language* 31.2 (1992), pp. 183–194.
- [187] Michael Willemys, Cynthia Gallois, Victor J Callan, and Jeffery Pittam. “Accent Accommodation in the Job Interview Impact of Interviewer Accent and Gender”. In: *Journal of Language and Social Psychology* 16.1 (1997), pp. 3–22.
- [188] Yang Xu and David Reitter. “An evaluation and comparison of linguistic alignment measures”. In: *Proc. Cognitive Modeling and Computational Linguistics*. Denver, CO: Association for Computational Linguistics, 2015, pp. 58–67. url: <http://www.david-reitter.com/pub/xu2015evaluation-alignment.pdf>.