

# **Robot Interaction Language**



**Omar Mubin**

**ROILA**

***RObot Interaction LAnguage***

Omar Mubin



# ROILA: RObot Interaction LAnguage

## PROEFONTWERP

ter verkrijging van de graad van doctor aan de  
Technische Universiteit Eindhoven, op gezag van de  
rector magnificus, prof.dr.ir. C.J. van Duijn, voor een  
commissie aangewezen door het College voor  
Promoties in het openbaar te verdedigen  
op woensdag 1 juni 2011 om 16.00 uur

door

Omar Mubin

geboren te Lahore, Pakistan





De documentatie van het proefontwerp is goedgekeurd door de promotor:

prof.dr.ir. L.M.G. Feijs

Copromotoren:

dr. C. Bartneck MTD Dipl. Des.

en

dr. J. Hu PDEng MEng

Typeset with  $\text{\LaTeX}$

Cover design by Christoph Bartneck

A catalogue record is available from the Eindhoven University of Technology  
Library

ISBN: 978-90-386-2505-8

# Acknowledgements

One may think that a PhD project is individual in its nature and solely the effort of a single person. To some extent that is true but I truly believe that the ROILA project could not have been possible without the contribution and hard work of not just me but several other people.

Firstly, I would like to thank the reading committee, comprising of Prof Michael Lyons, Prof Emiel Krahmer and Dr Jacques Terken for providing valuable comments to improve the content of the thesis. At this point I would also like to mention the generous support provided by Steven Canvin of LEGO Minstorms NXT who was so kind to donate 20 Mindstorms boxes to us. The Mindstorms kits were invaluable towards the development of the project. All my colleagues at the Designed Intelligence group deserve special acknowledgement, in particular Prof Matthias Rauterberg and Ellen Konijnenberg. I would also like to thank Alex Juarez for his help with the LaTeX typesetting and for various technical aspects related to the project.

The ROILA evaluation could not have been accomplished without the cooperation of the Christiaan Huygens College Eindhoven. Therefore I would like to thank all the science teachers, Marjolein, Arie and Geert. I would also like to thank the school administration and all the participating school children for making the ROILA curriculum so enjoyable. Moreover, the ROILA evaluation was also made possible due to the efforts of Jerry Muelver (President of North American Ido Society, Inc) and Hanneke Hooft van Huysduynen (from Eindhoven University of Technology). I would like to thank them for their assistance in the design of the ROILA curriculum.

I would also to mention two of my colleagues from the USI program: Abdullah Al Mahmud and Suleman Shahid, with whom I collaborated on several research projects over the past years and learned a lot about HCI in the process.

I would like to dedicate the thesis to my parents and family in Pakistan and also to my wife and my son Aaryan. Your encouragement and care has always spurred me on. Last but not the least, my sincere gratitude goes out to my promotor Prof Loe Feijs, thank you for always believing in me and offering a helping hand and also to my supervisor Dr Christoph Bartneck, for being a source of inspiration and always providing constructive and out of the box supervision. ROILA is yours as much as it is mine. Dr Jun Hu also receives special mention as the newest member of the ROILA team.



# Contents

<b>List of Figures</b>	<b>iv</b>
<b>List of Tables</b>	<b>v</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Speech in HCI . . . . .	1
1.2 How does speech recognition work? . . . . .	2
1.2.1 Why is Speech Recognition difficult . . . . .	3
1.3 Speech in HRI . . . . .	3
1.3.1 Difficulties in mapping dialogue . . . . .	4
1.3.2 Technological Limitations . . . . .	4
1.4 Research Goal . . . . .	4
1.5 Artificial Languages . . . . .	5
1.6 Thesis Outline . . . . .	7
<b>2 An overview of existing artificial and natural languages</b>	<b>9</b>
2.1 Proposing a language classification . . . . .	9
2.2 Overview Schema . . . . .	11
2.3 Morphological Overview . . . . .	11
2.4 Morphological Overview: Discussion . . . . .	16
2.5 Phonological Overview . . . . .	17
2.6 Phonological Overview: Discussion . . . . .	18
2.7 Conclusion . . . . .	20
<b>3 The design of ROILA</b>	<b>21</b>
3.1 ROILA Vocabulary Design . . . . .	21
3.1.1 Choice of Phonemes . . . . .	21
3.1.2 Word Length . . . . .	22
3.1.3 Word Structure: initial design . . . . .	24
3.1.4 Using a Genetic Algorithm to generate ROILA words . . . . .	24
3.1.5 Genetic Algorithm simulations . . . . .	26
3.2 Using a word spotting experiment to evaluate the ROILA vocabulary	26
3.2.1 Participants . . . . .	28
3.2.2 Material . . . . .	28
3.2.3 Pilots and Procedure . . . . .	29
3.2.4 Experiment Design and Measurements . . . . .	29
3.2.5 Results from the word spotting experiment . . . . .	30
3.2.6 Second iteration of evaluating the ROILA vocabulary . . . . .	32

3.2.7	Discussing the results from the word spotting experiment .	33
3.3	A larger ROILA vocabulary and giving it semantics . . . . .	34
3.4	Phonetic Modification of the vowels . . . . .	35
3.5	ROILA Grammar Design . . . . .	36
3.6	The rules of the ROILA grammar . . . . .	37
3.6.1	Parts of Speech . . . . .	39
3.6.2	Names and Capitalization of Nouns . . . . .	39
3.6.3	Gender . . . . .	39
3.6.4	Word Order . . . . .	39
3.6.5	Numbering . . . . .	39
3.6.6	Person References . . . . .	39
3.6.7	Tenses . . . . .	40
3.6.8	Polarity . . . . .	41
3.6.9	Referring Questions . . . . .	41
3.6.10	Conjunctions . . . . .	41
3.6.11	Punctuation . . . . .	41
3.6.12	What the ROILA grammar does not have & some alternatives	42
3.7	Grammar Evaluation . . . . .	43
<b>4</b>	<b>The Implementation of ROILA</b>	<b>45</b>
4.1	Speech Recognition for ROILA . . . . .	45
4.1.1	Overview of Speech Recognition Engines . . . . .	45
4.1.2	Adaptation of an Acoustic Model for ROILA . . . . .	48
4.1.3	Language Model/Grammar Representation for ROILA . . .	49
4.1.4	Setting up the Sphinx-4 Configuration file . . . . .	49
4.1.5	Executing Speech Recognition in Java using Sphinx-4 . . .	50
4.2	Speech Synthesis for ROILA . . . . .	50
4.3	LEGO Mindstorms NXT . . . . .	51
4.4	First ROILA prototype . . . . .	52
<b>5</b>	<b>User Aspects of Evaluating Constrained and Artificial Languages</b>	<b>57</b>
5.1	Case Study I . . . . .	57
5.1.1	Experimental Design . . . . .	58
5.1.2	Game Design . . . . .	58
5.1.3	Procedure . . . . .	58
5.1.4	Interaction in the game via Artificial Languages . . . . .	59
5.1.5	A Learnability Test for Artificial Languages . . . . .	59
5.1.6	Evaluating Emotional Expressions via a Perception Test . .	60
5.1.7	Discussion . . . . .	62
5.2	Case Study II . . . . .	63
5.2.1	Scenario 1: Constrained Languages . . . . .	63
5.2.2	Scenario 2: Artificial Languages . . . . .	67
<b>6</b>	<b>ROILA: Evaluation in Context</b>	<b>71</b>
6.1	ROILA Evaluation at the Huygens College Eindhoven . . . . .	71
6.2	LEGO Mindstorms NXT and ROILA . . . . .	71
6.3	ROILA Curriculum . . . . .	71
6.4	Technical setup . . . . .	73
6.5	ROILA Lessons . . . . .	73

6.5.1 Lesson 1 . . . . .	73
6.5.2 Lesson 2 . . . . .	74
6.5.3 Lesson 3 . . . . .	74
6.6 Homework, the Lesson Booklet and the ROILA exam . . . . .	76
6.7 Discussion: ROILA Curriculum . . . . .	77
6.8 Controlled Experiment comparing ROILA and English . . . . .	77
6.8.1 Research Questions . . . . .	78
6.8.2 Participants . . . . .	78
6.8.3 Experiment Design and Measurements . . . . .	79
6.8.4 Procedure . . . . .	82
6.8.5 Setup . . . . .	82
6.8.6 Game Design . . . . .	83
6.8.7 Results . . . . .	84
6.8.8 Game Performance . . . . .	84
6.8.9 SASSI Score Analysis and Results . . . . .	85
6.8.10 Recognition Accuracy . . . . .	90
6.9 Evaluating the learnability of ROILA . . . . .	95
6.9.1 Pre-test . . . . .	95
6.9.2 Main effects analysis . . . . .	96
6.10 Formulating the Efficiency of ROILA . . . . .	97
6.11 Discussion of Results from the controlled experiment . . . . .	98
<b>7 Conclusion</b>	<b>101</b>
<b>Bibliography</b>	<b>105</b>
<b>A Letter from Christiaan Huygens College Eindhoven</b>	<b>111</b>
<b>B English to ROILA dictionary</b>	<b>113</b>
<b>C ROILA Homework Lessons</b>	<b>125</b>
<b>D List of Publications related to this research</b>	<b>139</b>
<b>E Summary</b>	<b>141</b>
<b>F Omar Mubin: Curriculum Vitae</b>	<b>143</b>



# List of Figures

1.1 Graffiti: Handwriting language for Palm . . . . .	6
2.1 Language continuum . . . . .	10
3.1 Simulation graph for $N=G=W=200$ . . . . .	26
3.2 Graph relating average confusion and average word length where $N=G=W=300$ . . . . .	27
3.3 Recording setup . . . . .	29
3.4 Graph illustrating the relation between word length and recognition accuracy . . . . .	31
3.5 Bar chart showing mean errors difference between English and sec- ond ROILA vocabulary . . . . .	33
4.1 Examples of LEGO Mindstorm robots . . . . .	52
4.2 First robot used to prototype ROILA . . . . .	54
4.3 Blue snowflake microphone . . . . .	54
4.4 System Architechture . . . . .	55
4.5 LEGO resources provided by LEGO Mindstorms NXT, Billund . . . . .	56
5.1 Game Design . . . . .	60
5.2 Children involved in the game . . . . .	61
5.3 Game played in constrained Dutch . . . . .	64
5.4 Experimental setup . . . . .	66
5.5 Children interacting with the iCat in Case Study 2 - Scenario 1 . . . . .	67
5.6 Game played in ROILA . . . . .	68
5.7 Child interacting with the iCat in Case Study 2 - Scenario 2 . . . . .	69
6.1 Robot used in Lesson 1 . . . . .	73
6.2 Robot used in Lesson 2 . . . . .	75
6.3 Students interacting with the robots during the ROILA lessons . . . . .	78
6.4 Participant setup . . . . .	83
6.5 Game setup . . . . .	85
6.6 State Diagram of the game . . . . .	86
6.7 SASSI mean ratings bar chart . . . . .	89
6.8 Example of what the system output video looked like . . . . .	91
6.9 Bar chart indicating mean percentages for recognition accuracy mea- surements . . . . .	94
6.10 Scatter plots relating to ROILA exam scores . . . . .	96

# List of Tables

1.1	Examples of speech recognition errors . . . . .	3
2.1	Major Natural Languages of the World . . . . .	18
2.2	Set of Common Phonemes . . . . .	19
3.1	Initial ROILA phonetic table . . . . .	23
3.2	Vocabulary size spread based on word length . . . . .	23
3.3	Vocabulary size spread based on word length . . . . .	28
3.4	Means table for recognition accuracy for English and ROILA . . . . .	30
3.5	Means Table for recognition accuracy for English and ROILA across native language . . . . .	30
3.6	Means Table for recognition accuracy for English and ROILA across experiment order . . . . .	30
3.7	Means Table for recognition accuracy for English and ROILA across gender . . . . .	31
3.8	Means Table for number of users who got a type of word wrong . . . . .	32
3.9	Sample words from second ROILA vocabulary . . . . .	32
3.10	Means Table for recognition accuracy for English and two versions of ROILA . . . . .	33
3.11	Examples of ROILA words and their pronunciations . . . . .	36
3.12	Subsection of the QoC Matrix . . . . .	38
3.13	Sample ROILA sentence showing SVO Word Order . . . . .	39
3.14	Sample ROILA sentences related to Grammatical Numbering . . . . .	40
3.15	Sample ROILA sentence showing the use of pito (I) . . . . .	40
3.16	Sample ROILA sentence showing the use of liba (he) . . . . .	40
3.17	Sample ROILA sentence showing ROILA tenses . . . . .	40
3.18	Sample ROILA sentence showing the representation of polarity in ROILA . . . . .	41
3.19	Sample ROILA sentence showing the use of biwu in ROILA . . . . .	41
3.20	Sample ROILA sentence showing the use of sowu in ROILA . . . . .	41
3.21	Sample ROILA sentence showing how perfect tenses can be stated in ROILA . . . . .	42
3.22	Sample ROILA sentence showing how cases can be represented in ROILA . . . . .	42
3.23	Sample ROILA sentence showing how prepositions can be represented in ROILA . . . . .	43
3.24	Examples of ROILA and English sentences used in the grammar evaluation experiment . . . . .	43

3.25 Means Table for word accuracy for ROILA and English . . . . .	44
4.1 Overview of Speech Recognizers . . . . .	47
4.2 ROILA commands used in the prototype . . . . .	53
4.3 Specifications of the Blue snowflake microphone . . . . .	55
5.1 Constrained Dutch Sentences . . . . .	65
5.2 ROILA Sentences . . . . .	69
6.1 Vocabulary employed in Lesson 1 . . . . .	74
6.2 Vocabulary employed in Lesson 2 . . . . .	75
6.3 ROILA Exam Means table for selected and non selected students . .	79
6.4 Commands that could be used in the game . . . . .	84
6.5 T-Test result and means table for balls shot and goals scored . . . .	85
6.6 Cronbach Alphas for the 6 Factors . . . . .	87
6.7 Means table for SASSI ratings across gender and class group . . . .	87
6.8 ANOVA table for SASSI ratings across gender and class group . . . .	87
6.9 Means table for SASSI ratings across experiment order . . . . .	88
6.10 ANOVA table for SASSI ratings across experiment order . . . . .	88
6.11 ANOVA and Mean-Std.dev table for SASSI main effects . . . . .	89
6.12 ANCOVA table for SASSI main effects after including game perfor- mance as a covariate . . . . .	90
6.13 Means table for recognition accuracy measurements across gender and class group . . . . .	92
6.14 ANOVA table for recognition accuracy measurements across gender and class group . . . . .	92
6.15 Means table for recognition accuracy measurements across experi- ment order . . . . .	92
6.16 ANOVA table for recognition accuracy measurements across experi- ment order . . . . .	93
6.17 Results for regression model for days between last ROILA lesson and day of experiment and recognition accuracy measurements . . . . .	93
6.18 Means table relating total ROILA commands with number of days between 3rd lesson and day of experiment . . . . .	93
6.19 Means and ANOVA table for recognition accuracy analysis . . . . .	94
6.20 ANCOVA table for recognition accuracy main effects after including game performance as a covariate . . . . .	95
6.21 ROILA Exam Score Means and Std.devs across Gender and Class group	95

---

# Introduction

---

Robots are becoming an integral part of our life and state of the art research has already been contemplating into the domain of social robotics (Fong, Nourbakhsh, & Dautenhahn, 2003). Studies have investigated various factors, questions and controversial issues related to the acceptance of robots in our society. We are already at a juncture, where robots are deeply engrossed in our community and importance must now be levied onto how can we as researchers of Human Robot Interaction (HRI); provide humans with a smooth and effort-less interaction with robots. Organizational studies have advocated the fact that robots are a part and parcel of nearly every domain of our society and their use is growing in large numbers (Department, 2008). Robots are deployed for the in various diverse domains such as Entertainment, Education, Health, Search and Rescue Acts, Military and Space Exploration (Goodrich & Schultz, 2007). Given their increasing commercial value it is not very surprising that the emphasis in recent times has been to improve and enhance the user experience of all humans who are directly and indirectly affected by them. Speech is one of the primary modalities utilized for Human Robot Interaction and is a vital means of information exchange (Goodrich & Schultz, 2007). Therefore, improving the performance of speech interaction systems in HRI could consequently better user-robot-interaction. Before discussing speech interaction specifically with respect to Human Robot Interaction, it is worthwhile pondering over the domain of Speech Interaction and Dialogue based systems in the wider area of Human Computer Interaction (HCI).

## **1.1 Speech in HCI**

Within the context of Multimodal Human Computer Interaction, Speech Interaction is one of the interaction modalities. In principle, speech interaction is naturally intuitive, easy to use and natural language interaction requires little or no learning. But the scales are tipped over in a hurry when there is an error or a break down, leading to frustration and irritation from the user; consequently the expectations of the same user are not met. It has been pointed out in (Atal, 1995) that one of the biggest challenges faced by speech recognition is when the conversation being tracked is natural and spontaneous. Other limita-

tions pointed out include the main argument specified in (Shneiderman, 2000) that the use of speech as a modality interferes in the performance of other simultaneous tasks. Robustness and Accuracy are other issues which attract attention and critique (Chen, 2006). It would be very interesting to investigate when and if accuracy is the most desirable. In certain situations, such as in health related interfaces, accuracy would be of the utmost importance. However, speech recognition errors in the context of a game might in fact add an extra dimension to the game play.

There has been large debate pressing for and against the use of Speech in HCI (Human Computer Interaction) systems (James, 2002). Empirical research has analyzed and compared Speech over conventional tangible forms of input to a system (Koester, 2001), (Sporka, Kurniawan, Mahmud, & Slav, 2006). Another reason why Speech is brought forward as an interesting modality within HCI is that it is possible to represent emotions via speech (Cahn, 1990) and that it can have emotional impact if the interaction mechanism is designed accordingly. Furthermore, studies such as (Fry, Asoh, & Matsui, 1998) ascertain that speech is one of the most natural yet practical solutions, especially in situations where the user does not have to learn a formal programming syntax. Speech is a very effective interaction technique and mechanism in assistive technologies for the handicapped. Disabled such as the blind or the physically impaired can interact with various products by using speech only (Bellik & Burger, 1994), (Pitt & Edwards, 1996). Products that use a speech interface (comprising of both speech recognition and speech synthesis) are gaining in ascendancy. Their application domains are several, for e.g. Navigation systems for automobiles (Geutner, Denecke, Meier, Westphal, & Waibel, 1998), as tourist guides (Yang, Yang, Denecke, & Waibel, 1999), and telephone based information access systems (Rosenfeld, Olsen, & Rudnick, 2001).

### **1.2 How does speech recognition work?**

To understand that speech recognition is not a simple task we need to understand briefly how it works and what kind of challenges it faces on each step. The mechanism is summarized from (Lee, Soong, & Paliwal, 1996). Speech recognition initially involves a speech signal as input. The input is basically an analogue sound wave which is then processed by the recognizer so that it is converted into digital machine readable format. The input contains utterances by the user, if any and it may also include ambient sound or purely environmental noise. Needless to say this can hamper the recognition accuracy. The speech system then tries to find a suitable match based on the information it has about the language and the context. This is in the form of a grammar and an acoustic model. The grammar operates on a word level within sentences, i.e. it describes how words may complete sentences. The acoustic model operates on a syllable level, i.e. how individual sounds combine to produce complete words, where individual sounds are also called phonemes. In summary, there exist two outlets of erroneous recognition, either at the grammar or at the phonetic level. The recognizer then spurs out what it computes was said to it, i.e. it makes a guess as it can never be sure. It may also be that the system

concludes nothing was said when something was said or vice versa. Usually speech recognizers also state their confidence with every recognition guess that they make.

### 1.2.1 Why is Speech Recognition difficult

The limitations prevailing in current speech recognition technology for natural language is a major obstacle behind the unanimous acceptance of Speech Interfaces (Chen, 2006). Existing speech recognition is just not good enough for it to be deployed in natural environments, where the ambience influences its performance. Certain properties of natural languages make them difficult for a machine to recognize them. Homophones are prime examples of such dilemmas, i.e. words that sound almost the same but have different meanings. Note that this only means that when a word is said by a user the machine thinks another word was said which is acoustically similar. Other problems that a speech recognizer faces for natural languages is detecting where the word boundaries lie in a sentence because there are multiple ways to combine the sounds uttered by the speaker. Recognizing continuous speech is even more difficult when the machine has to deal with different dialects, i.e. users having different native languages. To give a perspective on the kind of consequences a user may find him/her self as a result of inaccurate recognition, we give some interesting examples-extracted from (Typewell, 2011), some of which are quite historical in being quoted in speech recognition technology research (see Table 1.1).

What was said	What was recognized
That's speech recognition	That's peach wreck in kitchen
Senior years	Seen your ears
It can't work	It can work

Table 1.1: Examples of speech recognition errors

It is clear and evident that while Speech provides an easy and non physical input modality, yet various issues arise pertaining to the applicability of speech, such as ambient noise, cultural limitations, dialect, cognitive overload, etc. In the next section, we will present an overview of Speech based systems in Human Robot Interaction, the predicaments faced by such systems and what the future holds in terms of designing a Robot Interaction Language.

## 1.3 Speech in HRI

Some researchers in HRI have concentrated on designing interaction which can provide or at least to some extent, imitate a social dialogue between humans and a robot. An overview of state of the art research in dialogue management systems unearths several hindrances behind the adoption of natural language for robotic and general systems alike. The challenges faced when using speech interaction would be the same regardless if the user talks to a robot, machine or a computer.

### 1.3.1 Difficulties in mapping dialogue

Dialogue Management and Mapping is one of the popular techniques used to model the interaction between a user and a machine or a robot (Fry et al., 1998). However the inherent irregularity in natural dialogue is one of the main obstacles against deploying Dialogue Management systems accurately (Churcher, Atwell, & Souter, 1997). A conversation in natural language involves several ambiguities that cause breakdown or errors. These include issues such as turn taking, missing structure, filler utterances, indirect references, etc. There have been attempts to solve such ambiguities by utilizing non verbal means of communication. As reported in (Hanafiah, Yamazaki, Nakamura, & Kuno, 2004), a robot tracks the gaze of the user in the case when the object or the verb of a sentence in a dialogue may be undefined or ambiguous. A second argument related to the difficulties in mapping dialogue is which approach to adopt when building a dialogue management system. Several approaches exist, such as state based, frame based and plan or probabilistic based, with an increasing level of complexity. A state based approach is one in which, the user input is predefined and so the dialogue is fixed. Consequently there is limited flexibility in a state based approach. On the other end of the scale are probabilistic approaches that allow dynamic variations in dialogue (Bui, 2006). It has been argued by (Spiliotopoulos, Androutsopoulos, & Spyropoulos, 2001) that for most applications of Robotics, a simple state based or frame based approach would be sufficient. However a conflict arises when it is important to support an interaction which affords a natural experience. In (Lopes & Teixeira, 2000) it is stated that a mixed initiative dialogue, that is more natural than a master slave configuration, can only be sustained by adopting a probabilistic approach, which is as stated before, more complex. The hardest dialogue to model is one in which the initiative can be taken at any point by any one.

### 1.3.2 Technological Limitations

The hardware platform of the robot and the speech recognition engine can be out of sync, causing uncertainty to the user (Kulyukin, 2006). This has been precisely the reason why some HRI researchers have concentrated more on using speech more as an output modality instead of as a form of input. As a direct after effect of un-synchronization, both speech recognition and generation or synthesis is far from optimal.

As a consequence of the prior discussed problems miscommunication occurs between the user and robot. The mismatch between humans' expectations and the abilities of interactive robots often results in frustration. Users are disappointed if the robot cannot understand them properly even though the robot can speak with its mechanical voice. To prevent disappointment, it is important to match the communication skills of a robot with its perception and cognitive abilities.

## 1.4 Research Goal

Recent attempts to improve the quality of the technology of automatic speech recognition for machines have not advanced enough (Shneiderman, 2000).

Generally in speech interfaces the focus is on using natural language (constrained or otherwise). Due to mainly technical difficulties the machine does not always have an easy time recognizing natural language resulting in a frustrating experience for the users. It is perhaps time to explore a new approach to the problem. We need to find a different balance between, on the one hand, allowing users to speak freely, which is good for the users, but difficult for the machines, and on the other hand, constraining the users, which is good for the machines, but difficult for the users. But we should not be dismissing the option of constraining the users too quickly. A speech system that constraints the users would offer a higher recognition accuracy, which in turn is also good for the users. The main question is if we can find a new balance that offers a better trade-off than the current state of the art systems.

This thesis presents such a new balance by proposing a new artificial language named ROBot Interaction Language (ROILA), created using the methodology of *research through design* (Zimmerman, Forlizzi, & Evenson, 2007). The two conflicting requirements for ROILA is to, on the one hand, be easy for humans to learn and speak, and on the other hand, be easy for the machines to recognize. An example for this conflict is the word length. Speech recognizers are more accurate for long words (Hämäläinen, Boves, & De Veth, 2005), which are difficult to learn and speak. Humans prefer short words, since they are more efficient and easier to remember.

In addition, in this project we do not extensively deal with Speech Synthesis. Providing text to speech with natural prosody is a complete research area in itself. Later on in the thesis we will reveal our efforts with Speech Synthesis in the project but this was only as a means of providing a wholeness to our prototype. To reiterate, our focus is on improving speech recognition accuracy by not providing new algorithms but by giving the machine or robot input which is easy to recognize. Another aspect that we did not wish to focus on extensively was the effect of contextual information on the accuracy of speech recognition. Therefore we aimed to design an artificial language that would not be dependent on semantics and consequently we could adopt any context of use for ROILA.

## **1.5 Artificial Languages**

An artificial language as defined by the Oxford Encyclopedia is a language deliberately invented or constructed, especially as a means of communication in computing or information technology. Recent research in speech interaction is already moving in the direction of artificial languages, as stated in (Rosenfeld et al., 2001), constraining language is an important method of improving recognition accuracy. Even human beings are known to vary their tone or prosody depending on the environmental circumstances. After all we know that humans alter their language when they talk to infants, pets or non-native speakers.

In (Tomko & Rosenfeld, 2004) the user experience of an artificially constrained natural language - Speech Graffiti was evaluated within a movie-information dialog interface and it was concluded that 74% of the users found



it more satisfactory than natural language. In addition, it was ascertained that Speech Graffiti was also more efficient in terms of time. The field of handwriting recognition has encountered similar results. The first recognition systems for handheld devices, such as Apple's Newton were nearly unusable. Palm solved the problem by inventing a simplified alphabet called Graffiti, which was easy to learn for users and easy to recognize for the device (see Figure 1.1).

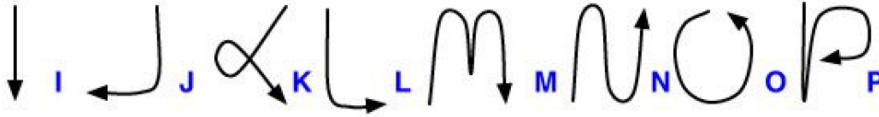


Figure 1.1: Graffiti: Handwriting language for Palm

In linguistics, there are numerous artificial languages (for e.g. Esperanto, Interlingua) which attempt to make communication between humans easier and/or universal. These languages also simplify the vocabulary and grammar, similar to the approach of Graffiti. To the best of our knowledge there has been little or no attempt to optimize a spoken artificial language for automatic speech recognition, besides limited efforts from (Hinde & Belrose, 2001) and (Arsoy & Arslan, 2004). Both endeavours were for vocabularies of limited size and no formal evaluations were carried out. Moreover one cannot term the afore-mentioned efforts as languages as they only comprised of isolated words and not sentences.

We acknowledge the trade-off factor of humans having to invest some energy in learning a new language like ROILA. Ofcourse it would be perfect if speech technology could understand natural language without any problems but this has not yet been achieved. However, by designing an artificial language we are faced with the effort a user has to put in learning the language. Nevertheless, we wish to explore the benefits that an artificial language could provide if it's designed such that it is speech recognition friendly. This factor might end up outweighing the price a user has to pay in learning the language and would ultimately motivate and encourage them to learn it. We could also argue that humans have adaptable instincts and would in the long term be able to use artificial languages to talk to machines or robots.

Another criticism that might be levied on ROILA is that many artificial languages were created already but not many people ended up speaking them. Where our approach is different is that we aim to deploy and implement our artificial language in machines and once a large number of machines can speak the new language it could encourage humans to speak it as well. With just one system update of the most common operating system, a critical mass of speakers could become available. In addition, ROILA does not necessarily have to be restricted to robots only, but it could also be applied to any behavioral products that employ speech interaction.

## **1.6 Thesis Outline**

The format of the thesis follows a standard HCI design approach, i.e., initial investigation, design, implementation and evaluation. The second chapter of the thesis overviews existing artificial languages and attempts to extract linguistic commonalities amongst them and also in comparison to natural languages. The third chapter details the design of ROILA and explains the various iterations involved within the design stage. The fourth chapter explains the implementation of the ROILA into prototypes and gives an introductory example. The fifth chapter ascertains subjective impressions of users while interacting in constrained or artificial languages. The sixth chapter describes the ROILA evaluation carried out at a local school in Eindhoven, The Netherlands, where high school children learnt ROILA in a specially designed curriculum and used it to interact with robots. The main contributions of the thesis and the future prospects are rounded off in the last chapter.



---

## An overview of existing artificial and natural languages

---

Before attempting to design our very own artificial language it was imperative that we overviewed existing artificial languages to gain an understanding about them and their properties. Therefore in this chapter we present a morphological and phonological overview of artificial languages, individually and also in contrast to natural languages. We chose nine major artificial languages as the basis of our overview and the majority of those were international auxiliary languages. Our selection of languages was based on their popularity and availability of authentic information about them, such as dictionaries or official websites.

We also tried to ascertain the design rationale of artificial languages, i.e. why were they created? Could we learn something from them specifically or the methods used to create them? We discovered that Artificial Languages have been developed for various reasons. The primary one being universal communication i.e. to provide humans with a common platform to communicate, other reasons include, reducing inflections and irregularity from speech and introducing ease of learnability.

The morphological overview showed that there are two major grammatical strategies employed by artificial languages. The phonological overview was done on the basis of a common phoneme set from natural languages. Most artificial languages were shown to have phonetic similarities with Germanic languages.

### 2.1 Proposing a language classification

As a first step in our research on languages, we wished to determine the various types of artificial languages and attempt to classify them. In order to accomplish this we analyzed various artificial languages and extending from (Janton, 1993) we proposed the following language continuum (see Figure 2.1). Constrained languages were determined to have two main categories which differed

by the manner in which the vocabulary was altered. In Type 1 with languages such as Basic English the vocabulary is just reduced in size but Type 2 languages adopt the strategy of changing the words within the vocabulary as in Pidgin or Creole languages. Examples of pidgin languages could be the fictional language for children by Kalle and Astrid, where the syllable structure of words is actually changed by inserting extra vowels.

Artificial Languages were observed to have four basic types, which are well described in (Janton, 1993). An artificial language can have naturalistic derivations or be completely artificial in nature. The first level of categorization is whether the artificial language in question inherits any linguistic properties from natural languages. If the artificial language is completely deviant from existing natural languages on all accounts (i.e. grammar and vocabulary) it is termed as A priori, for which a prime example could be Klingon. We discuss the traits of Klingon in detail later on in this chapter.

If an artificial language inherits some traits from natural languages it is termed as A posteriori. If the artificial language is completely based on natural languages it is termed as fully naturalistic, examples being Interlingua. If the vocabulary of the artificial language is based on natural languages but not its grammar it is termed as schematic, with Volapük an example. Artificial languages can also be partly naturalistic and partly schematic such as Esperanto. Note that this classification is quite broad and not distinctively comprehensive, i.e. a particular language may fall across two categories. In summary, a particular language could be placed in any of the eight categories.

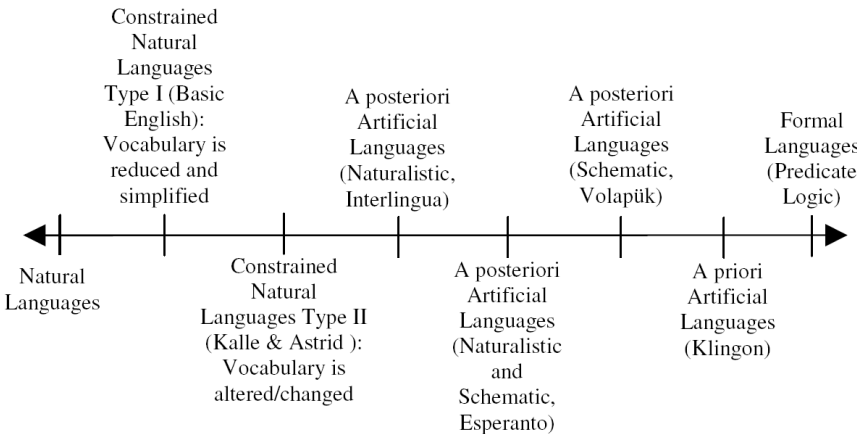


Figure 2.1: Language continuum

## 2.2 Overview Schema

The first step in the conducting the overview was to identify which languages would be considered in the analysis. We focused mainly on international auxiliary languages, i.e. languages that were designed to make communication between humans easier, especially if they did not speak the same language. This decision was based on the goals of ROILA, i.e. we wished to design a language which was easier to learn for humans, which we believed auxiliary languages were. Moreover, we also believe that human-robot interaction has some aspects which are similar to human-human interaction therefore auxiliary languages would be the way to go.

Once we had decided to delve into auxiliary languages the next step was to actually choose specific languages from them. Choosing the set of artificial languages was an important decision and this was based on a number of factors. These included selecting artificial languages that had sufficient information available about them from authentic sources for e.g. dictionaries, official websites, or if they had generated some research interest and/or had a reasonable number of speakers. Artificial languages that were merely constructs of a single author and spoken by hardly anyone besides the author were not considered. Therefore we selected the following artificial languages for our overview: Loglan (Brown, 2008), Esperanto (Janton, 1993), Toki Pona (Kisa, 2008), Desa Chat (Davis, 2000), Ido (ULI, 2008), Glosa (Springer, 2008), Interlingua (Mardegan, 2008), Volapük (Caviness, 2008) and Klingon (Shoulson, 2008). Klingon was the odd one out as it is not an *A posteriori* language.

The final step of the overview was to define a classification scheme, for which existing schemas for natural languages were borrowed and adapted to artificial languages. Various encyclopedias such as (David, 1997) define the major properties of a language, via which we divided our schema into two major categories: Morphology/Grammar and Phonology. Given the initial research we conducted we short listed the afore-mentioned nine artificial languages for further research. We first present the overview along the lines of morphology and subsequently we present a phonological overview.

## 2.3 Morphological Overview

Morphology is the study of the structure of solitary words. The smallest meaningful elements into which words can be analyzed are known as morphemes (David, 1997). Hence in very simple terms morphology can be stated as the grammar of the language and its syntax. The first step in classifying a language on the basis of grammar is stating its grammar type. As indicated by (Malmkjaer & Anderson, 1991), a language can have three grammar types. The first is Inflectional where affixes are added as inflections and they independently do not serve a purpose, for e.g. Latin which is heavily inflected and English which is less so. The second major grammar type is Agglutinating where every affix has one meaning, for e.g. Japanese, and the last major grammar type is Isolating where no suffixes or affixes are added, but in fact meanings

are modified by inserting additional words also known as word markers, for e.g. Chinese. An affix can indicate a wide degree of information, for e.g. aspect, case, numbering, tense, gender etc. We have utilized the overview schema presented in (David, 1997) for natural languages, to morphologically overview artificial languages. In it important grammatical variables are described which entail the grammatical properties of a language, for e.g. aspect, case, tense, number, mood, etc. In order to clarify the context of how we have interpreted the grammatical properties we briefly describe them next.

**Aspect** relates to the nature of the tense, referring to the duration, occurrence and completeness of the tense. Types of aspect include: Perfective (single occurrence that has occurred), Imperfective, and Prospective.

**Case** exhibits the role a noun plays in the sentence, in terms of who is the subject (direct or indirect), object or possessor. The inflection can take place through the noun itself or via pronouns or adjectives. The major types of case in most modern languages include: Subjective or Nominative Case (I, he, she, we), the accusative/dative case (me, him, her) and the genitive case which indicates possession (ours, mine). Older languages such as Latin have much more case types

**Gender** tends to inflect nouns in various languages. This is usually done via adding a suffix to the noun in the case of an inflecting language or in isolating languages it can be expressed by the verb or the pronoun. Nouns are classified into groups such as Male, female, inanimate, animate and neutral.

**Mood/Modality** describes the way the action took place (fact), if it indeed took place (uncertain), or should take place (likelihood). Modality is related to verbs only. Types of mood include: Indicative, Subjunctive (might or desired to happen), and Imperative (must happen).

**Number** is a grammatical category that highlights the total number of nouns/objects. It can be expressed by inflecting the nouns itself only or by inflecting nouns and verbs or pronouns. Typical categories of number include: Singular and Plural, others being dual or trial indications.

**Voice** refers to the relationship between the verb and the subject and object in the sentence. It refers to who did the action related to the subject: him/herself (active), or someone else (passive). Besides active and passive voice, other types of voice are: causative and neutral.

**Person** is an identification or reference to who is the speaker or addressee in a situation. It is typically represented by pronouns and affects verbs. It has the ability to represent the following participants: first, second, third or fourth person.

**Grammatical Tense** refers to the time at which the action of the verb took (past), is taking (present) or will take place (future). Variants also exist, e.g. of

the perfect or imperfect type.

**Grammatical Syntax** or Word Order determines the sequence of words within a sentence, with respect to the subject, verb and object. The possible combinations are: SVO, SOV, VSO, VOS, OVS, OSV and free order.

Now we describe each of the nine artificial languages and present an overview about them based on these grammatical properties wherever applicable. The source of information for all the nine artificial languages has been stated earlier on this chapter.

**Desa Chat** is an artificial language designed to be amenable to computer processing. It has been designed to make use of language processing techniques. It has a long term goal of supporting international communication. It has been mainly derived from Esperanto and attempts to remove whatever irregularities existing in Esperanto. It has a similar alphabet as English having 5 vowels and 21 consonants. Its vocabulary size has been estimated to be larger than 5000 words and it is known to have 105 phonemes. Nouns, verbs, adjectives, adverbs and pronouns make a larger part of the Desa Chat vocabulary. The grammar of Desa Chat is isolating in nature and it supplies references to the possessive case. The major classes of gender (male, female and inanimate) are prevalent in Desa Chat. Moreover, singular and plural indications of grammatical count exist. Desa Chat distinguishes between first, second and third person as well as between the past, present and future tense. It too adopts the common SVO word order.

**Esperanto** is unanimously the most known and spoken artificial language. It is known to have between 1-2 million speakers. It is also referred to as an auxiliary language as it attempts to achieve a goal of universal communication. It is based and derived upon several natural languages, mostly in the Germanic and Romanic groups. However, it is known to be a semi schematic and semi naturalistic language. Its alphabet consists of 5 vowels and 22 consonants with the number of phonemes being 34. It has all common word types: nouns, verbs, adjectives, adverbs, pronouns, prepositions and conjunctions. It has an interesting grammar type as it comprises of less heavy inflections as compared to natural languages. In most places its grammar is stated to be agglutinating. Grammatical aspect is not required in Esperanto. Grammatical case is fulfilled by the nominative and the accusative types. With respect to the representation of Gender, both male and female are supported but there is no category of an inanimate class. The most common modality is of the imperative type. Grammatical number is represented via singular and plural inflections and voice by active and passive references. Esperanto distinguishes between first, second and third person as well as between the past, present and future tense. Word order is rather flexible and there is no mandatory word order that is required to be adhered to. Word order in Esperanto is optional but whenever used it follows the SVO standard.

**Glosa** is also one of the auxiliary constructed languages promoting universal communication. It is well documented as an isolating language, since it is



free from inflections. Words in Glosa stay in their original format, regardless of whether they are nouns or verbs. Therefore the same word, unchanged can act as noun or a verb. Operator words or word order provide most grammatical functions, with every word being affected and modified by its predecessor. Its vocabulary is derived from Greek and Latin and has a sentence structure that is similar to English. It is an *a posteriori* language of the semi-naturalistic type. It has the standard set of 5 vowels and 21 consonants rendering a vocabulary size of between 1000 to 2000 words, having the usual word classes of nouns, verbs, adverbs, pronouns, prepositions, conjunctions, etc. Typically modifiers are used to indicate grammatical number and gender. As stated prior, Nouns or verbs are not inflected. Using modifiers the common categories of male/female references and singular/plural count are permissible. Similarly particles or modifiers allow expression of tenses and aspect. Particles exist for past and future tense, but there is none for the present tense. Individual particles occur for all three aspect conditions of perfective, imperfective, and prospective. By modifying the word order, passive voice is utter-able, as the receiver gets a mention at the beginning of the sentence. The conventional sentence has active voice emerging from its verb phrases. Modality of actions is possible in the imperative and subjunctive forms. Pronouns provide references to the first, second and third person. It is known that Glosa has a phonetic spelling and words are built on the consonant-vowel structure (CV, CVCV, etc), to ensure ease of pronunciation. Sentences in Glosa are also built using the SVO word order.

**Ido** is another auxiliary constructed language based on the goal of providing communication between speakers having different linguistic backgrounds. The design of Ido is based on Esperanto. Ido is a language of semi naturalistic type having influences from Romance languages. The number of speakers of Ido is believed to be around several thousand. It follows the conventional Latin alphabets, having 5 vowels and 21 consonants. The grammar of Ido is agglutinating. Ido has a grammar somewhat simplified from that of Esperanto and has no irregularities or special cases. It provides the standard word types of nouns, verbs, adjectives, adverbs, pronouns, etc. Generally, agreement in number and gender is not imposed in sentences of Ido, therefore adjectives and verbs do not vary depending on the number or gender of the context. Besides male and female references of gender, Ido also has a non-gender category for nouns. Grammatical number in Ido is represented via singular and plural inflections by adding appropriate affixes to the root noun. The pronouns of Ido provide references to the first, second and third person and singular and plural first person pronouns have been made phonetically more unique than Esperanto. All three levels of verb tense are expressed, as are modality of actions in imperative and subjunctive forms. Word order is generally typical of English word order, namely of the SVO model.

**Interlingua** is yet another one of the constructed languages developed for an auxiliary purpose. It is a language that is purely naturalistic in derivation, derived from various natural languages of the world, especially the Romance languages. The main aim of the language is to remove irregularity from natural languages. Interlingua comprises of 5 vowels and 21 consonants. The grammar

of Interlingua tries to free itself from inflections and is primarily agglutinating. Hence, verbs are not inflected by aspect or gender. Modality is covered by the indicative type only. Grammatical number is represented in nouns only with the affixes appended depending on the last consonant of the noun. Plural and Singular references are permissible. Pronouns are responsible for the two grammatical case inflections which are normally used: nominative and genitive. Pronouns also provide references to the first, second and third person. A gender distinction is present in the third person. Word order is again SVO.

**Klingon** is fictional in nature and belongs to the Star Trek fame. The design rationale behind Klingon was that every language must have a cultural ideology as its motivation and justification. Klingon is an apriori language and therefore does not have strong influences from natural languages. It has 5 vowels and 21 consonants, rendering a total of 26 phonemes. The grammar type of Klingon is agglutinating. The Klingon verbs do not represent tenses but grammatical aspect is represented in all three forms: perfective, imperfective and prospective. Klingon verbs also indicate two modalities: imperative and indicative. Grammatical number in nouns is characterized conventionally. Grammatical gender via nouns has a variant notion in Klingon, it does not indicate gender but rather three unique categories: can the object in question speak, is it a body part or neither. Both active and passive voice is present in Klingon. It is one of the rare languages that deviates from the SVO word order. It uses a reverse ordering of OVS.

**Loglan** is one of the well known constructed languages of the engineered type. One of the primary reasons why it was created was to test the Sapir-Whorf hypothesis. The hypothesis states that the language one speaks influences the cognitive thought process of the speaker. Moreover, Loglan was created on the basis of simplicity and aimed to incorporate the principles of phonetic spelling and regularity. The derivation type of Loglan is schematic. Loglan is also referred to a logical language in some quarters and it is known to derive its morphemes from natural languages using statistical methods. Loglan uses the latin alphabet having 17 consonants and 6 vowels. The size of its vocabulary is known to be between 9000 to 12,000 words. The phonemes existing in Loglan amount to 27. The grammar of Loglan is of the isolating type. Loglan formally does not make any distinction between nouns, verbs or adjectives and uses predicates instead. It is extremely flexible in the sense that Grammatical Person, Case and Gender are all optional and not required. Moreover, its predicate paradigm is also free from time and hence no tense forms are used. As far as Grammatical Number is concerned, the same word can refer to both singular and plural. Loglan does include both the active and passive voice. The primary word order that it uses is SVO.

**Toki Pona** has a design rationale of simplicity and attempts to focus on simple concepts only. It is known to have several hundred speakers. It derives some of its properties from natural languages but has adapted them and is therefore a schematically derived language. It has 14 phonemes only, 5 vowels and 9 consonants. Its total number of phonemes is also 14 as it does not distinguish between long and short vowels. The size of its vocabulary is also limited with

118 words. The grammar of Toki Pona is isolating in nature. The vocabulary of Toki Pona includes nouns, verbs, adjectives, adverbs, pronouns, conjunctions etc. Grammatical gender is absent in Toki Pona, as is Grammatical Mood, Voice and Number. Similar to some artificial languages it is time free and hence has no tenses. It does provide deictic references to the first, second and third person. The word order used in Toki Pona is the common SVO.

**Volapük** is without a doubt one of the first efforts to design an artificial or constructed language. A rough idea of its date of emergence is accounted to be in the late 1800s. It is thought that it once had 2 million speakers. It inherits some of its vocabulary from Germanic languages and French but is schematic in nature. It has 8 vowels (including special character vowels, such as those in German for e.g.) and 19 consonants. All common word forms of nouns, verbs, adjectives, adverbs, pronouns, conjunctions, etc are present in Volapük. It too has an agglutinating grammar type. There are four cases in Volapük: the nominative, accusative, dative, and genitive. For nouns where the gender is ambiguous, prefixes are added to indicate the particular gender category. Volapük verbs are capable of indicating all three types of major modalities: subjunctive, imperative and indicative. Nouns are also inflected on the criteria of number, i.e. is a noun plural or not. Prefixes added to verbs enable the depiction of all major tenses. It does provide deictic references to the first, second and third person in both active and passive voice. It is accepted that most of the afore-mentioned markings are optional and the verb can stay untouched.

### 2.4 Morphological Overview: Discussion

Clearly various interesting trends and patterns were revealed upon analyzing artificial languages and comparing them to natural languages. It was deduced that most artificial languages have an agglutinating grammar or in some cases isolated; this fact has also been presented previously (Peterson, 2006). In addition, we also summarize the main trends based on each grammatical property individually. Some artificial languages do not give much importance to grammatical aspect and others rely on tenses to represent information about aspect. It was observed that in both artificial and natural languages, if the nouns did not inflect, then there was no grammatical case.

Artificial Languages are divided over the issue of Gender, some including it with respect to the classification of nouns. However languages such as Toki Pona and Interlingua do not indicate the gender of nouns. Very few artificial languages use mood/modality of verbs up to or more than the basic 3 levels, whereas this grammatical category is much more detailed in natural languages. Some Artificial Languages such as Loglan and Toki Pona do not inflect their nouns based on grammatical number but rely on context to get the number information across. Eastern natural languages employ the strategy of a word counter, which is basically an auxiliary word meant to convey the quantity of the noun in question. Active and passive voice are the most common in most artificial and natural languages.

Most languages (natural and artificial) provide 3 basic references to people: 1st, 2nd and 3rd person (I, you, he/she). By analyzing the tense inflecting techniques employed by various languages and in particular artificial we notice two interesting solutions. The first is to have three basic levels of tense but without introducing irregularity and ambiguity. Verbs if inflected on tense must be inflected consistently for all verbs. The second technique is to persist with the existing form of words, not to change their form but introduce the notion of time by adding auxiliary words. The most common word order across both natural languages and artificial languages is by far SVO. Some natural languages provide flexibility and hence there exists more than one option.

In summary, there are two relevant approaches of morphological design amongst artificial languages: The approach of languages such as Toki Pona and Loglan is to have very few grammatical markings, leaving it to the interpretation of the speakers, word order or the context. The second approach is to have inflections but the grammatical rules are consistent across all words within each category. Consequently, most artificial languages have either isolating or agglutinating grammar types. Esperanto for one is an inflectional language but it has less heavy inflections as compared to natural languages. It is interesting to note that natural languages gradually evolve from the second to the first approach (Beekes, 1995). With the passage of time, some grammatical markings tend to be phased out. Older languages such as Latin and Sanskrit have much more grammatical markings as compared to modern languages.

## 2.5 Phonological Overview

In linguistics, the study of the phonology of a language entails the analysis of how specific sounds are pronounced in the language (Ladefoged, 2005). Vowels and consonants together constitute the segments or phonemes of a language. Moreover the phonology of a language describes how vowels and consonants are pronounced for that language. Vowels for e.g. can differ in their point of articulation, also known as the frontness of a vowel. Or they can also be different based on the position of the jaw during pronunciation, which is occasionally referred to as the height of the vowel. Similarly, consonants can differ in the manner of articulation, the point of articulation or whether they are voiced or unvoiced.

Extending from our research goal of designing an interaction language that is easy to learn for humans, we extracted a set of the most common phonemes present in the major languages of the world. We used the UCLA Phonological Segment Inventory Database (UPSID), see (Reetz, 2008) and (Maddieson, 1984). The database provides a large inventory of all the existing phonemes of 451 different languages of the world. The number of phonemes documented in the database amount to 919. Based on number of speakers worldwide the Ethnologue (Gordon & Grimes, 2005) classifies the following 13 spoken languages as major (see Table 2.1). All the major languages in the table except English were included as part of the UPSID. This was because of a specific quota policy that is followed to select languages for the database. The quota rule states that only

one language may be included from each small family grouping (e.g. one from West Germanic and one from North Germanic), but that each family should be represented. Therefore only German was selected from the West Germanic group and English was dropped.

Language	Total Number of Phonemes
Arabic	35
Bengali	43
English	35
French	37
German	41
Hindi-Urdu	61
Japanese	20
Javanese	29
Korean	32
Mandarin	32
Russian	38
Spanish	25
Vietnamese	36

Table 2.1: Major Natural Languages of the World

However we believed that in choosing a set of phonemes that lie under an umbrella of major languages, English would play an important role. Therefore we added English to the UPSID. A list of American English vowels and consonants as stated in (Ladefoged & Maddieson, 1996) were added. In total we could enter 35 segments for English, with roughly 5 vowels unaccounted for as their transcriptions are not present in the UPSID database (Epstein, 2000). None of the consonants were absent. After incorporating English to the database we generated a list of segments that are found in 5 or more, major natural languages of the world. This resulted in a net total of 23 segments (see Table 2.6). The notations for each phoneme and their individual description are extracted from the UPSID. We added a column to connect the UPSID notations to the International Phonetic Alphabet notations (Ladefoged, 2005). We selected the same pool of 9 artificial languages for our phonetic analysis and they were now analyzed on the basis of the set of major phonemes.

## 2.6 Phonological Overview: Discussion

Interesting trends were observed; Loglan had the fewest absentees from the list of major phonemes, with only 5 (~19% of its total phonemes). Esperanto, Interlingua and Volapuk had 6 missing phonemes (~18% of the total phonemes in Esperanto). Toki Pona had the highest true misses (13), which can be attributed to the fact that its phonetic size is considerably small (71% of its phonemes were in the common list). Relatively, Klingon had the most missing common phonemes, ~35% of its total phonemes. Two dental consonants dD and sD were observed not to be found in any of the 9 artificial languages.

One reason why this might have occurred is that most artificial languages stem from Germanic or Western languages, whereas the dental consonants such as sD and dD are found in Indic or Asian languages such as Arabic, Bengali, Korean and Hindi-Urdu. In addition, the voiced dental nasal consonant nD was found in only 2 artificial languages: Loglan and Klingon, whereas tD was only found in Klingon. Trends that have been observed in natural languages with regards to the most common segments were replicated for the case of artificial languages. The phonemes m, k, j, b and p were barring a few exceptions present in all artificial languages. Klingon was the only artificial language that does not have a k and Toki Pona was the only language that did not have a b. The consonant f was absent from Toki Pona and Klingon, for the former most likely for simplicity and for the latter reasons of uniqueness. The consonants m and p were the most frequently found segments in artificial languages. They were present in all of the nine artificial languages. Certain consonants that were not found in 5 or more natural languages of the world, were found to be very common amongst the auxlangs (absent in only 1 auxlang or in none). These were the following phonemes: t, s, n and l.

The mirroring effect between natural and artificial languages extended to vowels as well. Klingon was again the odd one out, as it was the only artificial language that was adjudged not to have an i or an e. Klingon had the lowered variant of the vowel i. The vowels a, o and u were found in all the nine artificial languages.

UPSID	IPA	Description	Present in how many Natural Languages	Present in how many Artificial Languages
m	m	voiced bilabial nasal	13	9
k	k	voiceless velar plosive	13	9
i	i	high front unrounded vowel	13	9
j	j	voiced palatal approximant	12	8
p	p	voiceless bilabial plosive	12	9
u	u	high back rounded vowel	11	9
tD	t̪	voiceless dental plosive	10	1
o	o	higher mid back rounded vowel	9	9
O	ɔ	lower mid back rounded vowel	9	2
b	b	voiced bilabial plosive	9	8
f	f	voiceless labiodental fricative	9	8
w	w	voiced labial-velar approximant	9	4
a	a	low central unrounded vowel	9	9
e	e	higher mid front unrounded vowel	8	9
nD	ɳ	voiced dental nasal	8	2
g	g	voiced velar plosive	8	4
sD	s̪	voiceless dental sibilant fricative	8	0
h	h	voiceless glottal fricative	7	8
tS	tʃ	voiceless post-alveolar sibilant affricate	6	4
dD	d̪	voiced dental plosive	6	0
x	x	voiceless velar fricative	5	3
v	v	voiced labiodental fricative	5	8
r	r	voiced alveolar trill	5	6

Table 2.2: Set of Common Phonemes

### 2.7 Conclusion

We have presented a morphological overview of artificial languages where, two primary grammar types were discussed. In the future, we aim to evaluate which of the afore-mentioned grammar types will be easier to learn for our intended artificial language and which will be less ambiguous. Our phonological overview has revealed a set of phonemes that might be desirable to include in an artificial language to render it conducive for human learnability, with the assumption that the learnability of an artificial language is correlated to the extent of the overlap between the phonology of the artificial language and the phonology of the native language. It was also revealed that artificial languages created prior were based on Germanic languages, at least phonetically. Our overview is based on only nine artificial languages, whereas there are hundreds in existence. Moreover, we did not consider many languages other than A posteriori languages or international auxiliary languages therefore our overview cannot be generalized to the entire spectrum of artificial languages. Our sampling method would ultimately have an effect on the design of ROILA. The more design trends that we found via the overview are incorporated in ROILA the more it would start resembling an auxiliary language, which would not be such a bad thing.

As a motivational drive to our design process we were lucky to lay our hands on a book entitled *In the land of the Invented Languages* (Okrent, 2010). The book discusses the subject of artificial languages but not with disdain or critique but rather lauds the efforts of the creators. The book acknowledges that up to now most artificial languages that were designed were not huge successes yet they have a rationale or philosophical thought process behind their creation. The fact that the book puts the whole subject of artificial languages in such positive light was a great source of inspiration and driving force for the ROILA design process.

---

## The design of ROILA

---

Our overview of languages (both natural and artificial) resulted in several trends and design guidelines that were already discussed in the conclusion of the previous chapter. We aimed to carry out a careful integration of such trends into the design of ROILA, with the rationale that the existence of such trends would ultimately make ROILA easier to learn. This claim is of course dependent on the assumption that whatever linguistic trend is common amongst several languages is easier to learn. We also aimed to ascertain the effect these linguistic features would have on speech recognition accuracy. The design trajectory that we took followed an ascending approach. We gradually worked our way from the level of phonemes to syllables to words and lastly to the grammar.

The actual construction of the ROILA language began with a phoneme selection process followed by the composition of its vocabulary by means of a genetic algorithm which generated the best fit vocabulary. In principle, the words of this vocabulary would have the least likelihood of being confused with each other and therefore be easy to recognize for the speech recognizer. Experimental evaluations were conducted on the vocabulary to determine its recognition accuracy. The results of these experiments were used to refine the vocabulary. The subsequent phase of the design was the design of the grammar. Rational decisions based on various criteria were made regarding the selection of grammatical markings. In the end we drafted a simple grammar that did not have irregularities or exceptions in its rules and markings were represented by adding isolated words rather than inflecting existing words of a sentence. We will now explain each aspect of the ROILA design process in detail.

### 3.1 ROILA Vocabulary Design

#### 3.1.1 Choice of Phonemes

The initial set of phonemes that we started off with was the 23 phonemes that we had identified in our phonological overview as described previously (Chapter 2). At this point, we started to trim and modify the total number even further. From this list we have dropped the dental consonants: tD, nD, sD, dD because



they are hardly present in artificial languages and are only found in certain Asiatic natural languages. We also added some phonemes to this list. These phonemes were found to be very common in the set of artificial languages that we overviewed. They were the following phonemes: t, s, n, l. We also chose the more common variant of vowels such as o and a. Moreover we did not want to include any diphthongs, which are those vowels which produce two articulations within the same syllable. Examples of diphthongs in English would be for e.g. boy, where the o contributes to two differentiating vowel sounds. We wished to have only solitary variations of each vowel, thereby simplifying the pronunciation process and also allowing the speaker to not worry about where the vowel occurred in the word.

As we moved on we observed that the behavior of h is indeterminate, as in some languages it tends to behave like a vowel as well and so it could result in ambiguity for speakers (Ladefoged, 2005). It is also known that v is confused with b for speakers of certain eastern languages and g has been acknowledged as difficult to articulate (Ladefoged & Maddieson, 1996). Therefore after excluding certain phonemes the final set of 16 phonemes that we wished to use for ROILA was: a, b, e, f, i, j, k, l, m, n, o, p, s, t, u, w or in the ARPABET notation (Jurafsky, Martin, Kehler, Vander Linden, & Ward, 2000) AE, B, EH, F, IH, JH, K, L, M, N, AA, P, S, T, AH, W a total of 5 vowels and 11 consonants. In summary, our choice of phonemes was based on our linguistic overview of languages, general articulation patterns and acoustic confusability within phonemes (especially the vowels). Another important consideration was that having too few phonemes could effect the diversity of the vocabulary. Note that at times, some aspects preceded others, for e.g. we decided to include both m and n, even though they are acoustically similar, mainly because they are found in artificial languages. We could have completely inherited the common phoneme list that we discovered. But that could have meant that the resulting alphabet would contain phonemes from different types of languages that not many people could pronounce in completeness. This was because the common phoneme list consisted of phonemes present in 5 or more natural languages. If we had tried to find a common phoneme list for all the natural languages that we considered the phoneme set would have been very small. What would be wiser would be to pick and choose from the common phoneme list that we extracted and add phonemes as we see fit. We also decided not to include any kind of phonetic stress in ROILA.

Below is the table of all letters used in ROILA (see Table 3.1). Also provided are the International Phonetic Alphabet (IPA) (Ladefoged, 2005) and ARPABET pronunciations. Since these vowels and consonants are also found in English, we include their pronunciations with examples from English.

### **3.1.2 Word Length**

Once we had identified our phoneme set the next step was to generate the vocabulary. Within creating the vocabulary the first design decision taken was the word length. For the initial design, we set the required word length as S syllables, where  $2 \leq S \leq 3$  and the number of characters as  $4 \leq C \leq 6$ . These

Letter	IPA	ARPABET	Example
a	ae	AE	fast
e	ε	EH	red
i	ɪ	IH	big
o	ɑ	AA	cot
u	ʌ	AH	but
b	b	B	buy
f	f	F	for
j	dʒ	JH	just
k	k	K	key
l	l	L	late
m	m	M	man
n	n	N	no
p	p	P	pay
s	s	S	say
t	t	T	take
w	w	W	way

Table 3.1: Initial ROILA phonetic table

constraints imposed on word length at first glance offers a reasonable balance between improving speech recognition and being easy to learn for humans (Hinde & Belrose, 2001). We could not have only longer words as they would be hard to remember and pronounce, similarly only having shorter words would be difficult for the recognizer. Moreover we also analyzed the vocabulary space such length of words would cover, as is shown in the table (see Table 3.2). The shorter the word length the less number of combinations would be possible and it would be more difficult to find the right acoustic uniqueness and balance. At the same time we did not want to deal with very long words which would affect the learnability of ROILA. We have chosen only a specific type of word structure in the shown table (see Table 3.2), i.e. words having only Consonant-Vowel (CV) units. Ultimately our ROILA vocabulary would have words comprising of CV units only, the reasons for doing so will become clearer later on this chapter.

Word Length	Word Structure	Total Words Possible
2	CV	$55 = 11 \times 5$
3	CVC	605
4	CVCV	3025
5	CVCVC	33275
6	CVCVCV	166375
7	CVCVCVC	1830125
8	CVCVCVCV	9150625

Table 3.2: Vocabulary size spread based on word length

### 3.1.3 Word Structure: initial design

Here we took inspiration from another artificial language: Toki Pona (Kisa, 2008). Toki Pona is built with the aim to promote simplicity and consequently is supposed to be easy to learn for humans. It has a vocabulary of 118 words and sufficiently caters for the needs of a simple language.

Upon analyzing the word structures of Toki Pona, we noticed the following structure, for all words where  $S < 3$ ,  $C < 5$  and where V represents a Vowel and C a consonant: V, VC, VCV, VCCV, VCVC, CV, CVC, and CVCV. It is fairly evident that vowels are not followed in succession to each other. Additionally, by adding consonants in between vowels strong distinctions are made between syllables. Using the word types of Toki Pona we designed our own word types. We started off with 8 word types and attempted to maintain a balance of learnability and appropriate word length. In the first iteration we selected the following word types: VCCVCV, VCVCV, VCVCCV, CVCVC, CVCVCV, VCCV, VCVC, and CVCV.

### 3.1.4 Using a Genetic Algorithm to generate ROILA words

The manner in which the words would be constructed would need to be carefully implemented as to render the vocabulary to be speech recognition friendly. Moreover, the method would need to be scalable as well to allow the generation of as many words as required at any time.

At this point we took inspiration from another similar approach as shown by (Hinde & Belrose, 2001). They too had utilized a genetic algorithm for generating a vocabulary; however we chose to make slight modifications to their method based on our requirements. These modifications are described later on in this section. We carried out the implementation of our genetic algorithm in Java.

In order to define the exact representation of the ROILA words we designed a genetic algorithm that would explore a population (sets) of words and converge to a solution, i.e. a group or dictionary of words that would have the lowest acoustic confusion amongst them and in theory be ideal for speech recognition. The algorithm was randomly initialized for a population of  $N$  dictionaries/plausible solutions each having  $W$  words or genes, where each word was any one of the afore-mentioned 8 ROILA word types, where each vowel could be any of the 5 possible and each consonant from the 11 possibilities. Moreover it was also ensured that during the initialization process, all of the  $N \times W$  genes were unique.

The algorithm was then run for  $G$  generations with mutation and cross over being the two primary offspring generating techniques. For a given dictionary its confusion was defined as the average confusion of its all constituent words or genes, i.e. pair wise confusions were computed for each word. In every generation, 6% of the best fit (elite) parents were retained and infants were reproduced to complete the population, so that for every population the number

of words would remain consistent, i.e. W per N would never change. Parents were selected for breeding using the standard roulette wheel selection (Houck, Joines, & Kay, 1995). Note that in absolute terms low fitness or low confusion was preferred, so the selection had to be reversed.

We now briefly describe our offspring generating techniques. As described earlier, the best 6% of the population would compete to take the role of the parents. A mutation was implemented by swapping one randomly chosen vowel with another unique vowel and swapping one randomly chosen consonant with another consonant. Unique in our reference means that the vowel or consonant was not present in the word before. It was also made sure that after mutating the word in question it would not take the shape of another pre-existing word in that particular vocabulary. This process was repeated for every word existing in the vocabularies of the two parents. Therefore every mutation would result in two new infants and the entire cycle was repeated a sufficient number of times till the least-fit population was replaced. Mutation was set to a standardized rate of 1% and on all other instances cross over would take place. Crossover would generate infants by simply swapping the vocabularies of the two selected parents after a randomly predetermined cutoff point.

The fitness function was determined from data available in the form of a confusion matrix (Lovitt, Pinto, & Hermansky, 2007), where the matrix provided the conditional probability of recognizing a phoneme  $p_i$  when phoneme  $p_j$  was said instead. The confusion matrix was generated via a phoneme recognizer using the TIMIT corpus for English words. We computed the confusion between any two words within a dictionary by computing the probabilistic edit distance, as suggested in (Amir, Efrat, & Srinivasan, 2001). The probabilistic edit distance was a statistical extension of the conventional Levenshtein distance algorithm (Gilleland, 2002). Insertion and deletion probabilities of each and every phoneme were also utilized from (Lovitt et al., 2007). In summary, the similarity of two strings was given by the edit distance, which was a sequence of possible transformations converting one string to another. The transformations could consist of insertions, deletions or substitutions. The probabilistic edit distance would hence be the joint likelihood of the transformations, which were assumed to be independent.

Our genetic algorithm was an adaptation of the technique presented in (Hinde & Belrose, 2001). Their method was more discrete in nature and it was based on the assumption that the confusion between two words in a vocabulary is simply the sum or product of all the inter-phoneme confusion probabilities. However, we relied on a stronger probabilistic measure which was based on the Levenshtein distance algorithm as we were fortunate to not only have substitution probabilities, as in the case of (Hinde & Belrose, 2001), but also insertion and deletion probabilities for every phoneme in question. Therefore we could adopt a Levenshtein distance based formula to compute a probabilistic edit distance where the sums were replaced by products.

### 3.1.5 Genetic Algorithm simulations

Several runs were executed of the algorithm to determine when the algorithm would converge to a solution and therefore what would be the appropriate settings for  $G$  and  $N$ . The algorithm was seen to converge for  $G > 150$  (see Figure 3.1). The confusion of the best vocabulary seemed to stay consistent after 150 iterations. We noted two main variables that played an important role in influencing the confusion of a vocabulary: firstly word length and secondly size of the vocabulary. The algorithm preferred longer words as was exhibited in the relation between word length and average confusion of every generation (see Figure 3.2, where the average confusion of a population is plotted against the average word length of all its constituent words). For an increase in average word length of a population a corresponding decrease in average confusion was found. A sample run resulted in the best fit population having 42% of words with length of 6 characters, 29% of words with length of 5 characters and 29% of words with in total 4 characters, for  $W = 100$ ,  $G = 200$  and  $N = 200$ , exemplifying the tendency of the algorithm to favor longer words. Another interesting trend that was observed was that an increase in total number of words ( $W$ ) led to a corresponding increase in the average confusion of all the populations ( $N$ ).

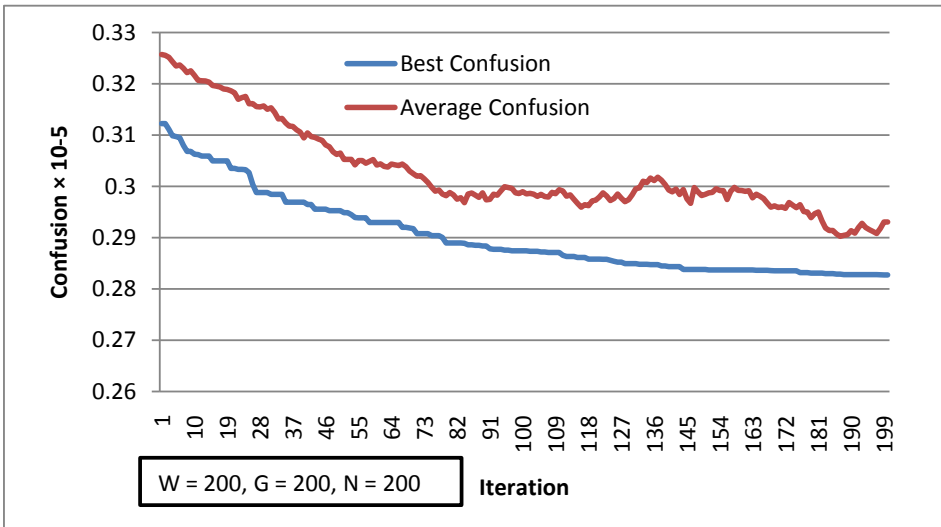


Figure 3.1: Simulation graph for  $N=G=W=200$

## 3.2 Using a word spotting experiment to evaluate the ROILA vocabulary

In order to adjudicate whether ROILA was indeed better than a sample counterpart English vocabulary we ran a word spotting test, where participants were

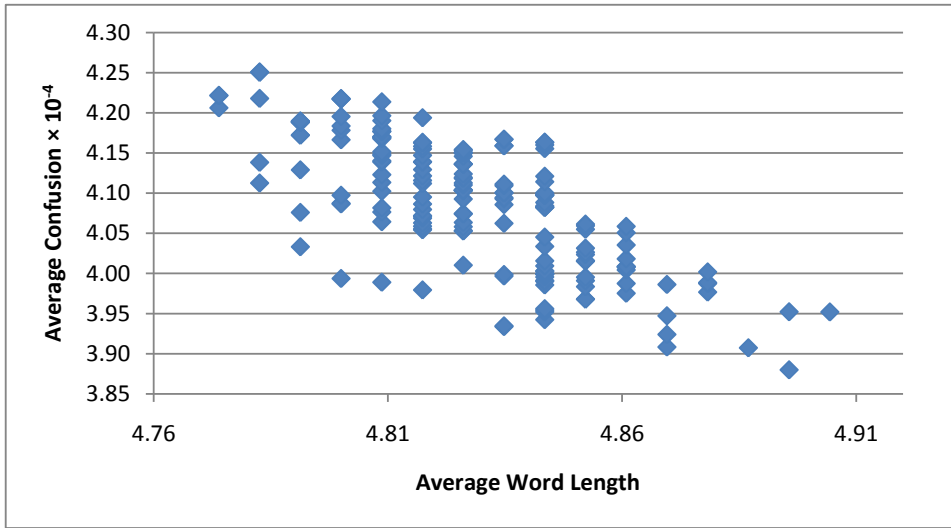


Figure 3.2: Graph relating average confusion and average word length where  $N=G=W=300$

asked to record samples of every word from both English and ROILA and the recordings were then passed offline through the Sphinx-4 (Lamere et al., 2003) speech recognizer.

We chose the size of the Toki Pona vocabulary as a starting figure for our first ROILA vocabulary, as the vocabulary size would be large enough to evaluate recognition accuracy and small enough to not cause practical problems in terms of conducting recording experiments with participants. Moreover it allowed us to easily determine the English vocabulary to compare against the ROILA vocabulary. In order to have a benchmark of English words to compare against in the subsequent word spotting test we set the English vocabulary as the meanings of all the 115 Toki Pona words. The average length of the English vocabulary thus obtained was 4.5 characters per word. Three Toki Pona words were not considered as they are used in the language as grammatical markers. The genetic algorithm was run to determine a vocabulary of  $W = 115$  ROILA words, where each word would be any one of the 8 mentioned ROILA word types. The first ROILA vocabulary was generated by running the algorithm for  $N = G = 200$ . Example words from this vocabulary are shown in Table 3.3. The ROILA vocabulary had 54 six character words, 31 five character words and 30 four character words, with the average length of the ROILA vocabulary was 5.2 characters per word. Any recognition improvement attained by ROILA would have to be considered carefully in light of this bias.

Word Type	Examples
CVCV (CV-type)	babo, wimo
VCVC	ujuk, amip
VCCV	obbe, uwjo
CVCVC (CV-type)	mejem, kutak
VCVCV	ofeko, ejana
CVCVCV (CV-type)	panowa, fukusa
VCCVCV	ukboma, emfale
VCVCCV	onabbe, emenwi

Table 3.3: Vocabulary size spread based on word length

### 3.2.1 Participants

16 (6 female) voluntary users were recruited for the recordings. Participants had various native languages but all were university graduate or post graduate students and hence had reasonable command over English. The total set of Native Languages of the participants was 10 (American English-3, British English-1, Dutch-5, Spanish-1, Urdu-1, Greek-1, Persian-1, Turkish-1, Bengali-1 and Indonesian-1). It is also worthy to point out that none of the participants spoke ROILA before and neither had they been exposed to it prior.

### 3.2.2 Material

Recordings were carried out in a silent lab with little or no ambient sound using a high quality microphone (see Figure 3.3). A recording application was designed that would one by one display the words to be recorded. Participants would record all the words from a particular language before moving on to the next language. Recordings of every participant were then passed through the Sphinx-4 Speech recognizer. The choice of speech recognizer was carefully ascertained keeping in mind the requirement that the speech recognition engine should be open source and allow for the recognition of an artificial language. Sphinx was tuned such that it was able to recognize ROILA by means of a phonetic dictionary; however the acoustic model that we used was that of American English. In the ideal circumstances, we would have liked to train a completely new acoustic model for ROILA by dictating a large corpus of ROILA sentences to it. This was obviously not possible due to practical reasons. Training a new acoustic model requires extensive amounts of corpus data.

Note that the ROILA words were generated from a confusion matrix that extracted its data from the basis of another speech recognizer (Lovitt et al., 2007) and not Sphinx; this might be a limitation but most speech recognizers operate on the same basic principles. In addition, we did not carry out any training on the acoustic model for ROILA. For further details about our choice of speech recognizer and how we implemented ROILA within it, please refer to (Chapter 4). In chapter 4 we mainly discuss live speech recognition of ROILA. However, in the word spotting experiment we utilized offline speech recognition of ROILA as the recordings were passed into the speech recognizer after the

recordings were complete. It should be noted that Sphinx-4 requires a specific format and type of audio recordings, i.e. type wav, 16 bit and mono.



Figure 3.3: Recording setup

#### 3.2.3 Pilots and Procedure

In order to ascertain the recognition of ROILA within Sphinx-4, we carried out some pilot recording sessions. We noticed that for certain American English speakers the recognition accuracy was relatively higher, an expected result due to the use of American English acoustic model. Therefore we chose an American English speaker and conducted several recording iterations until we had a pool of sample recordings of that voice that rendered a recognition accuracy of 100%. These sample recordings of every word would be played out once before other participants recorded their own pronunciations of each ROILA word. The participants had a choice of listening to the sample recording again. This was done to ensure that the native language of participants would not affect their ROILA articulations. We instructed participants to follow the sample recordings as much as possible. No sample voice was played out in English as all participants had sufficient knowledge of the English vocabulary employed in the experiment.

#### 3.2.4 Experiment Design and Measurements

The experiment was carried out as a mixed design with one within subject factor: language type (English, ROILA) and three between subject factors (gender, recording order and whether american english was the native language of the participants). The dependent variable was the number of errors in recognition by Sphinx. Words from both English and ROILA were randomly presented to



counter any training or tiring effects. The order of recording English or ROILA first was also controlled between participants.

### 3.2.5 Results from the word spotting experiment

We executed a single repeated measures ANOVA with recording order (ROILA or English first), gender and whether participants spoke American English (as a native language) as the three between subject factors and language type (ROILA or English) as the within factor. The ANOVA revealed that language type did not have a main effect  $F(1, 9) = 0.758, p = 0.41$  and there were no interaction effects either. Both ROILA and English performed equally in terms of accuracy of the number of words correctly recognized (see Table 3.4). Without any training, such recognition accuracy is expected from Sphinx on test data (Samudravijaya & Barot, 2003).

Language	Mean Accuracy (%)	Std. Dev
English	67.66	13.38
ROILA	67.61	10.89

Table 3.4: Means table for recognition accuracy for English and ROILA

From the between subject factors, the factor of whether participants spoke American English had a significant effect  $F(1, 9) = 6.25, p = 0.034$  as they achieved higher recognition accuracy for both ROILA and English (see Table 3.5).

Language	Native American English Speakers		Non-Native American English Speakers	
	Mean Accuracy (%)	Std. Dev	Mean Accuracy (%)	Std. Dev
English	82.90	7.29	64.15	11.98
ROILA	79.71	5.66	64.82	9.89

Table 3.5: Means Table for recognition accuracy for English and ROILA across native language

Recording order was not significant  $F(1, 9) = 0.019, p = 0.89$  and neither was Gender  $F(1, 9) = 1.07, p = 0.33$  as can be seen in Table 3.6 and Table 3.7 respectively. Consequently this meant that both recording order and gender were not influencing the recognition accuracy measurements.

Language	ROILA First		English First	
	Mean Accuracy (%)	Std. Dev	Mean Accuracy (%)	Std. Dev
English	66.85	14.55	68.48	13.05
ROILA	70.33	12.74	64.89	8.65

Table 3.6: Means Table for recognition accuracy for English and ROILA across experiment order

We carried out a second ANOVA with word length of ROILA words as the independent variable. It had 3 levels (4, 5 or 6 characters). The dependent

Language	Male		Female	
	Mean Accuracy (%)	Std. Dev	Mean Accuracy (%)	Std. Dev
English	63.91	13.08	73.91	12.41
ROILA	66.61	12.43	69.28	8.50

Table 3.7: Means Table for recognition accuracy for English and ROILA across gender

variable was the recognition accuracy within each category of every participant, as each category had a different total number of words. The ANOVA analysis revealed that Word Length had a significant effect on the number of recognition errors  $F(2, 18) = 20.97, p < 0.0001$ . Pair-wise comparisons (Bonferroni) between all three categories were significant ( $p < 0.05$ ). The average accuracy for 4, 5 and 6 character words was 52.6%, 69.33% and 77.7% respectively. Therefore longer words performed better in recognition, as is evident in the graph (see Figure 3.4).

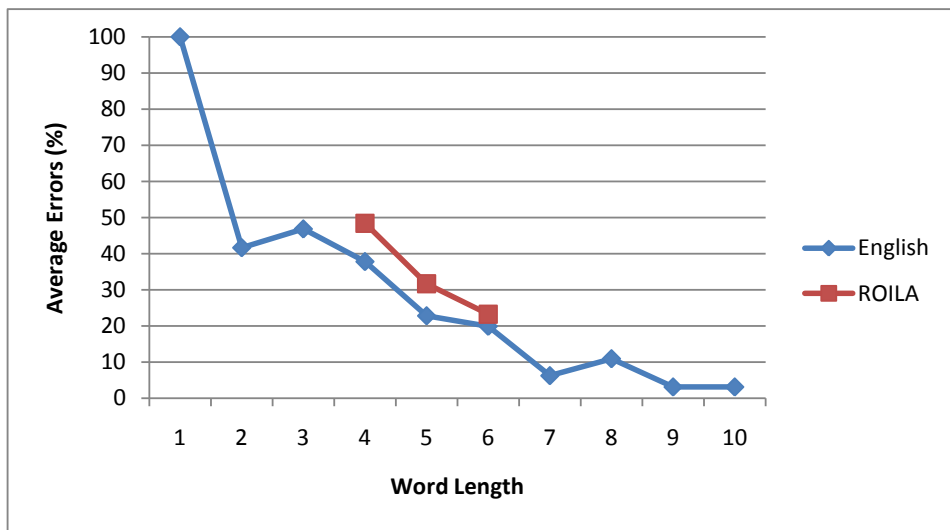


Figure 3.4: Graph illustrating the relation between word length and recognition accuracy

In order to understand if the word structure of ROILA words had an effect on recognition accuracy, we executed an analysis in which the type of word was the independent variable. This factor had 2 levels (CV or non-CV type, the former having three word types and the latter five-see Table 3.3. Our vocabulary had in total 73 non-CV type words and 42 CV type words. The dependent variable was the number of participants who got that type of word wrong. The

ANOVA analysis revealed a nearly significant trend  $F(1, 113) = 3.6, p = 0.06$ , where CV-type words performed better on recognition (see Table 3.8).

Word Type	Mean	Std. Dev
Non CV	5.75	4.28
CV	4.19	4.21

Table 3.8: Means Table for number of users who got a type of word wrong

### 3.2.6 Second iteration of evaluating the ROILA vocabulary

For our second iteration of the evaluation we generated a new vocabulary that comprised of CV type words only and in the second word spotting experiment our hypothesis was that the new ROILA vocabulary would be the English vocabulary in terms of recognition accuracy. In this iteration the vocabulary was set to include only the three CV word types (CVCV, CVCVC, CVCVCV). The genetic algorithm was run again with the same parameters  $G = N = 200$  and  $W = 115$ . The new vocabulary had an average word length of 5.1 characters (see Table 3.9 for sample words). We asked 11 (4 female) from the earlier 16 participants to carry out recordings of the new vocabulary using the same setup and procedure. We did not have them record the English words again. The same American English speaker as in the first setup was used as the sample voice, where the sample recordings of the new ROILA vocabulary had 100% recognition accuracy in Sphinx. The recordings from the 11 participants were run in Sphinx to evaluate the recognition accuracy of the new ROILA vocabulary. A repeated measures ANOVA with language type as the single within subject factor revealed that the new ROILA vocabulary significantly outperformed English  $F(1, 10) = 4.86, p = 0.05$  (see Table 3.10 and Figure 3.5). This vocabulary was hence declared as the first ROILA vocabulary. In summary, the first ROILA vocabulary was declared to have three word types: CVCV, CVCVC, and CVCVCV. They were referred as CV type words, where C could be any of the 11 consonants and V any of the 5 vowels. Note that the bar chart represented in Figure 3.5 is computed by taking into account the total number of recognition errors (from a pool of 115 words). Table 3.10 summarizes the recognition accuracy as a percentage. It is also pertinent to observe that since we had only one within subject factor (language type) we could have also executed a normal paired samples t-test to evaluate the differences between the new ROILA vocabulary and the original English vocabulary. The results would have been similar to the one way repeated measures ANOVA that we executed anyway.

Word Type	Examples
CVCV	bama, pito
CVCVC	fenob, topik
CVCVCV	simoti, banafu

Table 3.9: Sample words from second ROILA vocabulary

Language	Mean Accuracy (%)	Std. Dev
English	69.66	13.92
ROILA (1)	70.63	9.96
ROILA (2)	74.99	10.56

Table 3.10: Means Table for recognition accuracy for English and two versions of ROILA

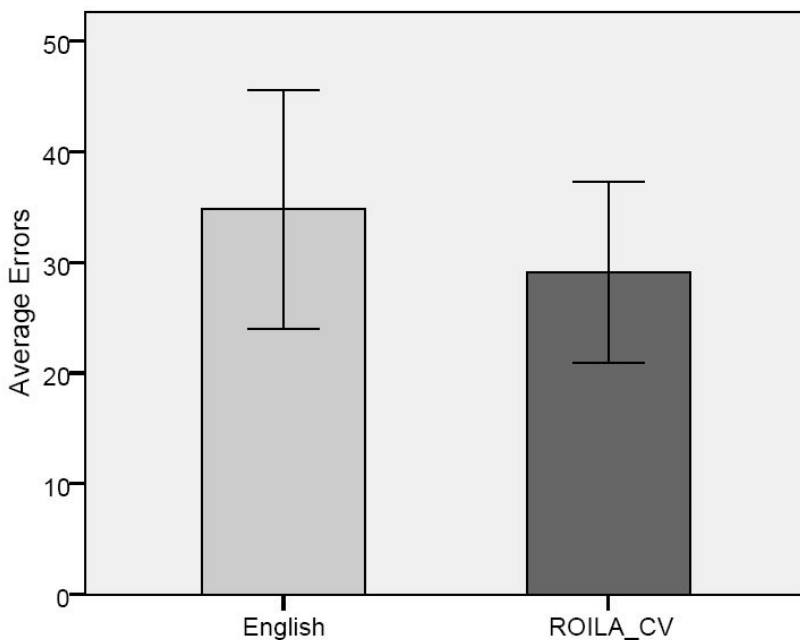


Figure 3.5: Bar chart showing mean errors difference between English and second ROILA vocabulary

### 3.2.7 Discussing the results from the word spotting experiment

Our results of the word spotting experiment revealed some interesting insights besides guiding the design of the first ROILA vocabulary. Firstly, we were able to achieve improved speech recognition accuracy as compared to English for a relatively larger vocabulary. Similar endeavors have only been carried out for a vocabulary size of 10 (Arsoy & Arslan, 2004). Secondly, we confirmed the result that longer words perform better in speech recognition (Hämäläinen et al., 2005). Thirdly, we quantitatively illustrated that CV type words perform better in recognition. This has only been discussed, for e.g. by (Hinde & Belrose,

2001), but was not empirically demonstrated. Co-articulation of CV syllables and hence easier and clearer pronunciation could be one explanation for their better recognition accuracy. Lastly, we showed that American English speakers significantly outperformed other speakers in our setup, due to our choice of the acoustic model in Sphinx, which was trained using American English speakers. To conclude we can state that recognition accuracy of the second ROILA vocabulary was significantly better compared to English despite using an acoustic model designed for English.

### **3.3 A larger ROILA vocabulary and giving it semantics**

On first sight, the Toki Pona vocabulary would seem to be sufficient for the purposes of using ROILA to interact with robots on a small scale. However, we did not want to limit ourselves and hence chose to work on further expanding the vocabulary. We could of course run the genetic algorithm to generate any size of vocabulary but the interesting challenge was to assign relevant and useful meanings to those words. We wished to cover a universally applicable vocabulary and not a vocabulary that is restricted to a specific domain. Therefore, the meanings of the ROILA words could be taken from Basic English (Ogden, 1944), (David, 1997) a concise version of the English language comprising of only 800 words (besides 3 word markers of our own, which we will explain in the grammar section). We ran the algorithm to generate 688 words without losing the 115 Toki Pona words, i.e. the 115 words were added in the computation executed by the genetic algorithm. The words of Basic English were first sorted based on frequency of occurrence using metrics provided by the Department of Psychology, Ghent (Brysbaert & New, 2009). The ROILA words were sorted by their length. The meanings of the more frequent words in Basic English were then associated to the shorter words in ROILA.

We did not carry out any further additional word semantic assignment. One way of doing this could have been to give participants a group of ROILA words and a group of possible semantics and let them carry out the assignment themselves. For example, participants might find implicit relations between ROILA words and English words. However semantics was not really in the heart of the ROILA project, therefore we kept the assignment as it was, i.e. initially logical and then semi-random. In summary, we did not care that amongst a certain category of ROILA words (for example 2 syllable words) which word got which meaning in English. In conclusion, we had two versions of the ROILA vocabulary, a short version consisting of 115 words and a longer version consisting of 803 words (also contained the 115 words).

Upon exposure of ROILA to the public via our website, which also resulted in considerable media exposure, we received several queries/comments on some important words that were missing in the initial ROILA vocabulary due to their absence in Toki Pona or Basic English. For e.g. there was no word for robot. We had apple, but no eat. We had shirt, but no pants, and no wear. Fortunately, ROILA words are generated automatically by a genetic algorithm, therefore to extend the vocabulary is only a matter of rerunning the algorithm with slightly

modified parameters so that the algorithm takes into account the existing vocabulary and does not discard it completely. Consequently we see the following approach can solve the problem of certain missing words.

We ran the genetic algorithm so that it generated a new vocabulary comprising of 100 *wild card* words besides the original 803 ROILA words, i.e. a vocabulary of 903 words. ROILA speakers could then assign any meaning not existing in Basic English or Toki Pona to any word from those 100 wild card words. The genetic algorithm would be responsible of ensuring that the new vocabulary has words which are acoustically different from each other. An English to ROILA dictionary (803 plus some wild card words) can be found in the Appendix of the thesis.

### 3.4 Phonetic Modification of the vowels

Upon the completion of our first round of vocabulary and grammar evaluations (described later on in this chapter), we observed that the pronunciation of certain ROILA vowels was confusing for some of the participants, especially those who were not native English. The most problematic vowel was the ROILA character o or in ARPABET terminology the phoneme AA. During the design process we had not anticipated that speakers would find it difficult to realize the subtle articulation differences. A more distinct AO seemed to be a more viable option as there was evidence based on the probabilities of the phoneme confusion matrix (Lovitt et al., 2007) that AO would be acoustically confused much less with AE (our choice of the character a). However at this juncture, it would have been difficult to discard our entire vocabulary which would have been the case had we simply interchanged the two vowels in question, i.e. swapping the AA with AO. This was mainly because by the time we had realized the acoustic similarity we had already created initial ROILA training material and we also had a group of people who had already started learning the ROILA vocabulary. As a workaround, we first tried to establish empirically if by swapping the vowels did the acoustic confusion of a vocabulary change significantly. We conducted a small experiment to test this.

Two simulations of our genetic algorithm were run. The first simulation used our original setup (i.e. vowel AA) and the second simulation used AO. Both simulations were each run 100 times, which therefore formed our sample size. For each simulation the size of each vocabulary was 50 words and each simulation iterated for 100 times over 100 solutions/populations, i.e.  $W = 50, N = G = 100$ . We chose the figure of 50 words because in our final evaluation the same figure would be the size of our evaluation set (see Chapter 6). We ran the simulations 100 times to ensure that the randomness of the algorithm was kept in check.

Therefore at the end of the two simulations we had  $2(AA, AO) \times 100$  sample points (confusion of best vocabulary). To empirically evaluate if there was a difference amongst these sample points we conducted a simple t-test. The t-test revealed that there was no significant difference in the confusion of the best vocabulary ( $t(198) = 1.32, p = 0.19$ ). Our empirical results gave us enough

confidence to alter the pronunciation of the vowel from AA to AO without breaking down our vocabulary and at the same time providing ease of pronunciation to prospective ROILA students. A sample ROILA pronunciation table is given (see Table 3.11). In the table the o is represented by the ARPABET symbol AO and not AA, due to reasons described earlier in this section. This was the only change made to the phonemes of ROILA as shown in an earlier table in this chapter (see Table 3.1).

ROILA	Meaning	Literal pronunciation	ARPABET
webufo	left (i.e. I turn left)	w eh b uh f oh	W EH B AH F AO
kanek	go	k aeh n eh k	K AE N EH K
botama	turn	b oh t aeh m aeh	B AO T AE M AE
koloke	forward	k oh l oh k eh	K AO L AO K EH

Table 3.11: Examples of ROILA words and their pronunciations

**3.5 ROILA Grammar Design**

In conjunction with conducting a phonological overview of artificial languages we also carried out a morphological overview of artificial languages individually and also in contrast to major natural languages of the world as described in the previous chapter (Chapter 2). This aided us in identifying grammar features which were popular in both natural and artificial languages. As described previously, we determined several grammatical categories based on properties defined in various linguistic encyclopedias. Gender, numbering, tense, mood and aspect are some examples. However within each category there were a number of options that we could choose from, for e.g. should we have gender? If yes, how many levels? How many tenses should we have? In order to make our choice we carried out a rationale decision making process by utilizing the Questions, Options and Criteria (QOC) technique (MacLean, Young, Bellotti, & Moran, 1991). For this purpose we defined the following important criteria for every grammatical property and they are described below. Appropriate weights were assigned to the criteria based on their importance, according to our subjective determination. The total sum of all the weights was 1. The weights are given in brackets in the definitions below.

Learnability (0.15) defines whether the grammatical marking in question would be easy to learn or not. This criterion was based on the assumption that the less number of rules a speaker would have to remember the easier it would be to learn.

Expected recognition accuracy (0.3): defines the effect the grammatical marking would have on the anticipated word error rate given that the more constrained a grammar (lower perplexity) is the better it would be for recognition (Makhoul & Schwartz, 1995).

Vocabulary size (0.1): describes the effect the grammatical marking would have on increasing or decreasing the vocabulary size.

Expressive Ability of the language (0.1): defines whether using the grammatical marking in question would actually enable speakers to express more concepts than they would have been able to do so otherwise.

Efficiency (0.1): simply relates the grammatical marking to how many words would be required to communicate any solitary meaning.

Acknowledgement within Natural (0.05) and Artificial Languages (0.1): states the popularity of the particular grammatical marking amongst each type of languages.

Relevance to the HRI context (0.1): would state the tendency of the marking being used in a conventional Human Robot Interaction context, which is after all the goal of ROILA.

Learnability and expected recognition accuracy were assigned higher weights with recognition accuracy being given twice as much weight as learnability. As stated on several occasions before, learnability and recognition accuracy contradicted each other. All relevant possibilities of each grammatical category were listed and each possibility was then ranked across the criteria by giving a number between 1 and 3 with 3 being the best fit. The category which yielded the highest output, i.e.  $(\sum rank \times weight)$  was then chosen to be as the grammar category of choice. As an example, a subsection of the table is given for the grammatical marking numbering (see Table 3.12).

After filling in a matrix we concluded firstly that the ROILA grammar would be of isolating type. Affixes would not be added as this might alter the word structure hereby reducing their efficiency for speech recognition. Therefore grammatical categories in ROILA would be represented by word markers, i.e. inflections will be represented by adding words from within the vocabulary to exhibit grammatical meanings. Moreover, the grammar of ROILA was intended to be regular. Hence there would be no exceptions and any grammatical rule would be applicable to all word types, for e.g. verbs or nouns.

At the end we arrived at the following basic grammatical properties: Gender (male, female) on the level of pronouns only, Numbering (singular, plural) on the level of nouns, Person references (first, second, third) on the level of pronouns, Tense (past, present, future) and word order would be SVO. In the subsequent section we describe the grammar rules in detail and give some examples.

### **3.6 The rules of the ROILA grammar**

In this section we describe every rule of the ROILA grammar and how it is represented in ROILA. To exemplify our point we give example sentences. Once again we would like to remind the reader that all our grammar formulations



Grammatical Numbering	Learnability	Recognition accuracy	Ac- Vocabulary size	Expressive ability	Efficiency	Relevance HRI	Conformity to within NL's	Conformity within AL's	Total Score
Optional	2	2	2	2	2	1	1	1	1.75
Not present	3	1	3	1	2	2	1	2	1.8
Singular, Plural	2	3	2	3	2	3	3	3	2.65
Singular, two in number, More than two	1	3	1	3	2	1	1	1	1.9

Table 3.12: Subsection of the GoC Matrix

were based on the results from the QoC matrix. For details on the definitions of the relevant grammatical categories please refer to (Chapter 2).

**3.6.1 Parts of Speech**

ROILA has the following parts of speech: Nouns, verbs, adverbs, adjectives, a couple of prepositions, conjunctions and 4 pronouns (I, you, he and she).

**3.6.2 Names and Capitalization of Nouns**

Wherever applicable, names of people will be used as they are. The first letter of names borrowed from natural languages will be capitalized as will be the first letter of every new sentence. Alternatively, pronouns can also be used.

**3.6.3 Gender**

Gender will not be marked in ROILA neither for animate beings nor for inanimate objects. Gender will only be marked for pronouns (he/she), as we will describe in the person references. Note that expressing gender in personal references was not our main priority, because after all we would expect most interaction between a single user and a single robot to not require gender markings for personal references. This was also the reason why we didn't mark gender in the first and second person references.

**3.6.4 Word Order**

Based on the results from the QoC formulation, it was decided that ROILA word order would be of the SVO type. An example is show in the table (see Table 3.13). The SVO structure is clear to see, where Pito is the subject, jasupa is the verb and jinolu is the object of the sentence.

English Sentence	ROILA Translation	Literal Translation
I will drop a ball	Pito jasupa jifo jinolu	I drop <word marker: future tense> ball

Table 3.13: Sample ROILA sentence showing SVO Word Order

**3.6.5 Numbering**

Grammatical Numbering will be represented for plural nouns. In the singular instance there will be no changes to the noun whereas for plural nouns the word tuji will be added after the noun (see Table 3.14). The meaning of tuji is very or many.

**3.6.6 Person References**

Person References are conveyed using I and you (pito and bama respectively). We assume that this would be enough as ROILA is primarily intended for Human Robot Interaction and references to he/she can be replaced by the usage

English Sentence	ROILA Translation	Literal Translation
I love fruit	Pito loki wikute	I love fruit
I love fruits	Pito loki wikute tuji	I love fruit <word marker: plural>

Table 3.14: Sample ROILA sentences related to Grammatical Numbering

of names of those people. However we still allow for the representation of gender in the third level of person references. This is accomplished by using liba for he and mona for she. We give an example of the use of Pito (I) and liba (he) in sentences (see Table 3.15 and Table 3.16 respectively).

English Sentence	ROILA Translation	Literal Translation
I can go left or right	Pito leto kanek webufo buno besati	I can go left or right

Table 3.15: Sample ROILA sentence showing the use of pito (I)

English Sentence	ROILA Translation	Literal Translation
He saw the bird	Liba make jifi mipuki	He see <word marker: past tense> bird

Table 3.16: Sample ROILA sentence showing the use of liba (he)

#### 3.6.7 Tenses

Tense are spread over the basic three levels: past, present and future. Present tense will imply normal sentences with no changes to the verb. Past will be represented by the addition of the word jifi and future by jifo after the verb in question, as shown in the table (see Table 3.17). At first sight it would seem jifi and jifo sound acoustically similar and one could question the working of our genetic algorithm. However, only jifi is the product of our algorithm and jifo is a self made similar looking variant to ensure ease of learnability for ROILA speakers.

English Sentence	ROILA Translation	Literal Translation
I am walking to the house	Pito fosit bubas	I walk house
I walked to the house	Pito fosit jifi bubas	I walk <word marker: past tense> house.
I will walk to the house	Pito fosit jifo bubas	I walk <word marker: future tense> house.

Table 3.17: Sample ROILA sentence showing ROILA tenses

### 3.6.8 Polarity

Polarity within sentences will be represented by yes/ok/good (wopa) in the case of affirmative or positive indications and no (buse) in the case of negative assertions. Wopa can also be used in response to questions or declaration of agreement or acknowledgement. Examples are shown in Table 3.18.

English Sentence	ROILA Translation	Literal Translation
Do not listen to her	Buse lulaw mona	No listen her
Listen to her	Lulaw mona	Listen her
Don't walk	Buse fosit	No walk
You are a good person	Bama wopa tiwil	You good person

Table 3.18: Sample ROILA sentence showing the representation of polarity in ROILA

### 3.6.9 Referring Questions

Questions in ROILA can be addressed using the word biwu which literally translates to what, see Table 3.19.

English Sentence	ROILA Translation	Literal Translation
What color is the museum?	Biwu wekepo buse kulil bubas?	What color not new house
Do you have boxes?	Biwu bama saki lujusi tuji?	What you have box <word marker: plural>

Table 3.19: Sample ROILA sentence showing the use of biwu in ROILA

### 3.6.10 Conjunctions

The three basic conjunctions that are part of ROILA include: sowu (and), buno (or), and kijo (because). An example is given in Table 3.20.

English Sentence	ROILA Translation	Literal Translation
Museums are dirty and bad	buse kulil bubas tuji topik sowu bujeti	not new house <word marker: plural> dirty and bad

Table 3.20: Sample ROILA sentence showing the use of sowu in ROILA

### 3.6.11 Punctuation

Every sentence will conclude with a full stop . with the first letter of every new sentences also capitalized. Question marks can be used in sentences where a question is asked and ROILA does not allow commas, apostrophes

and quotation marks: , ‘ ’ respectively. This is mainly done because we do not adopt any stress in ROILA articulations which would have been unavoidable had we used punctuation like comma and semi colons, etc.

**3.6.12 What the ROILA grammar does not have & some alternatives**

It is interesting to note that there is considerable freedom and flexibility for future speakers of the ROILA language. Given that we provide a good number of ROILA wild card words, speakers can add grammar rules and grammar categories which are initially not present in ROILA by using those wild card words. While designing the ROILA grammar we took into account several criteria which meant that some grammatical categories were dropped. These categories may be found in other artificial or natural languages. We discuss each briefly and suggest how we could make up their absence in ROILA.

**3.6.12.1 Certain Tenses**

Some form of Tenses such as perfect tenses are not fully supported. In some situations it would be possible to express perfect tenses by rephrasing the sentences. We give an example (see Table 3.21).

English Sentence	ROILA Translation	Literal Translation
I am going (I am about to go)	Pito kapim kanek	I about go

Table 3.21: Sample ROILA sentence showing how perfect tenses can be stated in ROILA

**3.6.12.2 Case**

Case is partially supported in ROILA by means of pronouns. Obviously the noun does not inflect. ROILA supports the expression of I, you, he and she as pronouns which covers the Subjective/Nominative case. For possession the sentences would have to be rewritten as shown in the table (see Table 3.22).

English Sentence	ROILA Translation	Literal Translation
This is Omar's book (This is the book of Omar)	Bamas fojato fomu Omar	This book of Omar

Table 3.22: Sample ROILA sentence showing how cases can be represented in ROILA

**3.6.12.3 Aspect**

Aspect is loosely interchanged with tenses in most languages. ROILA does not support it explicitly or implicitly.

3.6.12.4 Modality

Although modality is not directly support in ROILA, modality can be partially expressed by the usage of words such as may instead of might for the Sub-junctive type. We also added must by using one of our ROILA wild card words (waboki).

3.6.12.5 Voice

Active voice is the default scenario supported by ROILA so the subject or actor of the context is always before the verb. Therefore there is no direct support for expressing passive voice.

3.6.12.6 Articles

Articles are not part of the ROILA vocabulary, so there is no a, an or the.

3.6.12.7 Prepositions

ROILA does include a couple of prepositions but definitely not the whole variety. An example of the usage of *for* is shown in a table (see Table 3.23).

English Sentence	ROILA Translation	Literal Translation
He painted the house for an hour	Liba munune jifi bubas bi- jej kilu fulina	He paint < <i>wordmarker</i> : <i>pasttense</i> > house for one hour

Table 3.23: Sample ROILA sentence showing how prepositions can be represented in ROILA

3.7 Grammar Evaluation

In order to evaluate the grammar in terms of recognition we formulated some sample sentences ( $N = 30$ ) based on a hypothetical interaction scenario for a dialog system in English (see Table 3.24 for some sample sentences). These sentences were evaluated against their ROILA translation. Note that in this experiment we had not yet switched the vowel o from AA to AO in the ROILA pronunciations.

ROILA sentence	English sentence (Meaning)
Pito buse wetok	I am not sure
Biwu wekepo buse kulil bubas?	What color is the museum?

Table 3.24: Examples of ROILA and English sentences used in the grammar evaluation experiment

Sphinx-4 Language Models were created using the Sphinx Knowledge Base tool (Rudnický, 2010). For details on how to create ROILA language models in

Sphinx, please refer to Chapter 4.

To evaluate the ROILA grammar, an identical setup was followed as done in the evaluation of the vocabulary except that participants would now record sentences and not isolated words. Participants would once again hear a sample voice as a guide of how to pronounce sentences. This sample recordings had 100% recognition accuracy when passed offline in Sphinx. The dependent variable was word accuracy, a common metric to evaluate continuous speech recognition (Boros et al., 1996) with the independent variable yet again language type. For a detailed treatment of how word accuracy is computed in our context, please refer to Chapter 6. In the initial evaluation we conducted recording sessions with 8 participants. The results did not show any significant differences between ROILA and English; as indicated by the REMANOVA results  $F(1, 7) = 1.97, p = 0.21$ . The means of the word accuracy for both languages are shown in a table (see Table 3.25). The low number of participants should be kept in mind for the grammar evaluation experiment.

Language	Word Accuracy (%)	Std. Dev.
English	62.67	16.53
ROILA	60.39	14.68

Table 3.25: Means Table for word accuracy for ROILA and English

We must keep in mind several implications to our results obtained from both the word spotting and grammar evaluation experiments. Firstly, participants recorded ROILA words and sentences without any training in ROILA, whereas they were already acquainted with English. Potentially, by training participants in ROILA the accuracy could be further improved. This effect was observed to be more pronounced when participants had to speak ROILA sentences, which could explain the insignificant difference between ROILA and English in terms recognition accuracy. The acoustic models of Sphinx are trained with dictation training data and from what we observed the ROILA sentence articulations of participants did not fall within the domain of dictation speech. There were pauses between words and pronunciations were not smooth, which could have been caused by the inexperience of the participants in ROILA. In our future evaluations we gave extra training to participants in ROILA and achieved much improved results with respect to recognition accuracy (see Chapter 6).

---

## The Implementation of ROILA

---

This chapter will detail the implementation of the ROILA project. The first step in implementing ROILA was to enable speech recognition for ROILA. Once we had achieved that the subsequent step as to provide a speech synthesis facility for ROILA as well. Ultimately we would like our robots to talk back in ROILA. The last stage of the implementation was to choose a suitable robotic platform and integrate the afore-mentioned two modules within the platform. We now discuss each of the three main steps.

### 4.1 Speech Recognition for ROILA

The major component of the implementation was the speech recognizer. In an earlier chapter (Chapter 3) we gave some indications of our choice of speech recognizer; however we will attempt to fully elaborate on them in this chapter. To recall we selected the Sphinx-4 recognizer to test and evaluate ROILA.

#### 4.1.1 Overview of Speech Recognition Engines

From the outset of the project and in conjunction with the design of the language we were already deliberating the choice of speech recognizer. We would need to have open access to the speech recognizer since we would need it to be able to recognize ROILA. Therefore we had to be careful with our selection of the speech recognizer.

Since there were several options (commercial and non-commercial/research oriented) we would need to make an informed decision. We did not seek for a very detailed overview of speech recognition technology, rather at the end we wanted to make a logical decision amongst a few options, based on the pros and cons of each engine or platform.

Therefore we first drew some requirements of our intended speech recognizer. Potentially, the speech recognizer that we were looking for should ideally be:



- Open source or atleast inexpensive.
- The recognizer should not be a black box. For our application it is necessary to add a new language (ROILA), including a grammar, vocabulary, language models and acoustic models. This can be achieved by adding a completely new language or by adjusting an existing language.

As we found out, the requirements meant that any commercial system would go out of the window, regardless of its recognition accuracy estimations. We could have chosen an expensive commercial speech recognizer but mostly such systems do not allow easy extension with respect to the addition of a natural language let alone an artificial language like ROILA. Examples of such commercial systems would be Loqeundo (Loqeundo, 2011), Dragon Naturally Speaking and VoCon (Nuance, 2011). Although they are believed to have very high recognition rates but they are either not open source and hence in a black box. Mostly they are also very expensive reaching upto thousands of Euros (Loqeundo is one such). Moreover, we wished that our implementations be easily accessible to all those who would be interested so their access to the recognizer must not be a hindrance. In summary, the choice of speech recognizer was mainly dependent on the fact that the speech recognition engine should be open source and allow for the recognition of an artificial language. Therefore we continued our search for the most suitable speech recognizer and we came up with the following speech recognizers that satisfied at least some of our preset requirements (see Table 4.1).

Using the observations from the overview it was clear that the top front runners were Sphinx-4 and HTK. At this juncture, we took the help of speech recognition research to make our final choice. From (Samudravijaya & Barot, 2003) and (Guangguang, Wenli, Jing, Xiaomei, & Weiping, 2009) we found out that Sphinx was quantitatively evaluated as acoustically superior to HTK. Their empirical evaluations showed that Sphinx was found to have better acoustic models and it also achieved higher recognition accuracy rates. Therefore Sphinx-4 seemed to be the more intelligent option of the two. At this point, it is interesting to point out that Sphinx-4 also uses Hidden Markov Models as its theory of choice to carry out speech recognition.

Once we had reached to the conclusion that we would implement ROILA within Sphinx-4, we had to determine how we could do so. Before doing so, we would like to point out a limiting factor of our choice. The ROILA words were generated from a confusion matrix that extracted its data from the basis of another speech recognizer (Lovitt et al., 2007) and not Sphinx-4; this is a limitation but most speech recognizers operate on the same basic principles. As a matter of fact both recognizers use the same algorithm (Viterbi) as their searching technique (Lamere et al., 2003). Ideally we would have liked to use a phoneme confusion matrix that would have been generated from Sphinx-4 but we could not find data from prior research and constructing such large scale data would not be in the domain of our project. Another limitation was that we could have used a syllable based confusion matrix, but such a data set was not available and our choice of speech recognizer Sphinx-4 is also phoneme based

Table 4.1: Overview of Speech Recognizers

Speech Name	Recognizer	Descriptional Information	Pros	Cons
SPRAAK	<i>Speech Processing, Recognition and Automatic Annotation Kit</i> , 2010)	SPRAAK stands for Speech Processing, Recognition and Automatic Annotation Kit; "spraak" is also the Dutch word for "speech".	Open source, no license required. Satisfied most of our requirements	Extensive documentation not available as engine is in its infancy. Moreover due to the same reason no clear indication about its performance. Nothing in particular.
Hidden Markov Models Toolkit ( <i>HTK - The Hidden Markov Model Toolkit</i> (HTK), 2010)		Hidden Markov Models Toolkit is an established and well known speech recognition engine. The portable toolkit is also used in pattern recognition applications. Development is carried out in C/C++.	Extensive documentation and manuals available. Open source, widely used. Language modifications possible.	
CMU Sphinx (Lamere et al., 2003)		Sphinx-4 an open source recognition platform developed at the Carnegie Mellon University (CMU), USA. The Sphinx family also has several other versions known as Sphinx 1, 2 and 3. Sphinx-4 applications can be programmed in Java using the provided Sphinx-4 API.	Modular, Java based and hence platform independent. Excellent documentation and technical support. Open source and no license restrictions. Powerful API available.	Nothing in particular.
Microsoft Speech Technologies ( <i>Microsoft Speech Technologies</i> , 2007)		Microsoft provides a powerful Speech API that provides speech recognition and speech synthesis.	Open source with minimum license restrictions.	Intended mainly for web and telephony applications. Platform dependent.
Julius (Julius, 2010)		Julius is a speech recognition platform developed in Japan and hence primarily intended for Japanese. It is also based on Hidden Markov and uses statistical language models.	Open source.	Focus is mostly on speech recognition for Japanese. The quality of their acoustic models for English is not well established. Moreover, it was not clear if we could easily add ROILA to the engine.

in method. However an advantage in our favor was that the phoneme confusion matrix was based on recognition of American English and we would ultimately use an American English acoustic model in Sphinx-4, providing us with some comfort in extrapolating the use of the matrix to the genetic algorithm.

As stated earlier (Chapter 1), most speech recognizers work on the basis of two important parts: the acoustic model and the language model. The acoustic model essentially entails how combinations of letters are pronounced, also known as phonemes. The language model describes how the grammar looks like for example it would give the probability of word(X) being followed by word(Y), if it is probabilistic in nature. If the grammar is formal in nature it would be rule based, for e.g. context free grammars. Therefore the implementation of ROILA in Sphinx-4 would be concentrated on the modification and customization of the two afore-mentioned parts, by configuring Sphinx-4 so that it can recognize ROILA.

#### **4.1.2 Adaptation of an Acoustic Model for ROILA**

The first step in enabling speech recognition of ROILA was to setup an acoustic model for ROILA. We realized that there existed a potential of using an English acoustic model for the recognition of ROILA. There were several reasons for this: creating a new acoustic model customized for ROILA would require a lot of native ROILA training data and effort (see creating acoustic models for new languages (Liu & Melnar, 2006)), all the phonemes present in ROILA also existed in English (American and otherwise) and lastly the English acoustic model from Sphinx-4 was widely used and had been evaluated on several occasions. The acoustic model that we chose was one of the Wall Street Journal models (Carnegie-Mellon-University, 2008) provided by Sphinx. In essence, we made no changes to the acoustic model provided with Sphinx-4. The only adaptation that we had to do was to provide a phonetic dictionary with every ROILA application. The dictionary would list the ROILA words used in the context of the application and define their pronunciations in ARPABET format (Jurafsky et al., 2000). Sphinx requires every word of the dictionary to be broken down into ARPABET symbols. Most linguists are familiar with the International Phonetic Alphabet (IPA) standard (Ladefoged & Maddieson, 1996), ARPABET is basically an ASCII representation of the IPA standard for only American English phonemes. The ARPABET and IPA terminology was already introduced in (Chapter 3) and examples of ROILA transcriptions in ARPABET were also provided. For example the ROILA word KANEK in ARPABET would be written as K AE N EH K.

The dictionary defines how every word should be pronounced for successful recognition. There is also an option of providing alternate pronunciations for every word by adding a (2) after the word. For example the word FOSIT in ROILA could be inserted in the phonetic dictionary as follows:

```
FOSIT F AO S IH T
FOSIT (2) F AA S IH T
```

Here, we already see the potential of flexibility in ROILA in terms of pronunciation of its words. ROILA speakers can define their own customized pronunciations, a feature that finds itself in the realm of end-user programming. We elaborate on the future prospects of ROILA and the flexibility that it provides in Chapter 7.

### **4.1.3 Language Model/Grammar Representation for ROILA**

The next stage is to represent the grammar of ROILA in the form of a language model. Sphinx-4 supports three main types of grammars. Firstly, word grammar which is used for isolated word speech recognition as described in our word spotting experiment in Chapter 3, secondly, a n-gram probabilistic grammar and thirdly a formal rules based grammar, such as the Backus Naur Form (BNF) grammar. In our ROILA implementations we used the n-gram probabilistic grammar where the model was of bigram and trigram type. Sphinx-4 provides an online tool which generates the required n-gram language model once it has been given input of the list of transcriptions. Therefore before being able to create a language model the first step that needs to be accomplished is to identify the context of use, i.e. what do you want to talk about in ROILA? For e.g. in our first ROILA prototype we chose a navigation scenario and the list of sentences that we used to generate the language model are presented in see Table 4.2.

Once the list of ROILA sentences are identified a language model can be constructed by using the Language Modeling Tool (Rudnicky, 2010) provided by Sphinx. The language model is generated in ARPA format, a format specific to Sphinx-4. The Language Modeling Tool generates both bigram and trigram models. In summary the model comprises of probabilities of a word being followed by one word (bigram) or two words (trigram). We could have chosen a BNF grammar representation for ROILA but we went with the statistical n-gram representation due to two major reasons. Firstly, much more training data is required to determine network or formal grammars and secondly statistical models have been ascertained to achieve better recognition accuracy than network models (Weilhammer, Stuttle, & Young, 2006).

### **4.1.4 Setting up the Sphinx-4 Configuration file**

In order to have Sphinx-4 fully functional for an application a configuration file needs to be prepared in accordance with the two steps described earlier. The configuration file lists various parameters that the recognizer needs to know during the recognition process, such as the type of acoustic model and grammar type. Sphinx-4 provides various templates of configuration files and not a lot of changes are required to prepare it so that it can be incorporated in a ROILA recognition program. For our ROILA applications we used one such template and slightly tinkered with it. The main modification that needed to be done was entering the paths of the customized ROILA dictionary and language model.

These three files (phonetic dictionary, language model and configuration file) form the crux of speech recognition in Sphinx. The next step is to write Java code so that speech recognition can actually take place.

#### **4.1.5 Executing Speech Recognition in Java using Sphinx-4**

Sphinx-4 has a powerful Java API with extensive documentation that enables developers to write speech recognition applications. The basic input to the application is the user defined configuration file. The acoustic model is also inserted by referencing it within the application. The acoustic models of Sphinx-4 are available as java archive files (jar). The API is quite simple to use and can return the result of the recognition as plain text. The Sphinx-4 API also provides an option of carrying out additional parsing. As a developer, one can retrieve the list of plausible recognition results accompanied by their confidence scores. In our recognition applications described in (Chapter 6), we therefore carried out additional parsing using the confidence scores. The list of plausible recognition results were sorted based on their confidence scores and a choice was made based on the interaction context, in particular when the recognition result attained did not grammatically make sense or if was out of context.

### **4.2 Speech Synthesis for ROILA**

Once we had implemented speech recognition for ROILA, the next step was speech synthesis, i.e. given ROILA input as text a machine should be able to speak ROILA. In our situation we would want robots to speak ROILA.

We adopted the same process that we followed while choosing a speech recognizer. The requirements for a speech synthesizer were similar to the requirements that we had set for a speech recognizer. However, open source choices for text to speech (TTS) engines were limited, which made our task easier. The primary options were Festival (*The Centre for Speech Technology and Research*, 2008), FreeTTS (*FreeTTS 1.2*, 2005) (which extends from Festival) and rSynth (*rsynth - Text to Speech*, 2005). FreeTTS does not easily support addition of new languages as it is primarily designed for English and the voice quality of rSynth is not up to the mark. Therefore we employed Festival as our speech synthesizer of choice. Festival is a large platform offering speech synthesis support on a number of API levels, such as on the shell level, scripts or via C++ libraries. Festival is recommended to be built and setup on Linux, therefore we first attempted to setup Festival on Ubuntu even though our intended platform of choice was Windows. After a successful build of Festival we were fortunate to find a script which was initially written to provide speech synthesis for Lojban (Pomeroy, 2003), another artificial language. Lojban is another artificial language like Loglan that is based on predicate logic. Festival uses the Scheme programming language (dialect of the Lisp programming language) in its scripts. The phonetic information within the Lojban script had to be modified so that it would be according to the phonetic rules of ROILA. Since both ROILA and Lojban have phonemic orthographies, implementing them in Festival was easier than it would be for other languages. In summary, groups of

phonemes are pronounced as they are written in ROILA and the location of a phoneme in a word does not change the way it should be pronounced.

The correct way enabling speech synthesis for any new language would have been by recording native phonemes and diphones (i.e. two phonemes bunched together such as in syllables) and using them to train the synthesizer. This would also result in natural prosody (which is in itself a complete area of research in Speech Processing). Ultimately we felt that the afore-mentioned two aspects were out of the scope of the ROILA project.

The ROILA script could be loaded into Festival and then using the SayText Festival command we could enable the machine to speak ROILA using one of the default voices within Festival. However for our application we needed .wav files of ROILA transcriptions. This would also be helpful in case we implemented our ROILA application/setup on Windows. Festival provides a text2wave command with the functionality of piping its output to sound (wav) files. Therefore we used the following command to extract .wav files of ROILA sentences, where myFile.txt has one ROILA sentence and myScript is the ROILA Festival script and eval is a Festival tag. The eval tag lets Festival know that it must use the script and its properties for the text to speech.

```
text2wave myFile.txt -o myFile.wav -eval myScript.scm
```

In order to test the recognition accuracy of our speech synthesis setup we generated TTS recordings of the same 30 ROILA sentences as discussed in the grammar evaluation experiment described in Chapter 3. The recognition accuracy of the TTS sentence recordings was found to be very high (Word Error Rate < 10%). This figure was generated by passing recordings offline in the Sphinx-4 recognizer, similar to our evaluations described in Chapter 3. Therefore we established enough confidence to use our TTS recordings in the final evaluation of ROILA.

### 4.3 LEGO Mindstorms NXT

Once the technical layout of our ROILA application was ready in terms of the speech recognition and speech synthesis features the next issue to contemplate was the choice of a robotic platform. We had several options due to their availability in our department, for example the iCat robot, the Aibo, the Nao robot and LEGO Mindstorms, however LEGO Mindstorms was the most prevalent (Verbeek, Bouwstra, Wessels, Feijs, & Ahn, 2007).

The unanimous choice for us was LEGO Mindstorms NXT for three major reasons. Firstly, the use of LEGO NXT robots as the platform for ROILA automatically meant that in an instant there would be potentially thousands of robots that would attain the ability to recognize and talk in ROILA. Secondly, there is an option to program LEGO Mindstorms robots using the Lejos platform, which is a Java based firmware. Since our speech recognition platform was also Java based we would make communication between the two components easier. A third reason that will become clearer in the subsequent chapter

(Chapter 6) and which overlapped with our the goals of the evaluation of ROILA was the choice of our user group. We intended to involve young children in the evaluation of ROILA and the target group of LEGO Mindstorms is also young children.

LEGO has established a new branch of robotic elements in its company by the name of Mindstorms (LEGO, 2010). It consists of a programmable brick called the NXT which can be expanded with LEGO blocks, electrical motors and sensors. These days the LEGO Mindstorms are sold as educational kits but also as toys to have at home. The NXT brick is the brain of a LEGO Mindstorms robot and the NXT can be programmed to control the motors and operate the sensors. In this way the robots can become alive and perform different tasks (see Figure 4.1). The NXT can also be programmed in various programming languages, one of the popular being Java using the Lejos firmware (*LEJOS - Java for LEGO Mindstorms*, 2009). LEGO Mindstorms is modular allowing the user to bring his/her creativity into play by building and programming robots. It provides several sensors as add-ons hereby enhancing the interaction possibilities with the robots. Moreover, it is also Bluetooth enabled allowing communication possibilities between robots and computers.

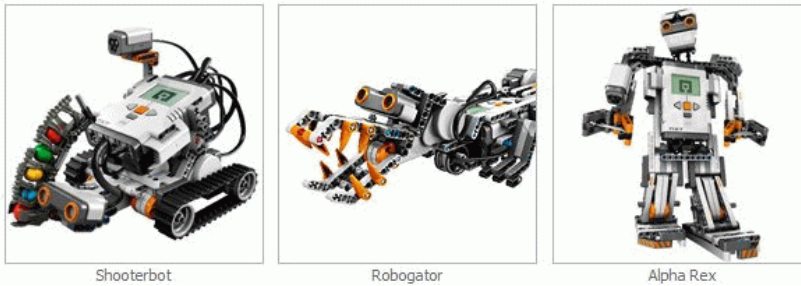


Figure 4.1: Examples of LEGO Mindstorm robots

#### 4.4 First ROILA prototype

As a conclusion to the design phase and also as a proof of concept we designed an initial prototype of ROILA by using the LEGO Mindstorms NXT platform. ROILA was demonstrated in use to instruct a LEGO robot to navigate in its environment, analogous to the principles of the turtle robot (Solomon & Papert, 1976).

The first decision that we had to make was regarding the system architecture. We had to determine what would be the communication setup between the speech recognition engine and the robot. At this point we would like to point out a certain limitation of Mindstorms NXT robots. The NXT brick has limited processing power and it would be surely unable to run the entire Sphinx-4 package. In the future the use of PocketSphinx (a version of Sphinx for embedded mobile devices) might be an interesting alternative for Mindstorms NXT.

However this was a research direction that we did not pursue in this project. Another limitation of the NXT brick is that it only has memory of 128KB for the JAVA virtual memory (Lejos), executables and other data so it could only playback a subset of sound files and that too with the sound quality compromised.

The limitations meant that we had to find an intelligent solution to establish communication between the recognizer and the LEGO robot. Bluetooth was the answer to our quandary. Sphinx-4 could be running on a computer and recognition results could be transmitted to the robot using the Bluetooth channel. The Lejos API provides efficient tools to support Bluetooth communication, therefore our task was made easier.

The next step was to select a context to implement our prototype. Adopting the theory of turtle robot semantics (Solomon & Papert, 1976) we built a simple turtle robot using LEGO Mindstorms (see Figure 4.2) that would navigate in the environment when it was given ROILA commands (see Table 4.2). A user could speak into a microphone, recognition and parsing would take place and whatever would be recognized would be sent over the Bluetooth channel as bytes. Therefore the computer was acting as the server and the NXT brick was the client. The NXT would then determine its action and on completion of the action or otherwise (robot did not know what to do) it would send back a byte to the server acknowledging the completion of processing. Speech recognition was carried using Sphinx-4 using the exact same specifications described earlier in this chapter with regards to customizing Sphinx-4 so that it could recognize ROILA. In this scenario (first ROILA prototype) the NXT robot could also playback ROILA TTS sound files due to the limited nature of the context but in more complex interaction scenarios (Chapter 6) we would have the server (computer) playback the ROILA transcriptions. Moreover, we did not implement any dialog management strategy for our first prototype as the interaction was quite trivial. However for our more advanced ROILA implementations as described in Chapter 6, we did carry out simple dialog management strategies based on representing the current system situation as states. The system flow would then traverse between the states as per the interaction. As a summary we also present our system architecture (see Figure 4.4).

ROILA Command	English Translation
fosit koloke	Walk forward
fosit kipupi	Walk slowly
fosit jimeja	Walk quickly
bobuja	Run
buse fosit	Stop
fosit nole	Walk backwards
fosit webufo	Walk left
fosit besati	Walk right

Table 4.2: ROILA commands used in the prototype





Figure 4.2: First robot used to prototype ROILA

We would also like to state the rationale behind our choice of microphone. All subsequent ROILA implementations (also those described in Chapter 6) after our first prototype would use the Blue snowflake microphone (see Figure 4.3) (*Blue Microphones*, 2009), the specifications of which can be seen in a table (see Table 4.3). The microphone provided the right balance between accuracy (observed from several pilots) and cost. Moreover it was mono-directional and hence less effected by ambient sound unlike other cheaper microphones.



Figure 4.3: Blue snowflake microphone

Transducer Type	Condenser
Polar Patterns	Cardioid
Sample/Word Rate	44.1 kHz/16 bit
Frequency Response	35Hz - 20kHz

Table 4.3: Specifications of the Blue snowflake microphone

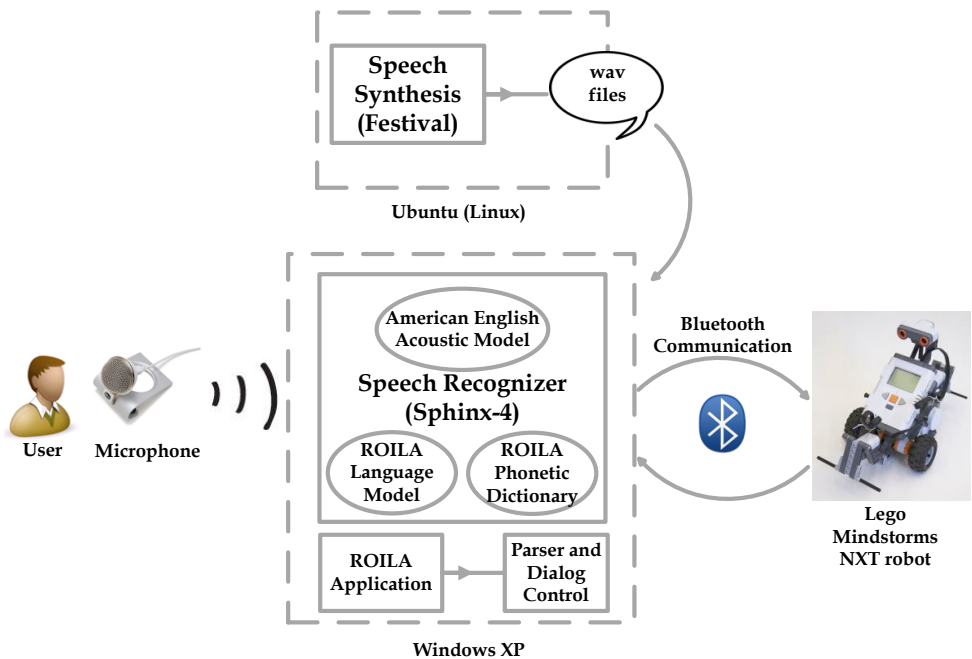


Figure 4.4: System Architecture

We were fortunate to be invited to visit LEGO Mindstorms headquarters in Billund, Denmark in April 2010 where we presented our first ROILA prototype (related to the navigational scenario) and discussed the concept of ROILA and our initial empirical results. The result of our engaging and positive discussion was the establishment of an informal sponsorship by LEGO, the output of which was 1 cubic meter of LEGO Technic and more importantly 20 LEGO Mindstorm 2.0 boxes (see Figure 4.5), all of which we could use to further develop ROILA. The Mindstorm 2.0 kit is slightly different from Mindstorm 1.0 as it has an extra touch sensor and the color and light sensor combined into one. In total it has 619 pieces including the sensors and the three motors (LEGO, 2010).



Figure 4.5: LEGO resources provided by LEGO Mindstorms NXT, Billund

The first ROILA prototype was also presented to a research assessment panel as part of the assessment of the department of Industrial design of our university in the mid of 2010. The concept and prototype was well received.

Once we had covered the three main angles related to the implementation of ROILA, we could now develop more complex interaction scenarios and move towards more comprehensive user evaluations. In such evaluations we would ask trained students of the ROILA language to interact with LEGO robots in ROILA and at the same time the recognition accuracy of ROILA would be recorded.

---

## User Aspects of Evaluating Constrained and Artificial Languages

---

From the very outset of the ROILA project, we were faced with a number of issues that dealt with the usability and user experience of participants while interacting in constrained or artificial languages. Prior to carrying out a real evaluation of ROILA we also conducted two large scale experiments where children were exposed to constrained or artificial languages and their subjective experiences were measured. In the first case study we also investigated measurement mechanisms to evaluate the learnability of artificial languages. This would prove to be helpful in the future when we would try to measure the learnability of ROILA. In the second case study we tried to determine the subjective user experience of children while interacting in constrained or artificial languages as compared to their native language.

### 5.1 Case Study I

In this section we articulate our efforts in measuring the learnability of a new language for humans. For this we have explored two objective measurement tools: self designed language tests and the intensity/level of emotional expressions while interacting in a language. A perception test (where independent judges analyze emotional response) was used to measure the level of emotional expressions. Our study was carried out in the context of a game for children, where they interacted in two artificial languages of varying difficulty: Toki Pona (Kisa, 2008) and Klingon (Shoulson, 2008). Our hypothesis was that an easier to learn language would result in children scoring higher on our version of the language test and expressing richer emotions.

The second hypothesis emerged from the following argument that, when people are interacting in their native language, they tend to be so spontaneous that they can very naturally express additional emotions. However, it could be that a difficult language reduces this possibility, because the use of the language is too cognitively demanding which leads to a very constrained interaction style with no room for the display of emotions. Various methods are

used to measure natural language learnability, for e.g. computer simulations (Lupyan & Christiansen, 2002).

Language Aptitude as defined in (Carroll, Sapon, & Corporation, 1959) is the ability of an individual to learn a foreign language. The use of various standardized tests for measuring aptitude in natural languages is fairly common (for e.g. the TOEFL). However in our case we required an engineering effort to adapt a similar tool for an artificial language. Numerous paradigms exist, that assist in the design of language tests, for e.g. MLAT, VORD and CANALF see (Parry & Child, 1990) and (Grigornko, Sternberg, & Ehrman, 2000). In (Parry & Child, 1990) a comparison concluded that the MLAT testing framework was the most appropriate and efficient instrument to predict language learnability. Consequently, adapting the MLAT testing methodology and utilizing it as a starting point for creating a test for artificial languages became one of the goals of our reported study.

### **5.1.1 Experimental Design**

As a scenario and case study option upon which artificial languages could be evaluated, we adopted the interaction mechanism in game play for children as suggested in (Shahid, Krahmer, & Swerts, 2008), where it was argued that games are an effective tool to elicit emotional response. To evaluate the user perspective with respect to the learnability of foreign languages an experimental study using a game was carried out in Lahore, Pakistan, with 36 children aged 8-12 years (Male=22, Female =14, average age =10.36, std dev=1.42). In total 18 game sessions were executed, each involving a pair of children. All children had sufficient knowledge of both English and Urdu: their native language.

### **5.1.2 Game Design**

We designed a simple wizard of oz card game where children would guess whether the subsequent number in a sequence would be larger or smaller than the previous number (see Figure 5.1). When the game would start, players would see a row of six cards on the screen where the number of the first card was visible and the other five cards were placed face down. All the cards ranged from 1 to 10 and a card displayed once was not repeated in a particular sequence. Once players would make a guess, the relevant card was revealed. Players were informed about the correctness or incorrectness of their answer via a characteristic non-speech sound (booing or clapping). A correct guess would earn positive points (+1) and an incorrect guess would result in losing the particular game. The children were encouraged to discuss with each other in order to attain a consensus about their final guess. All interaction in the game was speech based.

### **5.1.3 Procedure**

The game was run as a power point presentation on a laptop. The entire experiment sessions were video recorded. Written consent was given by the teachers and parents to use the recordings for research purposes. A pair of children

was asked to sit in front of a desk on which the laptop was placed (see Figure 5.2). Above the laptop, a camcorder was placed to record the children's faces and their upper body. A monitor connected to the laptop facilitated the wizard in controlling the game. The wizard was located out of the visual field of the game-playing children in another room. After an introductory round, the children were given game instructions and were informed about the points that they could win or lose. After this, the experimenter left the children's field of vision and started the game. At the end of the game session, the experimenter rewarded the children with gifts based on their points.

#### **5.1.4 Interaction in the game via Artificial Languages**

Pairs of children would play the number guessing game where the permissible set of game commands and utterances that could be said to the co-player ( 10 words such as larger, smaller, equal, go to the next number, etc) was predefined in an artificial language. There were two artificial languages chosen: Klingon (apparently difficult to learn) and Toki Pona (apparently easy to learn). Toki Pona is designed on a principle of simplicity and aims to reduce complexity. Its lexical and phonetic vocabulary is considerably small in size (118 words and 14 phonemes). In contrast, Klingon is a unique apriori language and therefore does not have a large scale influence from any existing natural language. In fact some of its phonemes are not found in any of the major natural languages, which is precisely the reason why Klingon has words that should be difficult to pronounce, besides being longer in length. This led us to set the assumption for our experiment that Toki Pona should be relatively easy to learn and Klingon in comparison more difficult. Each pair of children would play two rounds of the game each in either of the two artificial languages and in a natural language, in this case their native language Urdu. The four orders of language presentation were counter balanced. In total, 9 pairs of children played the game in Klingon and Urdu and 9 pairs in Toki Pona and Urdu. The wizard had knowledge of all the relevant game commands in each of the three languages and would direct the flow of the game accordingly. Prior to playing the game and during the explanation of the instructions phase, each child was given exactly 10 minutes to learn the game commands with the aid of audio clips. During the game, if the children would forget the commands of the artificial language, they could call out for help, but at the cost of exponentially increasing negative points.

#### **5.1.5 A Learnability Test for Artificial Languages**

One of the primary objectives of the study was to design and evaluate a self developed language learnability measurement test. Similar language tests were constructed for both Klingon (KN) and Toki Pona (TP), by adapting the framework as in (Carroll et al., 1959). The tests included ten questions each having only one correct answer. The questions tested the learnability of the artificial language in terms of vocabulary and pronunciation via semantics and rhyming respectively; two of the four language learning abilities advocated in (Carroll et al., 1959). The tests were handed out at the end of the game playing session. An independent samples t-test revealed that the children who learnt TP performed significantly better on their version of the test than those who played

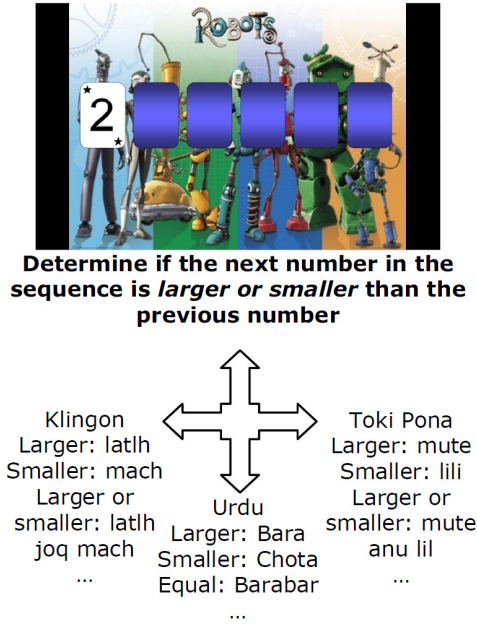


Figure 5.1: Game Design

the game in KN, ( $t(34) = 2.04, p < 0.05$ ). There was no order effect upon the test scores ( $F(3, 32) = 2.2, p = 0.11$ ). Note that the tests were scored absolutely with no partial credit.

### 5.1.6 Evaluating Emotional Expressions via a Perception Test

A secondary method to evaluate the learnability of languages was to run a perception test on the game videos. We hypothesized that a language that is easier to learn would be much more enjoyable to interact with and would hence elicit richer emotional expressions.

#### 5.1.6.1 Procedure and Participants

From the children that played the game, we selected video snippets of a random winning guess and a random losing guess from each individual game. This was done twice for each of the languages. In addition, from the clips we randomly selected one child from each of the 18 pairs by zooming in on his/her face. In this selection, half of the children sitting on the right chair and half of the children sitting on the left chair were selected. The stimuli were recorded from the moment the card in question was overturned till the primary response of the child was completed. This resulted in 72 stimuli:  $2 \times [\text{win/lost}] \times 2$  [(KN or TP) and Urdu]  $\times 18$  children. Stimuli were presented to participants in a random order, in vision-only format to avoid participants from relying on auditory cues. 30 Dutch adults, with a roughly equal number of men and women,



Figure 5.2: Children involved in the game

participated in the perception experiments. For the perception test, participants were seated in a classroom where the stimuli were projected on a wall. The participants were informed that they would see stimuli of children who had just won or lost a game. As viewers, they were instructed to guess from the children's facial expression whether the children had won or lost. Each stimulus was preceded by an ID and followed by a 6 second pause during which participants could fill on a form firstly whether they thought it was a winning or losing situation and secondly how expressive did they think the children were on a scale of 1 to 7, with 7 being the most expressive.

#### 5.1.6.2 Statistical Analysis and Results

Tests for significance were performed using a repeated measures analysis of variance (REMANOVAs) and the Bonferroni method was used for pairwise comparisons. The experiment had a within subject design with language type (levels: natural language Urdu-NL, TP and KN), being the independent variable. The percentage of correct classifications and level of expressiveness as ranked by the participants were recorded as the dependent variables. The REMANOVAs showed a significant main effect of language type ( $F(2, 58) = 143.220, p < .001, \eta p^2 = 0.832$ ) on correct classification. Pair-wise comparisons revealed a significant difference between the languages. The average of correct classifications was the highest for NL ( $M = 0.793$ ), followed by TP ( $M = 0.602$ )



and lastly KN ( $M = 0.546$ ). Similarly, for the variable 'level of expressiveness', a significant main effect of language type ( $F(2, 58) = 67.629, p < .001, \eta p^2 = 0.700$ ) was found. Pair-wise comparisons illustrated a significant difference between the languages. The level of expressiveness was ranked to be the highest for NL ( $M = 5.04$ ), lowest for KN ( $M = 4.05$ ), with TP ( $M = 4.40$ ) lying in the middle.

### 5.1.7 Discussion

The goal of our study was to explore two methods to evaluate the learnability of artificial languages. Firstly, our initial results illustrate that there is potential in utilizing a language test as an instrument to detect the learnability of an artificial language, because it was able to highlight the difference in ease of learnability between TP and KN. However our developed tests were not extensive and care must be taken while interpreting the results as the lexical vocabulary of each of the two languages was not of sizeable proportions. Moreover, mainly due to the aforementioned reason, our language tests did not test all the four cognitive abilities advocated in (Carroll et al., 1959), which can be termed as another limitation of our test.

Secondly, from the results of the perception test it is evident that the children were significantly more expressive in TP as compared to KN, which can also be adjudged by observing the higher number of correct classifications for the case of TP. In the case of TP, it was relatively easier to guess whether children have won or lost the game based on their facial expressions. Therefore our hypothesis that an easier to learn artificial language would elicit more emotions is confirmed. The differences in emotional expression and the accuracy of independent observers to perceive these differences is in line with existing work where emotional expressions of Pakistani children were used as a mean to judge fun or engagement in a game (Shahid et al., 2008).

The level of expressiveness and the average of correct classification were the highest for NL. We would expect this to be the case because, the children were quite comfortable while playing the game in their own language and there was no cognitive load in terms of recalling a new word from a new language which could hinder their expressiveness. The use of natural language in our design was mainly a control condition to check the expressiveness of children across two game sessions and the high expressiveness in natural language and relatively low expressiveness in artificial languages confirms our choice of not only this condition but also the perception test method. An interesting and potential limitation in the perception test was a cross-cultural element. The stimuli were of Pakistani children and the observers were Dutch adults. This cultural incongruence could have resulted in a winning guess being perceived as a losing one or a losing guess as a winning one, while perceiving the emotional reactions of children. It is known that complex emotional states e.g. shame and guilt in particular, are packed in cultural wrappers and it is sometimes difficult to judge such emotional expressions across cultures (Breugelmans & Poortinga, 2006).

## 5.2 Case Study II

In this section we discuss two scenarios which pertain to the separate evaluation of two artificially constructed languages: a constrained version of the natural language Dutch and our artificial language ROILA. We did not attempt to compare artificial and constrained languages but rather determine the user acceptance of both separately. Both experiments were carried out in an identical interaction context of a child playing a game in cooperation with a robot, as to provide the child with a scenario to interact in the new languages. Additionally, a gaming scenario is known to be a realistic context as children use games in their everyday social and educational life (Johnstone, 1996), (Lazzaro, 2006). By having children play a game we can subjectively measure fun by virtue of their gaming experience, a strategy which has been used before (Al Mahmud et al., 2007). Therefore we could ascertain if the children had more fun playing the game in a specific language. By adopting a specified fixed scenario we aimed to have much better control over the study as we were working in the domain of new languages. Moreover, we selected children as users because it is known that children are good learners of new languages as compared to adults (Rubin, 1975).

We used the iCat robot (Breemen, Yan, & Meerbeek, 2005) as the gaming partner of the children. The iCat robot is a cat like robot by Philips that has the ability to communicate to users via both verbal and non-verbal means such as by exhibiting facial expressions. We also speculated that due to its non-human like appearance the iCat robot would advocate lower expectations from the children (Bartneck, Kanda, Mubin, & Al Mahmud, 2009) and consequently they would be willing to learn new languages to interact with it. The iCat robot was controlled by the facilitators in a Wizard of Oz fashion and no speech recognition was taking place since our primary goal was to identify whether novel languages would be acceptable to users. In an actual setting; i.e. deployed with a speech recognizer, the accuracy of speech recognition would also affect the user experience. To summarize, the primary goal of this research was to evaluate whether children are comfortable with using constrained or artificial languages in comparison to natural languages and whether they are willing to invest some effort in learning such languages. If it would turn out that children are still relaxed, comfortable and have fun while communicating in constrained or artificial languages, then this finding would be a positive step forward in the implementation of speech recognition friendly languages, as the biggest objection with such new languages is that users would not be motivated or comfortable to interact in them.

### 5.2.1 Scenario 1: Constrained Languages

The first study was carried out in Tilburg, the Netherlands, where children either collaborated in their native language Dutch or in a restricted set of Dutch utterances that was suitable for the communication purpose (i.e. collaborating in a gaming scenario which we describe later).

### 5.2.1.1 Participants

92 children took part in this study, that were between 9 and 12 years old. From them, 52 took part in the natural language condition (Dutch); the other 40 conversed with the iCat in constrained Dutch. We balanced gender in both conditions. All children had prior written consent by their parents and teachers to participate in this study and to use the results and audiovisual data for research purposes.

### 5.2.1.2 Material

The game employed had shown to induce emotions within children (Shahid et al., 2008). In the game, the child would see a row of six cards on the computer screen, where each card had a number written on it and only the first card was shown initially (see Figure 5.3). The other cards were placed upside down. The players' task was to guess whether the next card contained a number that was bigger or smaller than the previous one. The available card numbers were between one and ten and every number could only appear once within a solitary sequence. When the player guessed a number, the card would become visible. Then, the child would hear a characteristic non-speech sound (booing or clapping) to inform them about the correctness of their guess. To win a particular game, the child was required to guess every number correctly. The child played seven rounds of this game and was encouraged to discuss every guess with the iCat. We used a Wizard of Oz method to simulate both the verbal and non verbal behavior of the iCat. The wizard was located out of the child's vision, behind a screen so that the children would not know about the wizard of oz setup. We received the input from the children through a camera and a microphone. The wizard could manipulate simple preprogrammed behaviors and animations that functioned as iCat's communicative response. A Dutch text to speech engine was also employed in order to elicit the responses of the iCat.



Figure 5.3: Game played in constrained Dutch

### 5.2.1.3 Design of the constrained language

For the composition of the constrained language we considered the children's language level and the probability of using certain utterances when playing a game in a natural situation. With these aspects in mind we composed a constrained language consisting of fifteen permissible commands (see Table 5.1).

When designing the constrained language, we also kept the difficulties in mind that speech recognizers would experience. For instance, the commands were designed such that they would have the least confusion amongst themselves. Moreover we used as few words as possible with which every sentence could be fully informative. Prior to interacting with the iCat, the children were given three minutes to study the constrained language so that they could recall and use it while playing the game. This was done out of the vision of the iCat, in a separate room. We checked whether the children could recall the commands, and whether they could easily restrict themselves to the defined vocabulary and recall appropriate commands. The duration of three minutes was finalized after conducting a few pilots which confirmed that it was long enough for the children to get acquainted with the language.

Constrained Dutch	English Translation
Wie is aan de beurt	Who's turn is it
Mijn beurt	My turn
Jouw beurt	Your turn
Wat denk jij?	What do you think?
Ik denk hoger	I think higher
Ik denk lager	I think lower
We hebben gewonnen	We have won

Table 5.1: Constrained Dutch Sentences

#### 5.2.1.4 Procedure

The children were seated on a bench, which was placed in front of a table with a computer screen on it. As shown in (see Figure 5.4), children sat beside the iCat. The iCat was positioned half diagonally, so that it could slightly turn its head for looking both at the screen and its game partner. When the children entered the room they first had an informal introduction with the iCat. After the game instructions the children played a practice trial together with the iCat. In this session, the experimenter was still present in the room in case there would be any questions. If the practice trial raised no further issues, the experimenter left the children's field of vision and started the game. After six game sessions, the experimenter guided the child back to another room where the child had to fill in his/her subjective evaluation. The experimenter (and iCat) was outside the children's view when they were filling in the questionnaire to avoid presence effects. Next, the experimenter asked the children some open questions, and rewarded the children with gifts. All sessions were video recorded. The video camera was placed on top of the monitor to record the child's face and upper body.

#### 5.2.1.5 Experiment Design and Measurements

The experiment was carried out between subjects with language type as the independent variable. To evaluate the children's social experience, we conducted several measurements by means of self-reports. We measured the fun that the



Figure 5.4: Experimental setup

children experienced during the game. For this, we adapted a Game Experience Questionnaire (GEQ (IJsselsteijn, Kort, & Poels, 2008)), which provides multiple questions to ask the children about the game's endurability, their engagement in the game, and whether their previous expectations about the game where met, all on a five-factor Likert scale. We also evaluated interaction in the constrained language, by means of the SASSI questionnaire (Subjective Assessment of Speech System Interfaces) (Hone & Graham, 2001), with which we measured factors such as cognitive demand and likability. Given that all interaction was carried out in the form of a wizard of oz scenario we extracted only the relevant factors from the SASSI questionnaire and factors such as system response accuracy and speed were excluded.

#### 5.2.1.6 Results

Independent samples t-tests were executed to ascertain if language type had an effect on the gaming experience. For the game experience, no significant differences between the language conditions were found for expectations ( $t(88) = 1.119, p = 0.26$ ), endurability ( $t(88) = 1.101, p = 0.27$ ) and engagement ( $t(88) = 1.537, p = 0.12$ ). Results from the SASSI questionnaire revealed some interesting trends. The constrained language seemed to cause only slight cognitive demand, as ratings on this scale were below 3 on a scale from 1 (no cognitive demand) to 5 (complete cognitive demand) ( $M = 2.06, SD = 0.96$ ). For the factor likability, a significant difference was found between Dutch and constrained Dutch ( $t(88) = 2.072, p < 0.05$ ) but the mean for both was rather high (Dutch = 4.65, Constrained Dutch = 4.40).

#### 5.2.1.7 Discussion

Overall, the children were generally expressive to iCat during the game, both verbally and nonverbally (see Figure 5.5). The children had few difficulties with using the commands in the constrained language. They felt comfortable with the constrained language while playing, as they did not evaluate the constraints on verbal communication negatively. Furthermore, the children deviated only

a few times from the constrained language while playing the game, and there were only three children that totally forgot one or more of the commands and could not continue with playing the game. However, the constrained language was not always used in its full potential, as at times children would only guess the card outcome and not argue the rationale of the choices made with iCat. For example, they would only use solitary words instead of complete sentences.



Figure 5.5: Children interacting with the iCat in Case Study 2 - Scenario 1

To evaluate whether effects of the constrained languages are indeed blurred by the minimalistic nature of the game, we could elicit a richer interaction by making a more extensive game where more choices could be made therefore leading to elaborate discussion. Therefore we decided to conduct a second study by adopting a different game where the interaction context was defined by an artificial language. It is worth pointing out that we could not have conducted a study with the three languages in question (natural, constrained and artificial) operating together as the three independent variables, primarily because of practical reasons. Consequently had this been the case, for a within subject analysis children would have to learn three languages and for a between subject analysis we would require a lot more participants for each condition.

### 5.2.2 Scenario 2: Artificial Languages

In this study children played a game with the iCat in either their native language or in an artificial language: ROILA. Primarily, based on our results from scenario 1, we used a different game. The main goal of this study was to determine if a certain artificial language hampered the user experience of the children. We realized that the game used in scenario 1 would not be entirely appropriate to evaluate artificial languages mainly due to its minimalistic nature. Despite losing the consistency between the two case studies we had no other choice but to design a new game where we could potentially explore the effect of artificial languages by having players engage in enhanced discussion.

#### 5.2.2.1 Participants

24 children took part in this study, that were between 10 and 13 years old. From them, 13 took part in the natural language condition; the other 11 conversed with the iCat in the artificial language ROILA. All children had prior written consent by their parents and teachers to participate in this study and to use the results and audiovisual data for research purposes.

### 5.2.2.2 Material

The game was a simple matching game in which the children had to match a given word with another word from a set of words based on some logical reasoning (see Figure 5.6). We anticipated that such a game would encourage the children to be much more verbally involved with the iCat as they would have to discuss the rationale of their choices and hence provide us with an opportunity to evaluate the artificial language with fairness. Therefore unlike the game employed in case study I they did not have one option but several and they had to discuss and deliberate their choices with the iCat. The children were only allowed one final guess and they would then be informed about the correctness of their guess. Yet again the iCat was controlled in a Wizard of Oz fashion with input from the children conveyed via a microphone and camera.

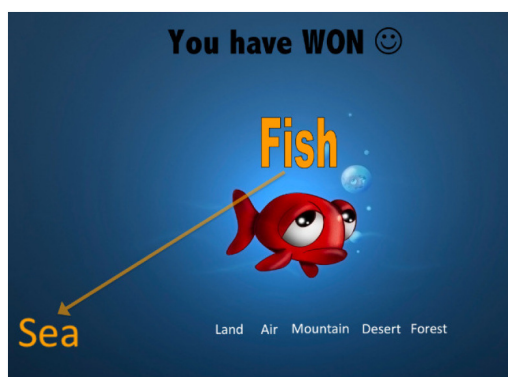


Figure 5.6: Game played in ROILA

### 5.2.2.3 Use of the artificial language ROILA

The children were asked to learn a set of roughly 25 commands in ROILA (see Table 5.2) and were given one day for training with regards to vocabulary and pronunciation. Before commencing with the experiment, necessary pronunciation checks were carried out for fluency. These commands were not fixed and the children could make changes to them based on the game situation. The design of ROILA has also been treated in (Chapter 3). The version of ROILA employed in this game, was one of the earlier versions. We had not yet restricted the word structure to CV-units only. This constraint therefore must be kept in mind while interpreting the results.

### 5.2.2.4 Procedure

The procedure and setup were identical to scenario 1 (see Figure 5.7). The text to speech engine of the iCat was replaced with audio recordings to ensure correct pronunciations. The recordings were slightly transformed to sound robot like.

ROILA	English Translation
babo etat ujuk	Who's turn is it?
ujuk ajne	My turn
babo ajne	Your turn
babo wimo	What do you think?
obat kutak	Which one should we pick?

Table 5.2: ROILA Sentences



Figure 5.7: Child interacting with the iCat in Case Study 2 - Scenario 2

#### 5.2.2.5 Experiment Design and Measurements

Language type was the main independent variable, as children played in either the natural or artificial language. Yet again we adopted the SASSI questionnaire (Hone & Graham, 2001) and used the following 3 factors from it: likeability, cognitive demand and habitability, where habitability is defined as the extent to which the user's conceptual model of the system agrees with what the system is actually doing.

#### 5.2.2.6 Results and Discussion

For the three factors we achieved Cronbach alphas of  $0.7 < \alpha < 0.8$ , which gives us sufficient reliability in the SASSI questionnaire. Language Type did not have an effect on any of the factors: likeability  $t(22) = 1.43, p = 0.17$ , cognitive demand  $t(22) = 1.23, p = 0.22$  and habitability  $t(22) = 0.22, p = 0.83$ . On average the natural language was ranked as the more preferred for the three factors but both languages were ranked on the positive end of the 5 level likert ranking scales (Likeability for native language = 3.31 and for ROILA = 3.00). Generally the children were quite positive in interacting with ROILA as was exemplified by their subjective rankings. However we did note that one day was not entirely sufficient for the children to achieve complete fluency in ROILA. Therefore they had to undergo a supplementary training session with the experimenters prior to playing the game.



### **5.2.2.7 Discussion and Conclusion**

We have presented an evaluation of how children interact with a robot using constrained and artificial languages. We compared these two languages with natural languages (the native language of participants) and showed that the children as users were comfortable in communicating with a robot using both languages. Similar results have been reported in (Tomko & Rosenfeld, 2004) where users preferred a constrained language over a natural language in a dialog information system however no studies have been carried out to ascertain the user acceptance of artificial languages. Moreover, the results that we obtained by evaluating constrained and artificial languages are quite positive as we concluded that there was no significant cognitive overload exerted by such languages in comparison to natural languages. Some significant differences were found, e.g. in terms of likeability, but both non natural languages were ranked positively. In hindsight we believe that it was a good decision to change the game for scenario 2 as it resulted in almost complete use of the available vocabulary of ROILA. In comparison to the constrained language condition where usually the children did not initiate a dialogue with the iCat or found it difficult to constraint themselves, in the case of ROILA the children did take an initiative and started deliberating with the iCat. We can conclude that at least in a wizard of oz situation children were not reluctant to learn new languages and there was no cognitive overload felt in doing so. Ultimately when a complete application is ready with speech recognition in place we can expect that children would be willing to invest some effort in learning a language that is optimized for speech recognition.

In this chapter we presented our attempts to evaluate artificial languages and constrained languages by keeping technology aside. The results of both our case studies showed that we could have some confidence in presenting such languages as options of an interaction medium to children. The results illustrated that there was no significant cognitive overload exerted in comparison to the native language of the children. Of course it must be kept in mind that interaction mediums were rather limited in nature and that when technology would be incorporated the results would be interesting to monitor.

---

## ROILA: Evaluation in Context

---

### 6.1 ROILA Evaluation at the Huygens College Eindhoven

As a final evaluation of ROILA we conducted a large scale experiment of the language. ROILA was exposed to Dutch high school students at the Christiaan Huygens College Eindhoven who spent three weeks learning and practicing the language. A ROILA curriculum was carefully designed for the students to aid them in their learning both in school and at home. In-school learning was more interactive and hands on as the students tested their ROILA skills by speaking to and playing with robots. In-home training was online readings based on short homework tasks which supplemented the material studied in class. At the end of the curriculum the students were asked to do a ROILA proficiency test and some of them were then invited to participate in a controlled experiment that compared ROILA against English. Throughout the whole learning process experiences of the students were measured and observed to determine if indeed ROILA was easy to learn for the students and easy to recognize for the machine. We will now delve into each aspect of the evaluation.

### 6.2 LEGO Mindstorms NXT and ROILA

As a test bed to prototype and evaluate ROILA we used the LEGO Mindstorms NXT platform (LEGO, 2010). The Lejos firmware for LEGO Mindstorms was used. This ensured a smooth connection between the Java based speech recognition system (Sphinx-4) and Lejos. LEGO A/S from Bilund Denmark were very kind to donate 20 Mindstorm NXT Version 2.0 kits to us.

### 6.3 ROILA Curriculum

The Christiaan Huygens College in Eindhoven, The Netherlands was pleased to offer their students for the ROILA activity. Moreover, the ROILA curriculum was merged into the Robotics module of their Science class. Therefore the ROILA lessons were not treated as an extra-curricular activity and hence the students would be expected to be more motivated. In total we worked with about 100 high school students (Age Range between 13-15 years old) who spent 3 weeks

learning ROILA in their Science lessons. The students were spread across 4 different classes. The four classes belonged to two groups, VWO and HAVO, categories used in the Dutch high school system. VWO study more theoretical subjects as a preparation for university study whereas HAVO are given more practical training, preparing for further study at a vocational school.

The choice of the age group was done keeping in mind several aspects. Firstly, we believed that children of an older age group would find it hard to learn a new language and on the other hand younger children would find it hard to use a computer or a robot. Moreover another key requirement was that, English should be a good second language of our sample of children (ROILA would be compared against English in the controlled experiment after the curriculum).

As was explained earlier in the thesis (Chapter 4), we would employ the use of an American English acoustic model for the recognition of ROILA for any implementation, evaluation or experiment. To recall, the reasons for doing so were several: Firstly, we do not have an acoustic model for ROILA and creating one would involve a lot of training data, secondly, an American acoustic model is available and widely used and lastly, the phonemes that are part of ROILA are also used in American English. These factors meant that if we had used native English participants (American English) in our evaluation the results would be biased against ROILA. This trend was also observed in our earlier results (Chapter 3). Therefore we decided to use children whose first language was Dutch but who also knew English as a good second language. Another key prerequisite of our intended participants was that they should be somewhat enthusiastic about playing with interactive LEGO.

The main idea behind teaching the students ROILA was that to fully establish the recognition accuracy of ROILA its speakers must have some prior knowledge of the language. Otherwise, they would not be able to pronounce the words, understand the grammar, etc. This would ultimately be harmful for the recognition and we could not execute a fair comparison with a natural language that the users would already know, for e.g. English in our case. The ROILA curriculum comprised of several components. The main tutoring source was a series of three lessons given by the author of the thesis, who is also the primary creator of the language. The lesson plans were carefully designed and constructed with the insights of the science teachers. We also ran pilot lessons to evaluate them. The lessons comprised of two parts: a theoretical part and a practical part. In the theoretical part, students were introduced to the linguistic elements of ROILA and in the practical part, we designed simple scenarios where the students could bring into practice whatever they learnt by talking ROILA to the robots. The fact that we had 20 Mindstorm kits meant that we could use several robots in one class session. As a token of appreciation we donated 10 Mindstorm kits to the Huygens College. At the end of the curriculum, the students attempted a ROILA exam and some of them were invited to participate in an evaluation experiment.

## 6.4 Technical setup

The LEGO robots were built with the help of voluntary students. These students were not participants in the ROILA curriculum. Up to 7 robots were used in one class, where each robot was associated with a separate laptop. The laptop was running the Sphinx speech recognition system, identical to our earlier implementations as described in (Chapter 4). The robot and the speech recognition system would communicate via Bluetooth. Each system was accompanied by a Snowflake microphone which was also used for our earlier ROILA implementations as described in (Chapter 4). The robots could also talk back in ROILA. For text to speech we used the Festival system. For details please refer to Chapter 4.

## 6.5 ROILA Lessons

Each lesson was conducted once every week for each class and lasted for 100 minutes. The language of instruction during the lessons was English. The science teachers were always present in class to help any of the students with understanding English and if necessary translating the sentence in question to Dutch. We will now succinctly describe the 3 ROILA lessons.

### 6.5.1 Lesson 1

The first lesson was an introductory lesson, where the students were taught briefly about robots and were given a quick introduction to ROILA. Pronunciation rules were explained and some basic words were learnt as part of the first vocabulary (see Table 6.1). The word structure of ROILA was also introduced. In the second part of the lesson the students were given some ROILA commands that they could use to instruct the robot to navigate in its environment (see Figure 6.1) in an open scenario. This meant that the students could immediately see a direct application of their efforts of learning the language. In reply, the robot would say phrases like: error (when it could not understand what was said), wrong way (when it bumped into a wall), etc.



Figure 6.1: Robot used in Lesson 1

Vocabulary List	Commands to be used by the students	What the robot can say
Start - bofute	Start/Go - bofute/kanek	wopa - ok, good
go - kanek	Go forward - kanek koloke	wekapa - error
turn - botama	Go backward - kanek nole	wolej nawe - other way (when the robot bumps into something)
forward	- Go/Turn left - kanek /botama	
koloke	webufo	
left - webufo	Go/Turn right - kanek /botama	
	besati	
right - besati	Turn back - botama nole	
backwards	- Run - bobuja	
nole		
quickly - jimeja	Go quickly - kanek jimeja	
run - bobuja	Go slowly - kanek kipupi	
slowly - kipupi	Stop (no go) - Buse kanek	
no, not - buse		
wopa - ok/good		
wekapa - er-		
ror, wolej nawe		
(robot)		

Table 6.1: Vocabulary employed in Lesson 1

### 6.5.2 Lesson 2

The second follow up lesson was slightly more detailed and it discussed the ROILA grammar. The main grammar elements were highlighted and examples were given. Students were taught how they could make simple sentences in ROILA (see Table 6.2). Once again in the second part of the lesson, the students were given an opportunity to bring into practice what they learnt during that lesson by playing a simple shooting game with a NXT robot (see Figure 6.2). The game required the students to place colored balls in front of a color sensor and if they uttered the correct color the ball would drop into a firing chamber and be fired. The robot would say phrases like: wrong color, wrong ball, OK, etc. The students could also use commands from the first lesson. The robot used in the second lesson was a simple extension of the robot used in the first lesson.

### 6.5.3 Lesson 3

The last lesson was much more theoretical in nature and various ROILA grammar exercises were given and practiced. This was meant as a revision for the ROILA proficiency test which would be administered in the second part of the lesson.

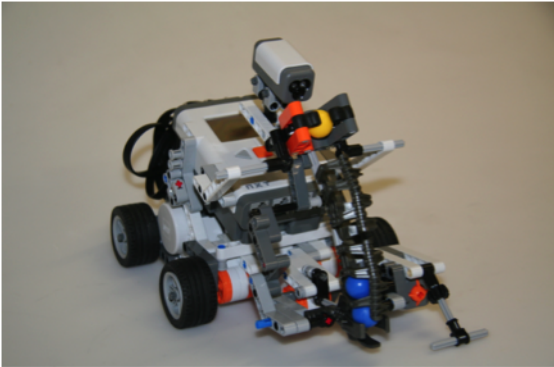


Figure 6.2: Robot used in Lesson 2

Vocabulary list	Commands to be used by the student	What the robot can say
jinolu - ball	I put (color) ball - pito	put (color) ball - to-bop (tifeke/kutuju/koleti/wipoba) jinolu
zero - pojos one - kilu	tobop(tifeke/kutuju/koleti/wipoba) jinolu	What - biwu (if robot does not understand what was said)
I - pito	(color)ball - (tifeke/kutuju/koleti/wipoba)jinolu	What color ball? - biwu wekepo jinolu
put - tobop		Wrong color - bemeko wekepo
color - wekepo	Other commands: from lesson 1	Wrong ball - bemeko jinolu
red - tifeke blue - kutuju green - koleti yellow - wipoba wrong - bemeko biwu - what		Everything ok - wopa

Table 6.2: Vocabulary employed in Lesson 2

## 6.6 Homework, the Lesson Booklet and the ROILA exam

In order to support in class learning the students were requested to carry out homework tasks. To support this we used the interactive e-learning tool Moodle (Moodle, 2010). 7 homework reading exercises were placed in Moodle, with the first four expected to be completed after the first in class ROILA lesson and the last three after the second lesson. The 7 homework exercises extended what was learnt in class and therefore provided the children with reference/-textbook material. The total vocabulary list that was covered in the curriculum comprised about 50 ROILA words (about 20 learnt in class, rest via homework).

The e-learning homework lessons were designed each to be completed within 20 minutes. The lessons presented a short vocabulary list of 6 to 10 words, reviewing vocabulary words presented in class as well as new words to be repeated in coming classes. The ROILA language features in the lessons were limited to the commands and interactions students needed for working (playing) with the LEGO robots. Practice exercises in the homework lessons emulated the classroom robot interaction game design, to make students comfortable with the process and provide incentive to study. Each lesson had a translation section that students answered in their lesson booklets, providing for both practice and the assembly of personal review materials. The homework lessons can be found in the Appendix.

A lesson booklet was also provided to the students where they could write down notes from both the lessons in class and also from the homework. The motivation of using such a booklet was as a self report diary, i.e. we wanted the students to record how much time it took them to complete the homework. Therefore specific space was provided to write down how much it would take them to do the homework. Before the start of the second and third lesson, we checked the diaries of as many students as possible to see if the self report was carried out. At the end of the third lesson the booklets were collected from the students.

The ROILA exam was also conducted in Moodle and it was conducted in the second part of the third ROILA lesson. Total time spent learning ROILA in relation to the exam scores would prove to be an essential variable in determining the learnability of ROILA. Since the ROILA lessons were part of the Science module of the students, the exam also carried weight towards their Science grade. This automatically meant that we could to some extent control the motivation factor of the students. Under normal circumstances, we would expect the students to be less motivated to learn a language. The exam was designed using the theoretical background as described in one of the earlier chapters. The same principles were applied as we did when we designed proficiency tests for other artificial languages such as Toki Pona and Klingon as described in Chapter 5. The exam was administered in one of the computer labs at the school and it was a closed book/closed notes type exam. The exam consisted of various questions that tested both vocabulary and grammar of ROILA. Sample questions were: vocabulary matching questions, completing ROILA sentences with the correct ROILA word and translating English sentences to ROILA. The

final exam for the course was based on the commands and instruction set learned in the seven homework lessons. Exam questions required students to manipulate language components to create new commands on the models they had learned, and to discern acceptable (correct) command structures for interaction with the robots. The exam was scored automatically by Moodle and the grades were scaled between a range of 1-10 (maximum = 10).

## **6.7 Discussion: ROILA Curriculum**

During the three lessons we observed the students and also talked to several of them to inquire about their learning experience. The VWO students as expected found the language relatively easier to learn as compared to the HAVO students. In addition, the VWO students also appreciated interacting with the robots, which was unexpected to us. This was primarily because VWO students are much more comfortable with theoretical tasks. In general all students enjoyed the practical part of the lessons the most, i.e. when they could talk to and play with LEGO robots. We observed a strong tendency amongst the students to assign their own meanings to ROILA words, i.e. they wanted to give ROILA words meanings of their own choice. Naturally some students were observed to be more enthusiastic than others. One of the bigger challenges that we faced during the lessons was in the practical session, when the children had to interact with the robots. Since there were at least 25 children in every class it meant that there was quite a racket when they were asked to play with the robots. Ultimately this affected the speech recognition and at times the children had to repeat their utterances. Based on qualitative feedback from the students we observed a general enthusiasm amongst the children of not only interacting with robots but also using a secret language to interact with them. Certain students also created their own customized robots and scenarios, for e.g. set up obstacles in a maze like fashion and make the robots navigate around them by giving ROILA commands. In summary, we (researchers, teachers, students) thought the practical part of the lessons was a fun activity because it was open and allowed for experimentation and creativity (see Figure 6.3).

## **6.8 Controlled Experiment comparing ROILA and English**

Now that we had imparted some prior knowledge of ROILA to the students of Huygens College Eindhoven we conducted a controlled experiment that would not only empirically evaluate the recognition accuracy of ROILA against English but also attempt to deduce the learnability of ROILA as a language. Moreover we also looked at the subjective impressions users would have after interacting in ROILA. The experiment was conducted in the week after the end of the ROILA lessons (4th week of the curriculum). In the experiment, the children were requested to play a simple game with a LEGO robot in both ROILA and in English.





Figure 6.3: Students interacting with the robots during the ROILA lessons

### 6.8.1 Research Questions

We had the following research questions that we aimed to answer using the results of the controlled experiment.

1. Do subjects perceive the use of ROILA to be easier than the use of English when talking to a speech system?
2. Is the speech recognition accuracy higher when subjects speak ROILA in comparison to when they speak English? If yes, by how much?
3. Under the assumption that ROILA has higher recognition accuracy, how long do the subjects have to interact with the speech system before their initial investment of learning ROILA pays off?

### 6.8.2 Participants

From the total group of about 100 students who took part in the ROILA curriculum (97 gave the ROILA exam), a selection of students ( $N = 35$ ) were invited for the experiment and all of them agreed to participate voluntarily. The selection was based on the following factors: enthusiasm shown in class, recommendations of the 3 science teachers, completion of homework and a fair mix of boys and girls. In summary, we wished to involve students who had taken somewhat of an active part in the entire learning process. At this juncture, we would like to discuss our selection of participants in slightly more detail. Obviously the ideal case would have been to include the entire pool of about 100 ROILA students. This was difficult firstly due to practical reasons. A random selection would have been the other way to go but this would bring with the main issue that we could end up with students who would not have taken part in the entire learning process. This would have been due to various reasons: absent from class, did not complete homework, etc. All of this would mean that such

students would not have basic proficiency in ROILA and this could potentially cloud the recognition accuracy comparison as these students already knew English. Therefore we used various criteria to determine which student would be used in the control experiment. Note that the score on the ROILA exam was not a factor and played no part in the selection process. Enthusiasm was mainly defined as interest shown in interacting with LEGO robots and not in ROILA. Our belief was that if a student was excited by the prospect of interacting with LEGO robots he or she would feel that way regardless of the language used. Nevertheless our approach is not fool proof and we acknowledge that our selection of participants would still be biased. To quantitatively determine if a selection bias was existing, we executed a simple between subjects ANOVA, where the between subject factor was whether a particular student was selected for the controlled experiment. The dependent variable was the ROILA exam score for that student. The ANOVA results were as follows:  $F(1, 95) = 2.8$ ,  $p = 0.10$ . The means are summarized in Table 6.3. It is evident that our selection of students for the controlled experiment performed better on the ROILA exam but this difference was not significant.

Selection	Mean Exam Score	Std. Dev
Selected students	6.87	1.73
Not selected students	6.24	1.79

Table 6.3: ROILA Exam Means table for selected and non selected students

### 6.8.3 Experiment Design and Measurements

We conducted a within participant experiment in which the within factor was language (ROILA or English) and the measurements were SASSI scores, number of commands, semantic accuracy, sentence accuracy, word accuracy. The SASSI scores were measured through various factors of the SASSI questionnaire (Hone & Graham, 2001) in the form of a self-report. The three accuracy measures and the number of commands were observed with the help of video recordings of the experiment sessions. In addition, we recorded several measurements to control for possible biases. These biases could be due to the characteristics of the participant (gender, class group), the experimental procedure (experiment order, number of days between last ROILA lesson and experiment) or the general game performance across ROILA and English. We will now define each of the measurements.

#### 6.8.3.1 SASSI Score

The Subjective Assessment of Speech System Interfaces (SASSI) questionnaire is a standardized questionnaire used in Speech Interaction analysis. It covers the important technical facets involved in speech interaction and also addresses the issues related to speech recognition accuracy as perceived by the participants. For example, did the participants think the system was accurate, was the response quick enough, did they like the interaction? The internal reliability of the questionnaire has been found to be high ( $> 0.7$ ) by its creators.

The standard version of the questionnaire comprises 34 items which are then merged across the following 6 factors described below for easy comprehension.

1. System Response Accuracy: Did the system recognize the user input correctly and hence did what the user intended and expected?
2. Likeability: Was the system appreciated and did the system induce some positive affect in the user?
3. Cognitive Demand: An indication of the perceived level of mental effort required to use the system and the feelings arising from this effort.
4. Habitability: The items in this factor deal with aspects such as, does the user know what to say and what the system is doing. High habitability would mean that there is a good match between the user's conceptual model of the system and the actual system.
5. Annoyance: General irritation and frustration pertaining to interaction with the system.
6. Speed: Performance of the system in terms of time taken to carry out a response.

Participants in our experiment were exposed to the SASSI questionnaire for every condition, i.e. separately for both ROILA and English. Items from the questionnaire were presented on 5 point Likert scales, where a 1 meant completely disagree and a 5 meant a completely agreed upon rating. Items were randomized so that the scores for every factor were not biased. Moreover, a Dutch translated form of the questionnaire was used so that comprehension would be easy for participants. The translated form of the questionnaire was made available from (Turnhout, 2007).

### **6.8.3.2 Number of Commands**

This measurement equalled the number of clean commands uttered by the participant that were considered in the analysis. During transcription of the interaction dialog of every participant we excluded commands which we thought were not related to the game context, for e.g. participant started talking to the facilitator, or in Dutch, etc. Since the microphone was continuously on, the system would mistakenly think something was said to it.

### **6.8.3.3 Semantic Accuracy**

Semantic accuracy measures if the system could understand what was said and extract the propositional context despite not having 100% recognition accuracy. For example in our context did the robot still carry out the correct action/response even if recognition was not 100% accurate? For example if the participant would say Turn Left in English, and if the system would recognize the utterance as Left it should still enable the robot to go left. This variable is also stated as concept accuracy in literature (Boros et al., 1996).

#### **6.8.3.4 Sentence Accuracy**

Sentence Accuracy states how many sentences were recognized 100% accurately.

#### **6.8.3.5 Word Accuracy**

Word accuracy is a standard accuracy for speech recognition (Boros et al., 1996) based on Levensthein distance (Gilleland, 2002). For the sake of comprehension, the reverse of word accuracy is also known as word error rate (WER), another commonly used measure in Speech Recognition.

$$\text{Word Accuracy}(\%) = 100 - \text{WER}(\%)$$

Where WER is the number of operations required to convert the recognized sentence to the reference sentence, or the sentence that was said by the user. This computation is in fact the Levenshtein distance. The operations can be of three types, namely: insertions, deletions or substitutions and they all have the same cost. We state the following equation to compute WER, where S = substitutions, D = deletions, I = insertions and N = number of words in the reference sentence.

$$\text{WER} = \frac{(S + D + I)}{N}$$

Word accuracy computation is based on Levensthein distance of two strings on a word level (the two strings being: what was said and what was recognized). Normally Levensthein distance operates on a character (byte) level but when it comes to Word accuracy for Speech Recognition the distance must be computed at a word level. This can be simply accomplished by recoding every unique word in the two sentences to a unique digit (1 byte) and then calculating the Levensthein distance on a character level.

#### **6.8.3.6 Gender**

In our experiment we employed both male and female participants.

#### **6.8.3.7 Class group**

The children invited to the experiment either belonged to the HAVO or VWO category as described earlier in this chapter.

#### **6.8.3.8 Experiment Order**

Participants in our experiment either interacted in ROILA first or in English. As stated earlier, they interacted in both languages.

#### **6.8.3.9 Number of days between last ROILA Lesson and day of Experiment**

The children were invited to the experiment in the fourth week of the curriculum, so there was a certain gap between their last ROILA lesson which was

in the third week and the day when they took part in the experiment. Even though the gap between the third lesson and the day of the experiment was particularly related to the ROILA condition, yet we believed that it could have affected the general performance for both ROILA and English due to a similar interaction context. For example it could be the longer the gap the more difficult the participants find to interact with the robots.

#### **6.8.4 Procedure**

All students were provided with a game handout by email a few days before the experiment session. This was done to save initial startup time and so that the students did not have several questions about the game or otherwise. We also requested the children to learn all the required vocabulary and commands for the game but later on we decided to provide all participants during the experiment with a game handout to ensure consistency. Every student played the game alone in a quiet room with little or no ambient sound. The students were first seated in the room and were explained the game rules by the facilitator. If they had no questions the game would start. As mentioned prior, the students were also provided with a help sheet which enlisted the commands that they could use for both ROILA and English. The order of playing the game in either English or ROILA first was balanced. Each game lasted for 10 minutes per language. At the end of every game the students would fill in the SASSI questionnaire related to their interaction experience with the language in question. The output of the recognition system was recorded as both log files (on disk) and as video recorded screen shots. This would help in transcribing the interaction dialogue for later on. Since we were recording video and also capturing audio we could code what said by a participant against what was recognized by the system. The three students who performed the best while playing in ROILA were awarded with souvenir gifts of the TU/e.

#### **6.8.5 Setup**

The software and the interaction required for the game were simple extensions from the class lessons and besides the game rules and the interaction vocabulary there were not much differences in context. The laptop was running the Sphinx speech recognition system, identical to our earlier implementations. The robot and the speech recognition system would communicate via Bluetooth. Each system was accompanied by a Snowflake microphone. The robots could also talk back in ROILA. The only change from the earlier described implementations as in Chapter 5 was that we also employed a dialog management strategy to control the flow of the game. The individual dictionaries used for the recognition of ROILA and English only comprised of the words that would be used in the game. The participants were seated in one of the corners of the room, from where they could easily see the robot and the playing space. The video camera was placed behind their right shoulder from where it would record the system output (see Figure 6.4).



Figure 6.4: Participant setup

### 6.8.6 Game Design

The game that was adopted as the interaction context for the experiment was an extension of the scenarios used in the first two ROILA lessons (same system). In addition, the vocabulary that was supposed to be used in the game came directly from the ROILA curriculum employed during the 3 weeks. The objective of the game was to put as many balls as possible in 4 colored goals (Red, Green, Blue, and Yellow), which were spread in a room (see Figure 6.5). The LEGO robot was placed at a start point and it would then choose one of the four possible colors. The sequence of colors was same for both ROILA and English for one participant. The robot would declare the color aloud, for e.g. toward red in English or kufim tifeke in ROILA. This meant that the student had to make the robot move to the red circle by giving navigation commands. If at any time during the game they were not sure or forgot which color they had to look for they could ask the robot, what color in English or biwu wekepo in ROILA. Once the robot had stopped on a circle (such that the color sensor of the robot was directly about the colored circle), the student had to give a command to the robot to sense the color, for e.g. see red in English or make tifeke in ROILA. If it was a wrong circle the robot would say wrong color in English or bemeko wekepo in ROILA. If the color recognition would go ok, the robot would say tobop jinolu or put ball and the children were allowed to shoot the ball in the goal, but they had to orient the robot towards the goal; so that there would be a chance of scoring. Once the shoot command was given, i.e. drop ball in English and jasupa jinolu in ROILA the ball was fired towards the goal. Subsequently, the robot would again choose a new color and the same process was repeated till the 10 minutes were over. The list of commands is shown in

the table (see Table 6.4). The game flow is also represented as a state diagram (see Figure 6.6).

Commands that could be used in the game		What the robot could say	
ROILA	English	ROILA	English
kanek koloke	Go forward	kufim tifeke	Toward red
kanek nole	Go backward	kufim kutuju	Toward blue
botama webufo	Turn left	kufim koleti	Toward green
botama besati	Turn right	kufim wipoba	Toward yellow
botama nole	Turn back	wopa	Good
bobuja	Run	wekepa	Error
kanek jimeja	Go quickly	bemeko wekepo	Wrong color
kanek kipupi	Go slowly	tobop jinolu	Put ball
buse kanek	Stop		
biwu wekepo	What color		
jasupa jinolu	Drop ball		
make tifeke	See red		
make kutuju	See blue		
make koleti	See green		
make wipoba	See yellow		

Table 6.4: Commands that could be used in the game

### 6.8.7 Results

In the analysis of our results, we first attempted to determine if any external biases were playing a role and once we had ascertained that we moved on to the analysis of the main effects. We attempt to follow this pattern in every category within the results section. Every pre-test was run as a mixed design ANOVA with language type as the within subject factor and the external biases as the between subject factors.

As mentioned earlier our total pool of children invited to the experiment was 35. However we dropped 4 participants from being considered in the analysis due to mainly technical breakdown in the experimental setup because of which the game session did not last for the entire 10 minutes. This happened because of a bug in our game software.

### 6.8.8 Game Performance

A possible source of a bias could stem from the success in the game itself. Participants that scored higher in the game could appreciate the speech system more compared to participants that scored lower. We therefore analyzed the game performance of the participants. We performed a paired sample t-test for the 31 participants. As indicated in Table 6.5 participants shot more balls and scored more goals in the ROILA condition, but this difference was not significant. Therefore we could assume that success in the game would not be a limiting factor for further analysis. However later on in the chapter while

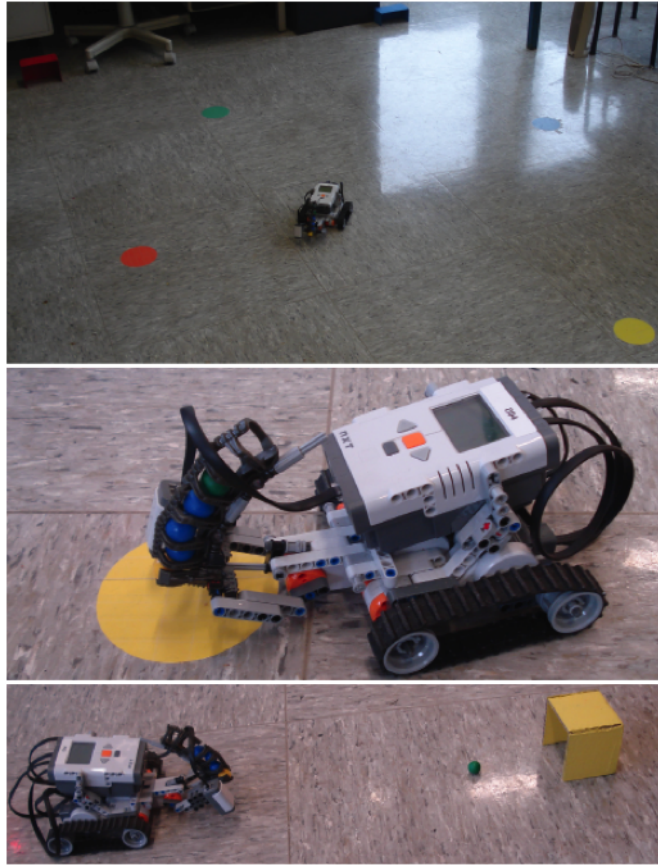


Figure 6.5: Game setup

presenting more advanced results we will get back to the possible bias of game performance and attempt to evaluate its effect.

Game Performance Measure	Language type		ROILA		English	
	t (30)	p	Mean	Std.dev	Mean	Std.dev
Balls Shot	1.06	0.30	1.00	1.01	0.79	0.92
Goals Scored	1.15	0.26	0.37	0.66	0.21	0.48

Table 6.5: T-Test result and means table for balls shot and goals scored

## 6.8.9 SASSI Score Analysis and Results

### 6.8.9.1 Coding the SASSI ratings

It is worth noting that the SASSI questionnaire is administered in such a way that some of the items within a factor have a negative weight. This ensures that



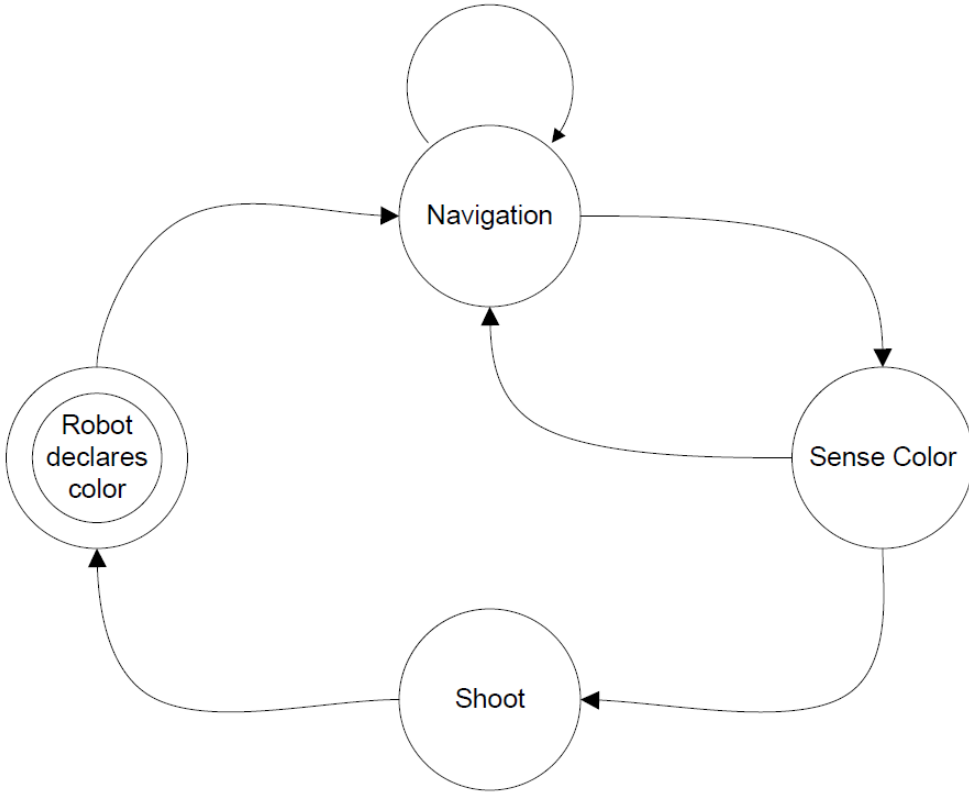


Figure 6.6: State Diagram of the game

the participants rank the items after thought and not just blindly or as put by (Hone & Graham, 2001) “to prevent respondents being tempted to simply mark straight down a column of responses”. Therefore while coding the rankings such items have to be reversed to ensure consistency. Final ratings for each of the 6 factors mentioned earlier are computed by averaging the corresponding items within that factor. Consequently after reversal of ratings and averaging we can state that the higher the rating for a factor, the more positive it was perceived to be. The analysis of the SASSI scores comprised of 31 participants, where each participant interacted with the NXT robot in both ROILA and in English for the complete 10 minutes. 17 participants interacted in ROILA first.

#### 6.8.9.2 Reliability Analysis

Presented first is the reliability analysis for the SASSI questionnaire. The Cronbach alphas do not raise any eyebrows and for all 6 factors are above 0.7. The alpha values are reported combined for both ROILA and English (see Table 6.6).

System Response Accuracy	0.80
Likability	0.85
Cognitive Demand	0.75
Annoyance	0.83
Habitability	0.71
Speed	0.77

Table 6.6: Cronbach Alphas for the 6 Factors

### 6.8.9.3 Assessment of possible biases

We executed a pre-test in the form of repeated measures ANOVA to determine if the characteristics (gender, class group) of the participants or the experimental order had an effect on any of the main measurements. As indicated in the results (see Table 6.7 and 6.8), we did not find any significant effects for either of two factors. Therefore we could exclude them from the analysis of the SASSI questionnaire. The second ANOVA was run to determine if experiment order had an effect on the measurements.

R=ROILA E=English	Class Group							
	VWO				HAVO			
	Female		Male		Female		Male	
Measurements	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev
R System Accuracy	2.59	0.45	2.77	0.45	2.53	0.58	2.96	0.77
E System Accuracy	2.26	0.32	2.65	0.78	2.39	0.71	2.42	0.43
R Likeability	3.06	0.53	3.48	0.23	3.27	0.47	3.48	0.49
E Likeability	3.07	0.28	3.18	0.85	3.15	0.68	3.00	0.80
R Cognitive Demand	2.47	0.61	3.40	0.59	2.88	0.55	3.42	0.56
E Cognitive Demand	2.30	0.10	3.33	0.37	3.20	0.97	3.08	0.92
R Annoyance	3.00	0.72	3.05	0.87	3.15	0.76	3.44	0.81
E Annoyance	2.97	0.87	2.80	0.28	3.10	0.66	2.62	0.63
R Habitability	3.08	0.76	2.38	0.72	3.27	1.04	3.15	0.58
E Habitability	2.92	0.63	2.77	0.73	3.27	0.93	2.68	0.66
R Speed	3.67	1.15	3.50	1.22	3.58	0.95	3.75	0.75
E Speed	3.00	1.32	3.58	1.20	3.38	0.91	2.95	0.83

Table 6.7: Means table for SASSI ratings across gender and class group

Factor Name	Gender		Class Group	
	F(1,28)	p	F(1,28)	p
System Response Accuracy	1.84	0.19	0.001	0.99
Likeability	0.31	0.58	0.001	0.99
Cognitive Demand	3.39	0.08	0.58	0.46
Annoyance	0.16	0.69	0.26	0.61
Habitability	2.54	0.12	1.45	0.24
Speed	0.01	0.91	0.03	0.87

Table 6.8: ANOVA table for SASSI ratings across gender and class group

The results revealed that experiment order had a significant effect on two subjective ratings, i.e. System Response Accuracy and Speed (see Table 6.9 and Table 6.10 for mean, standard deviation and ANOVA results). However,

R=ROILA E=English Acc=Accuracy	Order of experiment			
	English first		ROILA first	
Measurements	Mean	Std.dev	Mean	Std.dev
R System Accuracy	2.99	0.62	2.50	0.54
E System Accuracy	2.60	0.49	2.31	0.67
R Likeability	3.52	0.40	3.23	0.45
E Likeability	3.12	0.54	3.08	0.83
R Cognitive Demand	3.40	0.48	2.88	0.66
E Cognitive Demand	3.26	0.61	2.97	0.99
R Annoyance	3.19	0.61	3.22	0.91
E Annoyance	2.59	0.54	3.11	0.59
R Habitability	2.98	0.83	3.09	0.91
E Habitability	2.63	0.75	3.21	0.75
R Speed	3.32	1.05	3.88	0.74
E Speed	2.71	0.99	3.68	0.71

Table 6.9: Means table for SASSI ratings across experiment order

Factor Name	Experiment order	
	F(1,28)	p
System Response Accuracy	5.56	*0.03
Likeability	1.08	0.31
Cognitive Demand	3.34	0.08
Annoyance	1.94	0.17
Habitability	2.00	0.17
Speed	6.83	*0.01

Table 6.10: ANOVA table for SASSI ratings across experiment order

given the small effect size and the fact that we counterbalanced the experiment order, we assume that this possible bias does not invalidate the further analysis of the SASSI data.

#### 6.8.9.4 Main effect analysis

To determine if the language condition was having an effect on the subjective SASSI ratings of the children a repeated measure ANOVA was conducted. The within factor was language (English or ROILA) and the measurements were the six factors from the SASSI questionnaire: system response accuracy, likeability, cognitive demand, annoyance, habitability and speed. Table 6.11 shows the mean, standard deviation, F value, and p value for each measurement. The results are visualized in Figure 6.7. ROILA was evaluated as better on all six factors of the SASSI questionnaire. Three factors, namely System Response Accuracy, Annoyance and Speed were all significant in favor of ROILA. Likeability was touching significance.

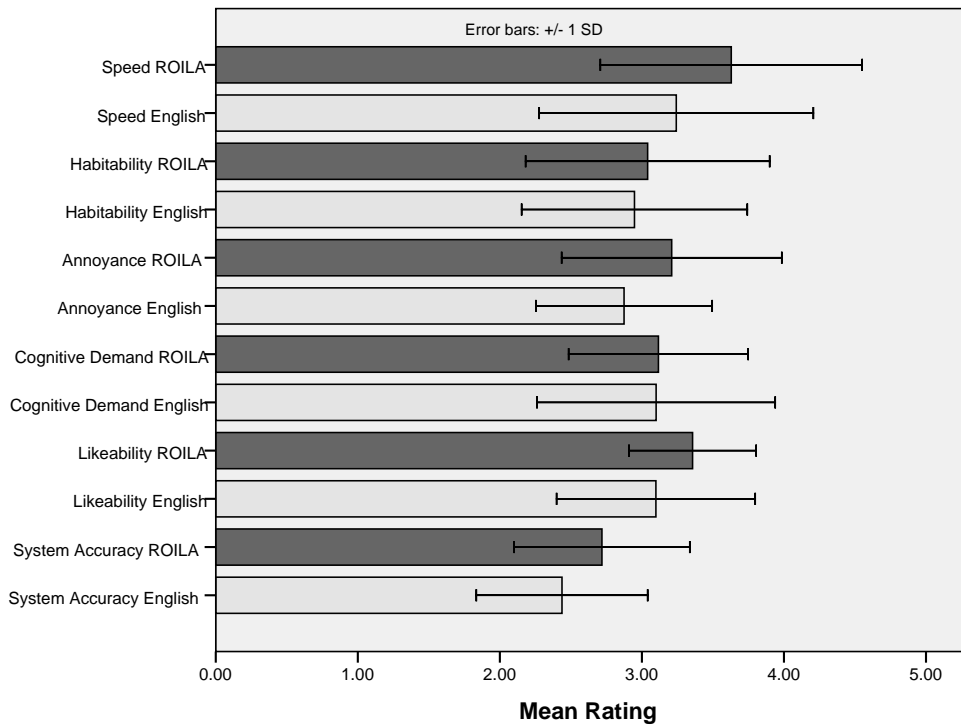


Figure 6.7: SASSI mean ratings bar chart

Factor name	Language type		ROILA		English	
	F(1,30)	p	Mean	Std.dev	Mean	Std.dev
System Response Accuracy	4.88	*0.04	2.72	0.62	2.44	0.60
Likeability	3.56	0.07	3.36	0.45	3.10	0.70
Cognitive Demand	0.01	0.91	3.12	0.63	3.10	0.84
Annoyance	4.94	*0.03	3.21	0.77	2.87	0.62
Habitability	0.29	0.59	3.04	0.86	2.95	0.79
Speed	10.44	*0.003	3.63	0.92	3.24	0.96

Table 6.11: ANOVA and Mean-Std.dev table for SASSI main effects

Earlier in the section we discussed the variable: game performance (balls shot and goals scored). Initially we concluded that it did not seem to be having a bias. For reassurance, we repeated the main effects analysis ANOVA but now with the game performance variables as 4 covariates. The new ANCOVA did not change our main effect results, so we still achieved significant trends for System Response Accuracy, Annoyance and Speed (see Table 6.12).

Factor name	F(1,26)	p
System Response Accuracy	4.52	*0.04
Likeability	3.60	0.06
Cognitive Demand	0.18	0.68
Annoyance	4.35	*0.05
Habitability	0.13	0.72
Speed	9.99	*0.004

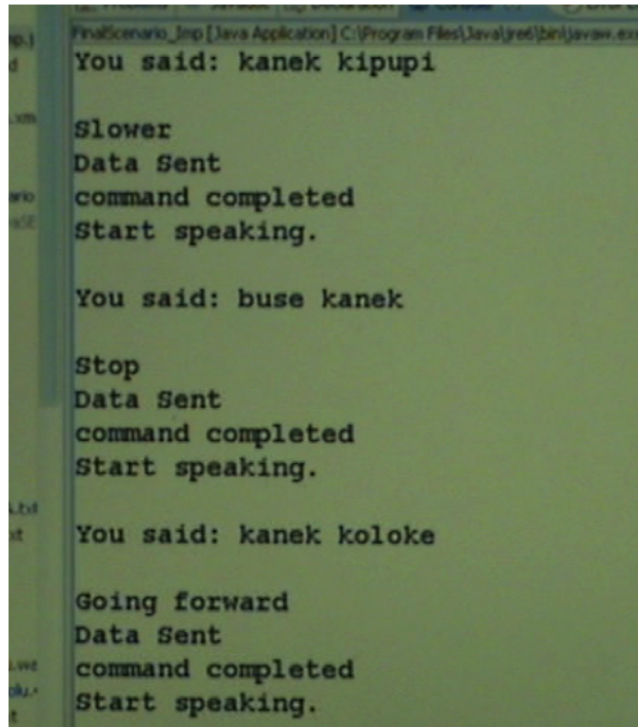
Table 6.12: ANCOVA table for SASSI main effects after including game performance as a covariate

### 6.8.10 Recognition Accuracy

#### 6.8.10.1 Data preparation

The interaction videos from the experiment were coded to transcribe the dialogue of every participant. Essentially for every participant we required a log of what was said and what was recognized by the system for both English and ROILA. We coded a pool of 24 participants from the total sample of 35 participants. 7 participants had to be dropped due to discrepancies in the videos or experimental setup. For example for some of the participants we did not have the entire 10 minutes of game play on video and for other participants the system microphone and the external microphone were both on, consequently meddling with the speech recognition on a subtle and almost unperceivable level. Four more participants were excluded who were the same as the participants also ignored in the SASSI analysis. They were dropped as they did not complete the experiment for 10 minutes. During the coding some dialogue from the participant was not considered. This was due to one of the following various reasons: participant started speaking in Dutch, participant started talking to facilitator, or participant started to speak in ROILA during the English condition and vice versa. As the microphone was always on, the system would react as if something was said to it in the language being tested (ROILA or English). Utterances by the participant were coded into a log file which already had the list of sentences recognized by the system. The videos were zoomed into the system output (see Figure 6.8 for an example), so that synchronization between audio and video was not an issue.

Ultimately the goal of the recognition accuracy analysis was to compute the recognition accuracy of ROILA, compared against English. To recall, we measured (observed) the following four dependent variables via the coding process: number of commands considered in the analysis, semantic accuracy, sentence accuracy and word accuracy. The number of clean commands was an absolute measure and the other three dependent variables were measured as percentages. Semantic and sentence accuracies were averaged across the total number of commands and word accuracy was averaged across the total number of words comprised in the total number of commands.



```

Practicalario_Jmp [Java Application] C:\Program Files\Java\jre6\bin\javaw.exe
You said: kanek kipupi

Slower
Data Sent
command completed
Start speaking.

You said: buse kanek

Stop
Data Sent
command completed
Start speaking.

You said: kanek koloke

Going forward
Data Sent
command completed
Start speaking.

```

Figure 6.8: Example of what the system output video looked like

#### 6.8.10.2 Assessment of possible biases

We conducted pre-tests to investigate the possible bias of either the characteristics of the participant (gender, class group) or the structure of the experiment (order of experiment, days between the exam and the experiment) on any of the measurements related to recognition accuracy.

First we investigated the possible bias from the characteristics of the participants (gender and class group). A repeated measures ANOVA with language type as the within subject factor and gender and class group as the between subject factors was run. The goal was to determine if gender and class group had an effect on the four recognition accuracy measurements that we described earlier in the measurements section. The results of this pre-tests show that the characteristics of the participants did not significantly influence the measurements (see Table 6.13 and Table 6.14).

A second repeated measures ANOVA was conducted to analyze the possible bias of the order of the conditions. The results revealed that experiment order was not influencing our recognition accuracy measurements. The results show the means and standard deviations across the order of the experiment (Table 6.15) as well as the ANOVA results (Table 6.16).

R=ROILA E=English Acc=Accuracy	Class Group							
	VWO				HAVO			
	Female		Male		Female		Male	
Measurements	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev
R Commands	68	4	83	14	63	14	79	15
E Commands	92	6	92	17	76	17	80	14
R Semantic Acc	82.6	5.2	72.3	11.5	60.0	12.3	71.9	9.9
E Semantic Acc	78.3	3.2	65.9	22.1	69.8	11.5	57.9	24.0
R Sentence Acc	63.9	20.6	55.7	14.0	45.7	13.6	59.0	13.2
E Sentence Acc	52.1	5.9	48.2	23.5	39.9	13.8	38.8	27.4
R Word Acc	71.7	16.1	67.8	11.4	57.8	15.5	63.5	17.5
E Word Acc	55.7	4.6	49.8	27.8	42.8	19.8	40.5	34.7

Table 6.13: Means table for recognition accuracy measurements across gender and class group

Dependent variable	Gender		Class Group	
	F(1,21)	p	F(1,21)	p
Number of Commands	2.94	0.10	2.24	0.15
Semantic Accuracy	0.33	0.57	1.89	0.18
Sentence Accuracy	0.17	0.68	0.99	0.33
Word Accuracy	0.0009	0.99	1.04	0.32

Table 6.14: ANOVA table for recognition accuracy measurements across gender and class group

R=ROILA E=English Acc=Accuracy	Order of experiment			
	English first		ROILA first	
Measurements	Mean	Std.dev	Mean	Std.dev
R Commands	78	12	71	19
E Commands	82	15	83	17
R Semantic Acc	72.4	12.5	66.4	11.8
E Semantic Acc	61.5	23.4	68.6	15.2
R Sentence Acc	60.1	14.6	49.4	12.7
E Sentence Acc	42.7	22.0	42.4	21.7
R Word Acc	67.3	16.8	59.8	12.9
E Word Acc	42.8	30.5	46.7	23.6

Table 6.15: Means table for recognition accuracy measurements across experiment order

As a last step in estimating the effect of any bias from the experimental setup, we wished to determine if the days between the third ROILA lesson and the day of the experiment had an effect on the recognition accuracy measurements. We conducted a regression analysis to answer this question. The average for the variable was 6.79 days with a standard deviation of 1.14 days. The summarized results of the regression model are presented in a table (see Table 6.17).

Dependent variable	Experiment order	
	F(1,22)	p
Number of Commands	0.27	0.61
Semantic Accuracy	0.01	0.92
Sentence Accuracy	0.79	0.39
Word Accuracy	0.05	0.82

Table 6.16: ANOVA table for recognition accuracy measurements across experiment order

Dependent variable	Regression results		
	t(22)	p	standardized $\beta$
Total ROILA Commands	1.91	0.07	0.38
Total English Commands	1.28	0.21	0.26
ROILA Semantic Accuracy	-0.70	0.49	-0.15
English Semantic Accuracy	0.54	0.60	0.11
ROILA Sentence Accuracy	0.21	0.84	0.04
English Sentence Accuracy	1.32	0.20	0.27
ROILA Word Accuracy	-0.57	0.58	-0.12
English Word Accuracy	0.38	0.71	0.08

Table 6.17: Results for regression model for days between last ROILA lesson and day of experiment and recognition accuracy measurements

The result of the regression analysis showed that the days between the third ROILA lesson and the day of the experiment did not significantly influence the recognition accuracy measurements. However the effect was nearing significance for Total ROILA commands. Upon analyzing the means, we see that the larger the gap between the last lesson and the day of the experiment the more commands were said by the user (see Table 6.18). Could this be because such students were less fluent in ROILA?

Number of days gap	Frequency	Total ROILA Commands	
		Mean	Std. dev.
3	1	74	-
6	8	66.75	12.09
7	8	70.5	16.24
8	7	87.71	13.38

Table 6.18: Means table relating total ROILA commands with number of days between 3rd lesson and day of experiment

### 6.8.10.3 Main effects analysis

We now performed repeated measures ANOVA to test our main effects. The results are summarized in the table and bar chart (see Table 6.19 and Figure 6.9



respectively). In the bar chart we do not include number of commands as it was an absolute measure and the other three measurements were reported as percentages.

Dependent variable	F(1,23)	p	ROILA		English	
			Mean	Std. dev	Mean	Std. dev
Number of Commands	9.13	*0.006	74.42	15.84	82.67	15.93
Semantic Accuracy (%)	1.03	0.32	69.41	12.26	65.07	19.61
Sentence Accuracy (%)	8.65	*0.007	54.72	14.47	42.55	21.36
Word Accuracy (%)	20.18	* < 0.001	63.58	15.14	44.74	26.76

Table 6.19: Means and ANOVA table for recognition accuracy analysis

The sentence accuracy and word accuracy in the ROILA condition was significantly above the accuracies in the English condition. There is no significant difference between the conditions with respect to the semantic accuracy but participants used significantly more commands in the English condition than in the ROILA condition.

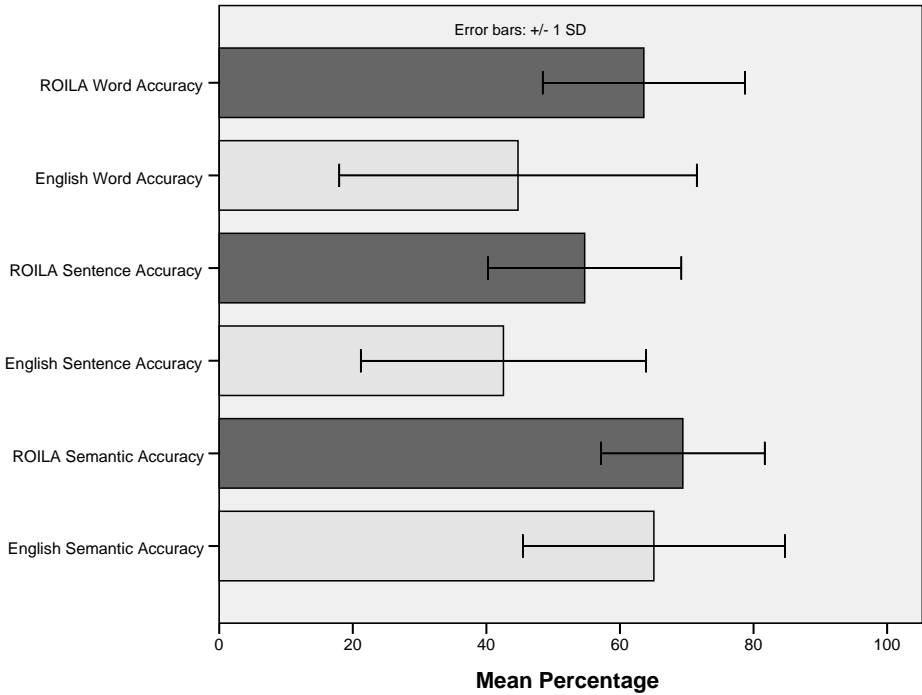


Figure 6.9: Bar chart indicating mean percentages for recognition accuracy measurements

As mentioned in the section of the SASSI scores main effect analysis we decided to carry out a last check to determine the effect of game performance by running an analysis of covariance (ANCOVA). Although one could argue that in comparison to the SASSI scores, game performance would not have as strong a bearing as on the recognition accuracy measurements. By including the 4 covariates our main effect trends did not change (see Table 6.20).

Measurement name	F(1,19)	p
Total Commands	8.39	*0.009
Semantic Accuracy	0.18	0.68
Sentence Accuracy	4.2	*0.05
Word Accuracy	14.36	*0.001

Table 6.20: ANCOVA table for recognition accuracy main effects after including game performance as a covariate

## 6.9 Evaluating the learnability of ROILA

In order to ascertain the learnability of ROILA we would have to analyze the performance of the students in the ROILA final exam.

### 6.9.1 Pre-test

Firstly, we wished to establish if the ROILA exam score was influenced by the characteristics of the participants. Therefore we conducted an between subjects ANOVA with gender and the class group as the independent variables and the ROILA exam score as the dependent variable. Both gender and class group were found to have a significant effect on the ROILA exam score,  $F(1, 21) = 5.53$ ,  $p = 0.03$  and  $F(1, 21) = 12.39$ ,  $p = 0.002$  respectively. The means and standard deviations are summarized in Table 6.21.

	Class Group							
	VWO				HAVO			
	Female		Male		Female		Male	
	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev
ROILA Final Exam Score	8.67	0.94	8.39	0.90	7.67	0.88	6.00	1.67

Table 6.21: ROILA Exam Score Means and Std.devs across Gender and Class group

Therefore it is clear to see that firstly, girls did better than the boys on the ROILA exam and that secondly, the VWO classes scored significantly higher than the HAVO classes. The second result was anticipated as VWO students are better at learning languages and they are already exposed to several languages in their curriculum.

### 6.9.2 Main effects analysis

To fully substantiate the learnability of ROILA and the general proficiency of each student we performed a linear regression analysis of the same 24 students who we considered in the recognition accuracy analysis. We looked into the variables related to the learnability of ROILA, i.e. time spent at home learning ROILA, exam score in ROILA (maximum possible = 10 points) and the average exam score across other languages (English, German, French and Dutch). The grades for other languages were provided by the school after protection of the identity of the students.

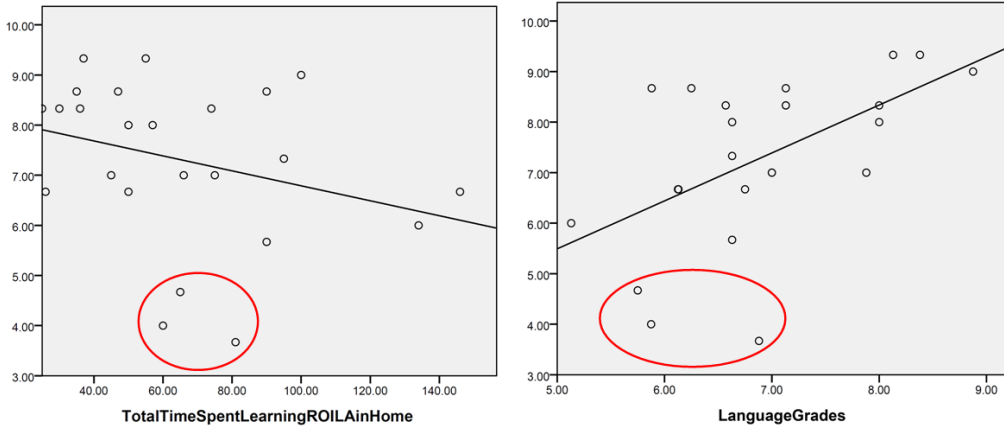


Figure 6.10: Scatter plots relating to ROILA exam scores

Illustrated (see Figure 6.10) are the initial scatter plots for the three variables in question, where the ROILA final exam score was the dependent variable. What we see firstly is that there exists a numerically inversely proportional relationship between time spent learning ROILA at home and the ROILA exam score, i.e. the more time a student spent the lower score he/she achieved. This could be explained perhaps by the general language proficiency of the student. However, note that this relationship is not statistically reliable. On average, the 24 students who we considered in the recognition accuracy part of our analysis took 65.4 minutes learning ROILA at home. Secondly, in the second plot we see a significant proportional relationship between language grades and the ROILA exam score.

We also observed the correlation data between the three variables in question by using the results from the multiple regression analysis. Using the enter method we did achieve a significant model  $F(2, 23) = 5.72$ ,  $p = 0.01$  but with a rather low  $R^2 = 0.35$  ( $R = 0.6$ ). Amongst the two coefficients language grades was significant,  $\beta = 0.52$ ,  $p = 0.008$  and time spent learning ROILA at home was not significant,  $\beta = -0.23$ ,  $p = 0.21$ . The model was not as strong as would expect initially. Therefore, it could be expected that certain outliers were clouding the results. Potential outliers are indicated in the red circles in the two graphs.

Upon further investigation there were found to be the same three participants for both graphs. We decided to repeat the regression analysis after excluding these three participants but the correlation results did not improve drastically. It should be acknowledged that after all the self report measure of time spent learning ROILA at home is a subjective measure and it may have been that the students carried out misreporting or erroneous judgments. In addition, language grades appear to be a much better predictor of the students performance in the ROILA exam as we found out that generally students who scored high in their grades for other languages also did well on their ROILA exam.

## **6.10 Formulating the Efficiency of ROILA**

In order to understand if it is worth learning ROILA, we would need to determine the value of every percentage of improvement in recognition accuracy. This would mean that for every gain in recognition accuracy how much is the gain in terms of interaction time? At the moment that we claim ROILA performed better than English in a controlled laboratory setting we could be faced with the critique that it comes with a cost, with an investment, an effort that must be put into learning the language. The children after all spent 3 weeks learning the language. Therefore we try to quantitatively explore if and when can it be worth learning ROILA. If a user is willing to invest some energy in the short term learning ROILA it will pay dividends in the long term, after prolonged consistent ROILA human robot or machine interaction. In order to carry out a cost analysis the first thing that we needed to know was how much does 1% recognition accuracy weigh in terms of time saved later on. Alternatively and in different words, how much does it cost if a user has to correct 1 speech recognition error in a dictation task of 100 words. Here we consulted the assistance of prior research (Munteanu, Baecker, Penn, Toms, & James, 2006).

Initially we discovered via such prior work that therein existed a linear relationship between an improvement in speech recognition accuracy and user performance. Though what we missed was a quantitative estimate. Other related research studies (VoCollect, 2010) and (Wald, Bell, Boulain, Doody, & Gerrard, 2007) provided quantitative measures which proved to be much more useful and valuable for us. Both studies discuss the time it would take for a user to correct a single speech recognition error (for e.g. in a dictation task for English) and using their results we can conclude that 3.5 seconds is a reasonable estimate. Note that this value of 3.5 seconds is an estimate, an assumption that would vary from system to system and from language to language. In an ideal case we would have liked to conduct a dictation experiment for ROILA as well, obviously not done due to practical constraints. From the results of our controlled experiment, we know that ROILA achieved 18.9% recognition improvement over English on the basis of word accuracy.

Therefore for a situation where a user has to say 100 words (not necessarily unique words) such as in a dictation task, we can expect that the user would save  $3.5 \times 18.9\% \times 100 = 66.2$  seconds ( $\simeq 1.1$  minute) for every task if he/she would interact in ROILA instead of English. Now let us assume that

the same user took 100 minutes to learn the necessary words in ROILA, required grammar and pronunciation rules. We also assume that he/she is already aware of those words in English and that after spending a certain time learning ROILA, the user is at an equal proficiency level in both ROILA and English, at least in terms of quality of pronunciation, as that is the main factor on which recognition depends on. We must also assume that the rate of speech (words per minute) for a user in both ROILA and English is the same, given that a user has underwent training in ROILA. Therefore after a total of  $100/1.1 \simeq 90$  dictation tasks the user would recover his/her time spent learning ROILA. From human behavior research we know that the normal rate of human speech is about 200 words per minute (Gould, Alfaro, Finn, Haupt, & Minuto, 1986). Therefore 90 ROILA dictation tasks of 100 words each would take  $(90 \times 100/200) + 90 \times (36.42 \times 3.5/60) = 236.21$  minutes, where 36.42 % is the average ROILA word error rate and 3.5/60 minutes is the cost incurred to correct one word error.

We now try to apply the accuracy relationship to our gaming scenario as employed in the controlled experiment. On average, the 24 students who we considered in the recognition accuracy part of our analysis took 65.4 minutes learning ROILA at home. At this point we include the time spent during class as that was identical for all students ( $3 \times 100 = 300$  minutes). The learning effort concentrated on the entire 50 words of the curriculum. The game required the students to understand 25 unique words, including both what they could say to the robot and what the robot could say to them. For the purposes of analyzing recognition accuracy we can exclude the words uttered by the robot, giving us a total of 20 unique words.

Although learning these twenty words might have taken less time, we continue our calculation with the more conservative estimation that to learn these 20 words it would have taken the participants the same total time it took to learn the entire curriculum ( $65.4 + 300 = 365.4$  minutes). In order to estimate how many words were said on average by a player in ROILA during the game we can use the average number of commands which are 74.2 (see Table 6.19) where each command has an average of 1.93 ROILA words (see Table 6.4). Therefore during the 10 minutes a player said on average  $74.2 \times 1.93 = 143.21$  clean words. For every 100 words we know that ROILA saves 1.1 minutes, so for 143.21 words a user would save 1.58 minutes. Therefore to recover their learning effort the children would have needed to play the game  $365.4/1.58 = 231.4 \simeq 232$  times in ROILA, where each game would last for 10 minutes. In other words, on average the children would have needed to play the game for 2320 minutes or 38 hours and 20 minutes before their learning effort would pay off.

### **6.11 Discussion of Results from the controlled experiment**

In this section, we would like to contemplate on our results and their implications. Firstly we would like to interpret our results in light of the research questions that we stated earlier in the chapter. Firstly, ROILA was indeed perceived to be as more user-friendly than English based on the subjective rat-

ings from the SASSI questionnaire and we could conclude that ROILA did not cause any extra cognitive overload as compared to English. We found significant differences for three factors and another factor was touching significance. Generally, most participants accepted that ROILA performed better than English for them. Secondly, ROILA was found to perform significantly better than English when it came to recognition accuracy. Thirdly we were able to state a relationship pertaining to when it would become worth learning ROILA in a practical situation such as in a dictation task and we also extended the same relationship to our gaming scenario as employed in the controlled experiment. We concluded that the real benefits of ROILA lie in long term use.

There is an initial investment but with persistent use within speech interaction therein lies potential. For example in a futuristic best case scenario students spend a certain time in their curriculum learning ROILA but then spend several years of their practical life using it in speech interaction. We also established that in our gaming scenario the children would have needed to play the game much longer in ROILA as compared to a normal dictation task (an almost 10 fold difference), to recover their learning effort. Therefore we believe that the time required to cover the learning effort also depends on the number of words said in an interaction scenario, i.e. the context holds certain importance. In the game scenario the children said less words (they didnt have to as they would be waiting for the robot to complete its movement) so the time required to cover the learning effort went up. The time it takes before it pays off to learn ROILA of more than 38 hours (in the gaming scenario) does seem large at first sight. But so does the hours it takes to learn how to write with twelve fingers. But over the years of usage, it certainly pays off. We would expect the same to hold true for ROILA. Upon more frequent use in everyday life, it is likely that learning ROILA will pay off.

There were a number of other interesting observations throughout the evaluation process. The recognition accuracy results were heavily in favor of ROILA, more so than the SASSI results. We have a plausible explanation for this. If we look closely at the recognition accuracy results, we see that the semantic accuracy or concept accuracy for ROILA and English does not have a significant difference. Therefore, more times than not, the robot would execute the same behavior for a particular uttered command for both languages, i.e. the behavior would not be drastically different. Obviously the participants would not be concerned with measurements such as word accuracy, as they were not directly affecting the behavior of the robot.

The number of commands is an interesting measurement and we found out that participants significantly said more commands in English than in ROILA. Could it be due to the fact they were more proficient in English and hence quicker in interaction? Or could it be due to the robot being less responsive to utterances in English because of low recognition accuracy and so the participants had to repeat their commands? We can only speculate about the causation of the trend. Significant differences for sentence accuracy and word accuracy indicate that at least on these levels ROILA performed much better than English.

In summary, we observed that the recognition accuracy of English was poor. It must be accepted that the participants were after all native Dutch speakers so English was a second language for them. Moreover they would speak English with an accent, some more than others. For example, pronouncing the “r” for students belonging to the south of Netherlands was difficult. The aspect of dialect comprised the speech recognition. Then again, such constraints would exist for any speech recognition system, unless a tailored acoustic model is used. In our situation we used an untrained American English acoustic model with the assumption that both ROILA and English would hence be on roughly the same footing. This could also explain the generally low recognition accuracies that we observed.

Besides the speech recognition friendly features of ROILA, another reason that could have contributed to its improved speech recognition was the average word length of the ROILA words used in the game played in the controlled experiment. The average length of the words for English was equal to 4.38 and for ROILA it was equal to 5.25. From our earlier results, as seen in Chapter 3 and from literature (Hämäläinen et al., 2005) it is known that the longer the word the better it is for recognition.

Closely monitoring the SASSI ratings leads to the conclusion that system response accuracy achieved negative ratings overall ( $< 3$ ,  $<$  neutral rating). Therefore we can rephrase the trend as that English was ranked much worse than ROILA. A cause for this trend could be that the students were mostly quite impatient while interacting with the robots and would want an immediate response without allowing for processing time. An important limitation of our result is that firstly we conducted the controlled experiment with a specific user group and with a limited number of participants. Secondly, only a subset of ROILA was evaluated. Therefore we cannot term our evaluation as large scale.

We would also like to contemplate the game scores observed in the controlled experiment (see Table 6.5). At first sight, the numbers indicate that general success level was low in the game, given that on average there was only 1 ball shot per game when the language was ROILA and even less in English. This is true and one reason is low recognition accuracy. However that is only one side of the story as during gameplay what we observed was that the children had trouble crossing the sensing color part of the game. As it may be recalled, this was when the children had to bring the color sensor of the robot ontop of the colored circles. To allow for a certain level of challenge the circles were not too big and therefore the children found it difficult to be accurate in their navigational instructions. Nevertheless, the low scores in the game cast a minor doubt over the applicability of this particular game scenario for ROILA.

To end on a positive note, we were extremely pleased with the success of the ROILA curriculum and the enthusiasm we encountered. The Huygens College appreciated our efforts and has inducted ROILA as a permanent part of their Science curriculum after certain revisions. As an illustration of this we attach an evaluation letter from them in the Appendix.

---

## Conclusion

---

As a conclusion to the research of ROILA we would like to summarize the results obtained, their implications and possible future avenues of research in the project.

The project began with a daunting goal of designing an artificial language which would be used to talk to robots and be easy to learn and easy to recognize. Initially we were faced with the plausible option of adopting a constrained language and modifying it to suit our needs. This would come with the added benefit of being easy to learn. For example using Basic English to talk to robots would fall in this domain. However, our interpretation was such that since Basic English already exists and if it would be efficient for speech recognition it would be used in speech applications. Moreover, if we would modify Basic English to make it speech recognition friendly we would already enter the category of artificial languages. After all, various artificial languages sound and look like other natural languages.

Throughout the design process, we always felt this tussle of maintaining a balance and keeping the trade-off to a minimum. Various design elements of ROILA were inherited from the languages overview that we carried out. The result of the overview was a set of design dimensions across phonology and grammar that were simply common trends found in a certain subset of artificial or natural languages. In the next stage of the project we carefully choose amongst these trends and incorporated some of them in the creation of ROILA. For example, we did not simply replicate the common phoneme list but rather rationally chose phonemes from and outside the list. A genetic algorithm was implemented which was the cornerstone of the ROILA vocabulary. The algorithm relies on a confusion matrix of phonemes as its fitness function and attempts to determine a vocabulary which is acoustically optimal. The strongest aspect of the algorithm is that it is scalable and hence practically any size of vocabulary can be generated. The grammar was designed with less automation and more thought. The vocabulary initially did not outperform English in terms of recognition accuracy but a second version of the vocabulary ultimately significantly outperformed English. We concluded that words having



Consonant-Vowel articulations are easier to articulate and hence easier to recognize for a machine.

A rational decision making technique was adopted to select grammatical markings for ROILA. Various criteria were involved which effected the choice of the markings with yet again ease of learnability and ease of recognition as the two main stakeholders. Our grammar evaluation did not reveal a significant effect in favor of ROILA and we speculated that this was due to lack of training or simply because we had a considerable variety of test subjects. The subjects did not have a common mother tongue and some of them had English as their first language.

Once we had the linguistic blocks in place we proceeded to implement and represent the language in the form of a prototype. LEGO Mindstorms NXT was our chosen test bed; the reasons of doing so have been elaborated earlier in the thesis. Sphinx-4 was our choice of speech recognizer, as it easily afforded the addition of ROILA to its setup. For the purposes of our evaluations we were also able to implement basic speech synthesis for ROILA using the Festival system.

The evaluation of ROILA as a whole followed multiple road paths. Initially we carried out a study where we established measurement tools to determine the learnability of artificial languages. We used Klingon and Toki Pona as the two test cases. These tools included proficiency tests which we would use later in other experiments as well. A second study used a wizard of Oz setting to judge if children would feel any extra cognitive overload while interacting in artificial or constrained languages in comparison to their native languages. By virtue of their self report we did not find any exertion of cognitive overload.

The subsequent phase of the evaluation was conducted at a Dutch high school, where 100 teenagers spent three weeks learning ROILA. We designed a special curriculum to aid in both at home and in school learning and the learning experience of the children was also recorded. Later on, some of the ROILA students were invited to take part in a controlled experiment that compared the recognition accuracy of ROILA against English. The results of the experiment showed that not only was ROILA evaluated better than English with respect to the self report of the students but also achieved significantly better recognition accuracy. We also quantitatively tried to determine the value of the recognition improvement, i.e. for how long must have the students played the game before their efforts in learning ROILA would pay off. Although we achieved promising results but there are still limitations if we try to generalize them. The controlled experiment was conducted with a group of children, who were native Dutch. Moreover we only tested a subset of the ROILA language as the curriculum concentrated on only 50 words and not on the entire vocabulary of 803 words. Our results could have been positively or negatively different if we had used participants who had a different native language or if we had evaluated ROILA in its entirety.

The fascination of the children towards ROILA as a secret language might also explain the positive self report ratings. Had we used adults as ROILA

---

students, we could have found them to be much more critical of learning ROILA. Nevertheless we concluded that the ROILA curriculum was an enjoyable learning activity for the children, where they learnt not only ROILA, but also something about robotics, building LEGO robots, about LEGO components, the problems with speech interfaces, etc.

We are of the opinion that ROILA offers an added dimension in the form of a fun providing element, especially to young children. This is an interesting add-on besides offering improved recognition accuracy. In order to discuss the implications of ROILA with respect to fun for children, we take the help of game heuristics (Malone, 1981), according to which ROILA provides the following benefits: Firstly, ROILA allows for fantasy, purely due to being an unknown and secret language. Throughout the ROILA activity at the school, several children enjoyed this aspect of ROILA and would immediately start talking in ROILA to each other. The nearly significant likeability SASSI rating points in the same direction. Many children took to ROILA personally as a language that only they knew and not everyone around them. Secondly, ROILA also allows for suspense and uncertainty, for example when speech recognition does not work. The children did not know what recognition error will occur and when. Of course this applies to any speech application as long as it is in the realm of gaming and within limits. Thirdly, ROILA supports creativity, such as giving semantics to a word of choice. ROILA provides this flexibility as speech recognition is not constrained by semantics but only by word structure. Therefore by changing the semantics of a ROILA word should not have an effect on the recognition, but it will obviously effect the understanding of the robot. Many children wished to assign meanings of their own choice to the ROILA words, as indicated by their qualitative feedback during the ROILA in class lessons.

We foresee two distinct, perhaps even mutually incompatible road maps that could be followed for the future development of ROILA. The first is a more strict standardized approach and the second is more free and customized. At this point we do not argue which would be better or appropriate.

The first approach deals with maintaining very strict control over ROILA. There should only be a single forum or platform where ROILA is discussed or modified. Only certain people are allowed to make changes to the language, whereas other ROILA speakers can discuss various issues and request for amendment to the original form of ROILA. A ROILA book would be a good stepping stone to accomplish this as it would lay down the basic linguistic principles of the language. So far the current implementation of ROILA has followed this trajectory. We have maintained a single website - <http://roila.org>, which is open to comments but only we as the creators of the ROILA language have right of exercising approval.

The second approach is much more liberal and it follows the wiki mentality. ROILA speakers would be allowed to make changes and contribute to the language as they desire. This ofcourse makes it harder to control the language but it provides subjective benefits to the speaker. We would also like to contemplate how speakers could make changes to the language. In relation to what we

have mentioned prior, i.e. ROILA supports creativity, ROILA also offers some flexibility. The heart of the language is its vocabulary and the vocabulary of ROILA comes from a scalable genetic algorithm. Therefore speakers have the option of adding as many words as they want to an existing word set. Moreover, by adding words they can also add grammar rules as the new words can take the role of new grammatical markings. Speakers also have the freedom of assigning meanings to the ROILA words. We do not specify any restriction of how meanings should be assigned besides, shorter words getting meanings of more frequently used words. With the genetic algorithm there is also a possibility of adding word types to the vocabulary. We recommend that words should follow the CV structure but a speaker might also want to have words such as CV or CVCVCVC or CVCVCVCV. These three types of words currently do not exist in the ROILA vocabulary. Another key option of manipulating and playing around is adding alternative pronunciations to ROILA words. So speakers could add pronunciations of ROILA words based on their native language. Our choice of speech recognizer (Sphinx-4) allows this option, as we indicated earlier. However this could affect the recognition accuracy as the words are created by the genetic algorithm keeping in mind only the original and solitary pronunciation.

We also contemplate on what the future might hold for ROILA in terms of its incorporation into LEGO Mindstorms. At this moment the recognition does not take place on the NXT itself due to resource restrictions, however embedded speech recognition engines could be an alternative, such as Pocket Sphinx. Moreover the modularity of LEGO could be taken advantage of, if for example a brick is created which has all the necessary technology and resources to implement ROILA. That brick could then be placed on potentially every LEGO robot to have it understand and talk back in ROILA.

The future of ROILA also holds promise for the field of Human Robot Interaction in general. ROILA does not have to be restricted to LEGO Mindstorms only but it would also be very relevant for various service robots such as the Nao (Aldebaran-Robotics, 2011) and Roomba. In addition we anticipate that ROILA could also have a societal impact, for example as an interaction training tool between robots and autistic children (Barakova, Gillessen, & Feijs, 2009).

---

## Bibliography

---

- Aldebaran-Robotics. (2011). *The creators of nao*. (<http://www.aldebaran-robotics.com/>)
- Al Mahmud, A., Mubin, O., Octavia, J., Shahid, S., Yeo, L., Markopoulos, P., et al. (2007). amazed: designing an affective social game for children. In *Interaction design for children* (p. 56-59). ACM.
- Amir, A., Efrat, A., & Srinivasan, S. (2001). Advances in phonetic word spotting. In *The tenth international conference on information and knowledge management* (p. 580-582). ACM New York, NY, USA.
- Arsoy, E., & Arslan, L. (2004). A universal human machine speech interaction language for robust speech recognition applications. In S. e. al.[SKP04] (Ed.), *International conference on text, speech and dialogue* (p. 261-267).
- Atal, B. S. (1995). Speech technology in 2001: new research directions. *Proceedings of the National Academy of Sciences of the United States of America*, 92(22), 10046-10051.
- Barakova, E., Gillessen, J., & Feijs, L. (2009). Social training of autistic children with interactive intelligent agents. *Journal of Integrative Neuroscience*, 8(1), 23-34.
- Bartneck, C., Kanda, T., Mubin, O., & Al Mahmud, A. (2009). Does the design of a robot influence its animacy and perceived intelligence? *International Journal of Social Robotics*, 1(2), 195-204.
- Beekes, R. (1995). *Comparative indo-european linguistics: an introduction*. John Benjamins Publishing Co.
- Bellik, Y., & Burger, D. (1994). Multimodal interfaces: new solutions to the problem of computer accessibility for the blind. In *Conference on human factors in computing systems* (p. 267-268). ACM.
- Blue microphones. (2009). (<http://www.bluemic.com/snowflake/>)
- Boros, M., Eckert, W., Gallwitz, F., Grz, G., Hanrieder, G., & Niemann, H. (1996). Towards understanding spontaneous speech: Word accuracy vs. concept accuracy. *Arxiv preprint cmp-lg/9605028*.
- Breemen, A., Yan, X., & Meerbeek, B. (2005). icat: an animated user-interface robot with personality. In *Fourth international conference on autonomous agents and multi agent systems*. Utrecht.
- Breugelmans, S., & Poortinga, Y. (2006). Emotion without a word: Shame and guilt among raramuri indians and rural javanese. *Journal of Personality*

- and Social Psychology*, 91(6), 1111-1122.
- Brown, J. (2008, December 9, 2008). *Welcome to loglan.org*. (Vol. 2008) (No. December 16). (<http://www.loglan.org/>)
- Brysbaert, M., & New, B. (2009). Moving beyond ku era and francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for american english. *Behavior Research Methods*, 41(4), 977.
- Bui, T. H. (2006, January 3, 2006). *Multimodal dialogue management - state of the art* (Technical Report). University of Twente, Enschede, The Netherlands.
- Cahn, J. E. (1990). The generation of affect in synthesized speech. *Journal of the American Voice I/O Society*, 8(1), 1-1.
- Carnegie-Mellon-University. (2008). *Sphinx-4* (Vol. 2009) (No. July 20). (<http://cmusphinx.sourceforge.net/sphinx4/>)
- Carroll, J., Sapon, S., & Corporation, P. (1959). *Modern language aptitude test*. Psychological Corp New York.
- Caviness, K. (2008, 2002). *Volapuk* (Vol. 2008) (No. December 16). (<http://personal.southern.edu/caviness/Volapuk/>)
- The centre for speech technology and research*. (2008). (<http://www.cstr.ed.ac.uk/projects/festival/>)
- Chen, F. (2006). *Designing human interface in speech technology*. Springer-Verlag New York Inc.
- Churcher, G. E., Atwell, E. S., & Souter, C. (1997). *Dialogue management systems: a survey and overview* (Tech. Rep.). University of Leeds.
- David, C. (1997). *The cambridge encyclopedia of language*. Cambridge University Press.
- Davis, P. (2000). *An outline of the world language - desa chat* (Vol. 2008) (No. December 16). (<http://www.users.globalnet.co.uk/vidas/desa.htm>)
- Department, I. S. (2008). *World robotics survey* (Tech. Rep.).
- Epstein, M. A. (2000). *All the sounds of all the worlds languages*. (Tech. Rep.).
- Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems*, 42(3-4), 143-166.
- Freetts 1.2*. (2005). (<http://freetts.sourceforge.net/docs/index.php>)
- Fry, J., Asoh, H., & Matsui, T. (1998). Natural dialogue with the jijo-2 office robot. In H. Asoh (Ed.), *In proceedings of the 1998 ieee/rsj international conference on intelligent robots and systems* (Vol. 2, p. 1278-1283). Victoria, B.C., Canada.
- Geutner, P., Denecke, M., Meier, U., Westphal, M., & Waibel, A. (1998). Conversational speech systems for on-board car navigation and assistance. In *Fifth international conference on spoken language processing*. ISCA.
- Gilleland, M. (2002). Levenshtein distance. *Three Flavors*..
- Goodrich, M. A., & Schultz, A. (2007). Human robot interaction: A survey. *Foundations and Trends in HumanComputer Interaction*, 1(3), 203-275.
- Gordon, R., & Grimes, B. (2005). *Ethnologue: Languages of the world*. SIL International: Summer Institute of Linguistics.
- Gould, J., Alfaro, L., Finn, R., Haupt, B., & Minuto, A. (1986). Why reading was slower from crt displays than from paper. *ACM SIGCHI Bulletin*, 17(SI), 7-11.

- 
- Grigornko, E., Sternberg, R., & Ehrman, M. (2000). A theory-based approach to the measurement of foreign language learning ability: The canal-f theory and test. *The Modern Language Journal*, 84(3), 390-405.
- Guangguang, M., Wenli, Z., Jing, Z., Xiaomei, Y., & Weiping, Y. (2009). A comparison between htk and sphinx on chinese mandarin. In *International joint conference on artificial intelligence (ijcai 2009)* (p. 394-397).
- Hämäläinen, A., Boves, L., & De Veth, J. (2005). Syllable-length acoustic units in large-vocabulary continuous speech recognition. In *Proceedings of specom* (p. 499-502).
- Hanafiah, Z. M., Yamazaki, C., Nakamura, A., & Kuno, Y. (2004). Human-robot speech interface understanding inexplicit utterances using vision. In *Chi '04 extended abstracts on human factors in computing systems* (p. 1321-1324). Vienna, Austria: ACM.
- Hinde, S., & Belrose, G. (2001). *Computer pidgin language: A new language to talk to your computer?* (Tech. Rep.). Hewlett-Packard Laboratories.
- Hone, K., & Graham, R. (2001). Towards a tool for the subjective assessment of speech system interfaces (sassi). *Natural Language Engineering*, 6(3, 4), 287-303.
- Houck, C., Joines, J., & Kay, M. (1995). A genetic algorithm for function optimization: A matlab implementation. *NCSU-IE TR*, 95(09).
- Htk - the hidden markov model toolkit (htk). (2010). (<http://htk.eng.cam.ac.uk/>)
- IJsselsteijn, W., Kort, Y. de, & Poels, K. (2008). *The game experience questionnaire: Development of a self-report measure to assess the psychological impact of digital games*.
- James, F. (2002). Panel: Getting real about speech: Overdue or overhyped. In *Conference on human factors in computing systems chi 2001* (p. 708-709). Minneapolis, Minnesota: ACM, New York.
- Janton, P. (1993). *Esperanto: language, literature, and community*. State University of New York Press.
- Johnstone, T. (1996). Emotional speech elicited using computer games. In *Fourth international conference on spoken language processing*. ICSA.
- Julius. (2010). (<http://julius.sourceforge.jp/en.index.php>)
- Jurafsky, D., Martin, J., Kehler, A., Vander Linden, K., & Ward, N. (2000). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition* (Vol. 163). MIT Press.
- Kisa, S. E. (2008). *Toki pona - the language of good* (Vol. 2008) (No. December 16). (<http://www.tokipona.org/>)
- Koester, H. H. (2001). User performance with speech recognition: a literature review. *Assistive Technology*, 13(2), 116-130.
- Kulyukin, V. A. (2006). On natural language dialogue with assistive robots. In *Acm conference on human-robot interaction* (p. 164-171). Salt Lake City, Utah, USA: ACM.
- Ladefoged, P. (2005). *Vowels and consonants* (Second Edition ed.). Blackwell Publishing.
- Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Blackwell Publishers.
- Lamere, P., Kwok, P., Gouva, E., Raj, B., Singh, R., Walker, W., et al. (2003). The cmu sphinx-4 speech recognition system.

- (IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP 2003), Hong Kong)
- Lazzaro, N. (2006). Why we play games: Four keys to more emotion without story. *Hmtad den*, 18.
- Lee, C., Soong, F., & Paliwal, K. (1996). *Automatic speech and speaker recognition: advanced topics*. Kluwer Academic Pub.
- LEGO. (2010). *Lego mindstorms nxt*. (<http://mindstorms.lego.com/en-us/Default.aspx>)
- Lejos - java for lego mindstorms. (2009). (<http://lejos.sourceforge.net/>)
- Liu, C., & Melnar, L. (2006). Training acoustic models with speech data from different languages. In *Multilingual speech and language processing*. Cite-seer. (Multilingual Speech and Language Processing)
- Lopes, L. S., & Teixeira, A. (2000). Human-robot interaction through spoken language dialogue. In A. Teixeira (Ed.), *In proceedings of ieee/rsj international conference on intelligent robots and systems, 2000. (iros 2000)* (Vol. 1, p. 528-534). Takamatsu, Japan.
- Loquendo. (2011). *Loquendo: global supplier of speech recognition and speech synthesis technology*. (<http://www.loquendo.com/en/>)
- Lovitt, A., Pinto, J., & Hermansky, H. (2007). On confusions in a phoneme recognizer. *IDIAP Research Report, IDIAP-RR-07-10*.
- Lupyan, G., & Christiansen, M. (2002). Case, word order, and language learnability: insights from connectionist modeling. In *24th annual conference of the cognitive science society* (p. 596601).
- MacLean, A., Young, R., Bellotti, V., & Moran, T. (1991). Questions, options, and criteria: Elements of design space analysis. *Human-computer interaction*, 6(3), 201-250.
- Maddieson, I. (1984). *Patterns of sounds*. Cambridge University Press.
- Makhoul, J., & Schwartz, R. (1995). State of the art in continuous speech recognition. *Proceedings of the National Academy of Sciences*, 92(22), 9956.
- Malmkjaer, K., & Anderson, J. (1991). *The linguistics encyclopedia*. Routledge.
- Malone, T. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science*, 5(4), 333-369.
- Mardegan, A. (2008, 2008). *Union mundial pro interlingua*. (Vol. 2008) (No. December 16). (<http://www.interlingua.com/>)
- Microsoft speech technologies. (2007). (<http://msdn.microsoft.com/en-us/speech/default>)
- Moodle. (2010). *Open source community-based tools for learning*. (<http://moodle.org/>)
- Munteanu, C., Baecker, R., Penn, G., Toms, E., & James, D. (2006). The effect of speech recognition accuracy rates on the usefulness and usability of webcast archives. In *Sigchi conference on human factors in computing systems* (p. 493-502). ACM.
- Nuance. (2011). *Dragon naturally speaking*. (<http://www.nuance.com/dragon/index.htm>)
- Ogden, C. (1944). *Basic english: a general introduction with rules and grammar*. K. Paul, Trench, Trubner.
- Okrent, A. (2010). *In the land of invented languages: A celebration of linguistic creativity, madness, and genius*. Spiegel and Grau Trade Paperbacks.

- 
- Parry, T., & Child, J. (1990). Preliminary investigation of the relationship between word, syntax and language proficiency. In *Symposium on language aptitude testing* (p. 30-67). Prentice Hall. (Language Aptitude Reconsidered: Papers Presented at the 11th Invitational Symposium on Language Aptitude Testing, Held Sept. 14-16, 1988, at the Foreign Service Institute Language School, Arlington, Virginia, USA)
- Peterson, D. (2006). Down with morphemes. In *First language creation conference*. Berkeley, USA.
- Pitt, I. J., & Edwards, A. D. N. (1996). Improving the usability of speech-based interfaces for blind users. In *Proceedings of the second annual acm conference on assistive technologies* (p. 124-130). Vancouver, British Columbia, Canada.
- Pomeroy, S. (2003). *Lojban text-to-speech*. (<http://staticfree.info/blog/lang/LojbanText-to-speech.comments>)
- Reetz, H. (2008). *Upsid info* (Vol. 2008) (No. December 18). ([http://web.phonetik.uni-frankfurt.de/upsid\\_info.html](http://web.phonetik.uni-frankfurt.de/upsid_info.html))
- Rosenfeld, R., Olsen, D., & Rudnicky, A. (2001). Universal speech interfaces. *Interactions*, 8(6), 34-44.
- rsynth - text to speech*. (2005). (<http://rsynth.sourceforge.net/>)
- Rubin, J. (1975). What the good language learner can teach us. *Tesol Quarterly*, 41-51.
- Rudnicky, A. (2010). *Sphinx knowledge base tool*. (<http://www.speech.cs.cmu.edu/tools/lmtool-new.html>)
- Samudravijaya, K., & Barot, M. (2003). A comparison of public-domain software tools for speech recognition. In *Workshop on spoken language processing* (p. 125-131). ISCA.
- Shahid, S., Krahmer, E., & Swerts, M. (2008). Alone or together: Exploring the effect of physical co-presence on the emotional expressions of game playing children across cultures. In *Fun and games: Second international conference* (p. 94-105). Eindhoven, the Netherlands: Springer.
- Shneiderman, B. (2000). The limits of speech recognition. *Commun. ACM*, 43(9), 63-65.
- Shoulson, M. (2008). *The klingon language institute*. (Vol. 2008) (No. December 16). (<http://www.kli.org/>)
- Solomon, C., & Papert, S. (1976). A case study of a young child doing turtle graphics in logo. In *Proceedings of the national computer conference and exposition* (p. 1049-1056). ACM.
- Speech processing, recognition and automatic annotation kit*. (2010). (<http://www.spraak.org/>)
- Spiliotopoulos, D., Androutsopoulos, I., & Spyropoulos, C. D. (2001). Human-robot interaction based on spoken natural language dialogue. In *Proceedings of the european workshop on service and humanoid robots (servicerob 2001)*. Bari, Italy.
- Sporka, A. J., Kurniawan, S. H., Mahmud, M., & Slav, P. (2006). Non-speech input and speech recognition for real-time control of computer games. In *Proceedings of the 8th international acm sigaccess conference on computers and accessibility* (p. 213-220). Portland, Oregon, USA: ACM.
- Springer, M. (2008). *The language glosa*. (Vol. 2008) (No. December 16). (<http://www.glosa.org/en/index.html>)



- Tomko, S., & Rosenfeld, R. (2004). Speech graffiti vs. natural language: Assessing the user experience. In *Proceedings of hlt/naacl*.
- Turnhout, K. (2007). *Socially aware conversational agents*. Unpublished doctoral dissertation.
- Typewell. (2011). *About speech recognition*. (<http://www.typewell.com/speechrecog.html>)
- ULI. (2008, March 18, 2008). *The language ido improved esperanto*. (Vol. 2008) (No. December 16). (<http://www.idolinguo.com/>)
- Verbeek, J., Bouwstra, S., Wessels, A., Feijs, L., & Ahn, R. (2007). Johnny q. In *Design and semantics of form and movement* (p. 182-183).
- VoCollect. (2010). *Speech recognition technology choices for the warehouse*.
- Wald, M., Bell, J., Boulain, P., Doody, K., & Gerrard, J. (2007). Correcting automatic speech recognition captioning errors in real time. *International Journal of Speech Technology*, 10(1), 1-15.
- Weilhammer, K., Stuttle, M., & Young, S. (2006). Bootstrapping language models for dialogue systems. In *Ninth international conference on spoken language processing (interspeech 2006)*. ISCA Archive.
- Yang, J., Yang, W., Denecke, M., & Waibel, A. (1999). Smart sight: a tourist assistant system. In *The third international symposium on wearable computers* (p. 73-78).
- Zimmerman, J., Forlizzi, J., & Evenson, S. (2007). Research through design as a method for interaction design research in hci. In *Proceedings of the sigchi conference on human factors in computing systems* (p. 493-502). ACM.

*Appendix A*

---

**Letter from Christiaan Huygens College  
Eindhoven**

---

We (the teachers of the Christiaan Huygens College Eindhoven) were very pleased to entertain and welcome the ROILA course as part of our science class. The whole activity was very interesting and it was an exciting learning experience for all of us. So much so, we plan to repeat the ROILA curriculum in the future with some adjustments and modifications.

In this evaluation letter we would like to reflect on the entire experience as there were some positives and some learning points that we would wish to address in the future.

Firstly we must acknowledge that the children had a lot of fun interacting with the LEGO robots. At times we felt that they enjoyed this much more than actually learning the ROILA language. By introducing ROILA as a language made specifically for robots the children were also given an opportunity to learn about robotics.

Initially the children did not realize the value of ROILA but once they started interacting with the LEGO robots in English they quickly learnt that ROILA is better understood by the robots as compared to English.

The use of Moodle was also encouraging and it proved to be a good add-on to the in class teaching. Most of the children struggled to learn the entire 50 words of the ROILA vocabulary, so we think this number was too much for them. In the future it might be a good idea to restrict this number.

In the future we would like to make some adjustments to the curriculum. We plan to introduce how speech understanding works for machines as this is similar to how humans perceive sound. Moreover in the classroom we would like to concentrate more on the practical elements, i.e. give more opportunities to the children to interact with the robots. During the practical part of the lessons, we used several robots at one time, where each robot was controlled by a group of children. At times, interferences from other groups reduced the understanding of the system. We also felt that the interaction system was not very user friendly but this is probably because it was a research prototype. In the future it would be nice if the students are also taught how to use the system and make changes to the software.

All in all, we were very positive about the ROILA activity and hope to make it a permanent part of our science curriculum.

We wish the ROILA research team all the best for the future!

On behalf of the Science Teachers for Grade 3 at the Christiaan Huygens College Eindhoven

Marjolein Seegers

*Appendix B*

---

**English to ROILA dictionary**

---

English	ROILA
able	fumela
about	kapim
accident	fituje
account	menoka
acid	watenu
across	josoko
act	jilufe
addition	wosebi
after	bafema
again	wikef
against	fifufu
agreement	similu
air	wifawe
all	bomem
almost	fipolu
among	nelobi
amount	pemona
and	sowu
angle	supofa
angry	likoja
animal	popike
answer	fokofo
any	sojan
anything	kela
apple	pikulo
approval	wiweno
argument	sisana
arm	lanuwe
army	kalutu
art	kowonu
as	muwak
at	kufim
attack	kisate
attempt	sekuja
attention	jobeku
authority	pubowu
awake	pakume
away	bufak
baby	besesu
back	nole
bad	topik

English	ROILA
bag	jotaka
balance	sumafu
ball	jinolu
band	lomuto
base	nasupi
basket	temini
bath	nolosu
be	mufe
beautiful	bujufi
because	kijo
bed	fisama
bee	tusani
before	wetow
behavior	pofuko
belief	woseti
bell	mojefi
bent	wusiwu
between	folifo
big	kasok
bird	mipuki
birth	nuwofo
bit	fatimu
bite	minino
bitter	tomeku
black	japipa
blade	tibawu
blood	fitani
blow	jomemi
blue	kutuju
board	lefabi
boat	jonofi
body	finos
bone	pebafu
book	fojato
boot	toloji
bottle	lupuma
bottom	lawuti
box	lujusi
boy	belutu
brain	kinili
branch	wanowi

---

English	ROILA
brass	tobaka
bread	nususa
breath	mekobi
brick	wajipi
bridge	masete
bright	mesewi
broken	konowi
brother	bufofe
brown	lesaka
brush	tajumu
bucket	wapisi
bug	lamine
building	jisape
burn	lisune
burst	witofi
business	bolemo
but	fojib
butter	punolo
button	kawuba
by	pofan
cake	mebiju
camera	limuna
can	leto
card	kapuwe
care	besowu
carriage	wulelu
cart	wenosi
cat	lakowo
cause	bomefu
certain	kebase
chain	pufoni
chance	faliwu
change	fanajo
cheap	nafoju
cheese	mokafu
chemical	tojiki
chest	mimujo
chief	kinelu
chin	tikulu
church	kumake
circle	poteto

---



---

English	ROILA
clean	jekute
clear	fomuja
clock	likopu
cloud	tofine
coat	mikafa
cold	bosipu
collar	tupebe
color	wekepo
come	fiki
comfort	sinefa
committee	pomofe
common	menomo
company	jakapu
competition	pulawu
complete	lukana
complex	tewola
condition	motaju
connection	pabipe
control	jasinu
cook	mawobu
copy	lufimu
cotton	tafuwo
cough	wikito
country	fulinu
cover	josipi
cow	pekumi
crack	nimawu
crazy	patap
credit	masabi
crime	kosino
cruel	sepawi
crush	sisumo
cry	lanila
cup	lukile
current	sipoja
curtain	tutuko
cut	fawofo
damage	nijofa
danger	metoto
dark	junobi
daughter	fonebe

---

English	ROILA
day	wepip
dead	biboko
dear	febiti
death	bukuf
debt	tafite
decision	litifa
deep	kinipi
degree	susita
delicate	wefipu
design	sopoma
desire	puwase
destruction	tofomu
detail	sawuli
development	tinalo
device	wemete
difference	juwike
different	fesisi
direction	pesiko
dirty	bujeti
discovery	tofumu
discussion	titame
disease	pawume
distance	pejabu
division	senoto
do	bimuj
dog	fipuko
door	bowata
doubt	lelaka
down	petej
drain	wineno
drawer	tibawi
dress	junuku
drink	fabutu
driving	kiwulo
drop	jasupa
dry	mewuse
dust	petiju
ear	nipepo
early	jimope
earth	jiseku
east	luwali

English	ROILA
edge	pileki
education	samala
effect	petamu
egg	pefawa
electric	sinuju
end	pekot
engine	nipumo
enough	besili
equal	tekafo
error	wekapa
even	tawut
event	pasote
ever	wufiw
every	bekusa
everything	jusaw
example	nuboka
exchange	safile
existence	tofiwu
experience	lepepu
expert	pomino
eye	lifelu
face	bubime
fact	fomose
fall	jetepe
false	pujowi
family	biwome
far	fejupu
farm	notupu
fat	kifefi
father	batuwa
fear	lajaso
feel	wetok
feeling	fopena
female	nojafu
field	kulami
fight	wikab
finger	mupine
fire	nejoj
first	titib
fish	busenu
fixed	ninebi

---

English	ROILA
flag	sikufe
flame	welona
flat	pawoba
flight	lewute
floor	jipipi
flower	piwaja
fly	kefoji
fold	winusu
food	pabulu
foolish	sijulu
foot	sotuja
for	bijej
force	kufake
fork	weseto
form	mififa
forward	koloke
frame	talamu
free	fojane
friend	biketa
from	nikuf
front	womamu
fruit	wikute
full	fosefi
fun	babot
furniture	kelupu
future	jipime
game	simoti
garden	panope
general	jifaka
get	butij
girl	batuno
give	bufo
glass	leneno
glove	wanita
go	kanek
goat	tumofu
gold	kifiso
good	wopa
government	lasobo
grass	sobinu
great	toton

---



---

English	ROILA
green	koleti
grey	welipu
grip	wawiji
group	makuni
guide	sibapo
gun	fekopu
hair	funuka
hammer	timofi
hand	jiwos
hanging	lokatu
happy	bojiju
harbor	tosanu
hard	bomuwo
harmony	wobiju
hat	lebije
hate	fekaku
have	saki
he	liba
head	babej
healthy	pelite
hearing	memopa
heart	fafaku
heat	mipute
help	tabob
here	fiwas
hi/hello	jabami
high	fijefi
history	kekisi
hole	fibale
hollow	tipeja
hook	mopepu
hope	bolowe
horn	pulaso
horse	jubusi
hospital	jawawo
hour	fulina
house	bubas
how	lipam
humor	sokewi
I	pito
idea	bitute

---



English	ROILA
if	kenet
ill	pubemu
important	fesuni
in	bafop
increase	wipamu
industry	tofoba
ink	wujine
inside	pawop
instrument	wofenu
insurance	momibu
interest	lunema
iron	sewulo
island	mofetu
it	supa
jelly	wutusa
jewel	wunufa
join	kekulo
journey	saluwe
judge	kesomu
jump	kuloja
keep	babusa
key	jusuti
kick	kokafo
kind	batapa
kiss	jeleno
knee	suweno
knife	makose
know	bati
knowledge	pekiwi
land	seseme
language	nawobu
last	wonat
late	buloku
laugh	lekulo
law	jewomo
lead	kelali
learning	pusike
leather	tawopu
left	webufo
leg	linolo
let	lutot

English	ROILA
letter	kemamu
level	luleje
library	pitebe
life	fenob
lift	nelete
light	foteli
like	jutof
limit	tewusi
line	fewiti
lip	tufona
liquid	woseta
list	kesalo
listen	lulaw
little	kute
living	funite
lock	linapi
long	lepol
look	nokit
loose	mimane
loss	nufoki
loud	mobino
love	loki
low	likobi
machine	kukoma
make	pisal
male	nelifi
man	losa
manager	mitika
map	nitojo
mark	kemoli
market	muwase
married	fasife
mass	sinaju
match	mababi
material	pokule
may	bemotu
meal	nulano
measure	tupaji
meat	mewite
medical	lobifa
meeting	janoka

---

English	ROILA
memory	mafeje
metal	sawena
middle	jumila
military	mimena
milk	mifini
mind	bibiji
mine	busibe
minute	bitipa
mixed	powitu
money	fesis
monkey	newuwa
month	jopofi
moon	komupe
morning	bijina
mother	fikuj
motion	semeje
mountain	nasali
mouth	jojitu
move	biliju
much	sufal
muscle	tawole
music	fupuma
must	waboki
nail	semeji
name	wafub
narrow	wuweme
nation	pumeti
natural	mijafi
near	kelumo
necessary	manuta
neck	lifelo
need	pubow
needle	tobako
nerve	pisulo
net	subefe
new	kulil
news	fowewo
night	telal
no	buse
noise	nefina
normal	kukeme

---



---

English	ROILA
north	lejoku
nose	kululu
not	bimuw
note	lokubo
now	lamab
number	felit
nut	sosisu
observation	wuneja
of	fomu
off	pusif
offer	kisume
office	fifiji
oil	mimubu
old	bajato
on	bawot
one	kilu
only	kopo
open	mifuf
operation	mesuku
opinion	mikuwi
opposite	sukisu
or	buno
orange	pokoku
order	fumene
organization	tewowe
other	wolej
out	kaben
outside	bajike
oven	wepuka
over	pofop
owner	pisosa
page	mowewe
pain	jomala
paint	munune
paper	banafu
part	busamu
past	jejale
paste	kewopo
payment	tulofo
peace	kuwuje
pen	peneko

---

English	ROILA
pencil	watule
person	tiwil
physical	pabeji
picture	witajo
pig	mojuwo
pin	sopale
pipe	sawusa
place	bajeji
plane	jonubo
plant	tiwaba
plate	pefuse
play	biwasa
please	sapup
pleasure	kepila
pocket	nanube
point	fasoli
poison	pepima
polish	weboko
political	posapu
poor	jatola
porter	sosufe
position	komete
possible	jijaba
pot	pobapo
potato	tojufe
powder	sopunu
power	fasewa
present	jukafi
price	lonope
print	sajeku
prison	lamela
private	kipupu
process	nutemi
produce	topuse
profit	totuba
property	nifese
protest	wijite
public	koteja
pull	jalaju
pump	tilose
punishment	tefumo

English	ROILA
purpose	nebaja
push	kufoli
put	tobop
quality	semife
question	fifupo
quick	jimeja
quiet	jetupo
quite	fifiko
rain	mabosu
range	pitele
rat	nimeto
rate	peliku
ray	jonawo
reaction	sopoko
reading	lelejo
ready	bipiki
reason	filaja
receipt	wunelu
record	kamoju
red	tifeke
regret	pajafe
regular	nenepo
religion	tapafu
request	pejujo
respect	koniko
responsible	mefibo
rest	felapo
reward	sewetu
rhythm	towuno
rice	sulese
right	besati
ring	jufebu
river	lisimi
road	jikafe
robot	lobo
rod	tujojo
roll	lekalo
roof	nanume
room	bifabe
root	tupoli
rough	mufisu

---

English	ROILA
round	lajeja
rub	sujelu
rule	mafusu
run	bobuja
sad	lekaja
safe	jalawe
sail	tatewa
salt	sanena
same	jebab
sand	putiba
say	palak
scale	wejeta
school	bokubo
science	mulisi
screw	mowapa
sea	lesuma
seat	kijopo
second	bufawa
secret	jilaja
secretary	nijabi
see	make
seed	wufobi
seem	jalinu
self	takanu
send	fofuki
sense	janulu
separate	posiji
serious	jabiju
servant	tiwine
sex	nafewu
shake	mofata
shame	mimeme
sharp	pibuba
she	mona
sheep	tejoni
ship	jitufi
shirt	malifu
shock	nupawo
shoe	nonisa
short	kamipu
shut	bulumu

---

English	ROILA
side	tuwun
sign	jamawu
silk	watuwi
silver	nofiju
simple	julewa
sister	fofala
size	malula
skin	mesuba
skirt	watemu
sky	melifu
sleep	masup
slip	pefowe
slow	kipupi
small	jatuwe
smash	wiwofe
smell	kelowo
smile	lililo
smoke	lanuna
smooth	seseфу
snake	pojeja
snow	nojebu
so	jobew
soap	sukumi
society	nikajo
sock	wepipo
soft	ninota
solid	fewisu
some	nutat
something	puku
son	bipane
song	jowatu
sort	futaba
sound	mifemo
soup	pelake
south	latibo
space	lalaso
special	futatu
spoon	wosufo
spring	nojime
square	nobuki
stage	mebame
star	kepetu

---

English	ROILA
start	bofute
statement	nubuno
station	kifeji
stay	tipet
steam	tebewe
steel	sepata
step	jesime
stick	jonefu
stiff	wabewa
still	wimut
stitch	wukewu
stomach	nesamu
stone	mipesi
stop	babalu
store	kepete
story	fefule
straight	jekesa
strange	jutota
street	jabube
stretch	tafali
strong	jupusa
structure	wakuse
substance	wutobo
such	bubafo
sudden	nibofo
sugar	motiwi
suggestion	tineti
summer	kikama
sun	fokibu
support	luteka
surprise	junapu
sweet	lenesi
swim	niwaku
system	jufifo
table	jineme
tail	pesuna
take	nomes
talk	seni
tall	nimewi
taste	luluno
tax	tafapu

English	ROILA
teaching	samafe
test	kefoju
than	wobap
that	pimo
then	pikik
theory	nupike
there	fopaf
thick	tanepu
thin	safufi
thing	sowob
this	bamas
though	fiwoju
thought	tukaj
throat	nanipi
through	bebibe
thumb	tofepa
thunder	tekuka
ticket	matubu
tight	lunupa
till	fosonu
time	nojob
tin	wiloki
tired	jijoso
toe	tijumi
together	bitabu
tommorrow	bojifu
tongue	noleka
too	peka
tooth	tawuno
top	jalule
touch	jajujo
town	buwija
trade	natuta
train	jopofo
transport	tokijo
tray	wofuba
tree	latewi
trick	majefi
trouble	fawufi
true	busapa
turn	botama

---

English	ROILA
twist	timabi
two	seju
umbrella	wujone
under	buneka
unit	nakelo
universe	tasesa
up	kape
use	seput
value	powako
very, <i>word marker for plural</i>	tuji
vessel	wejewi
view	molali
violent	sojufu
voice	kafame
waiting	fepasa
walk	fosit
wall	kubitu
want	jiwi
war	folopi
warm	lujesu
wash	minuba
waste	lopapi
watch	bolapo
water	tejim
wave	pubito
wax	wenafe
way	nawe
weather	nejapa
week	fapiko
weight	musoko
well	lukot
west	lepasso
wet	mojemu
what	biwu
wheel	paketa
while	bofabi
whip	temuwi
whistle	sujosi
white	fepaka
who	mumub
why	mojuf

English	ROILA
wide	pewebe
will	nibif
wind	lijowe
window	kajona
wine	lepoba
wing	puwitu
winter	pataje
wire	nuwoma
wise	nupomu
with	bopup
woman	nipib
wood	pakula
word	fatatu
word <i>marker for future tense</i>	jifo
word <i>marker for past tense</i>	jifi
work	towo
worm	wamilu
wound	papebe
writing	linufu
wrong	bemeko
year	buliwi
yellow	wipoba
yesterday	joninu
you	bama
young	fafobe

*Appendix C*

---

## **ROILA Homework Lessons**

---



# Learning ROILA - Level 1

[Introduction](#)

[001 - Alphabet](#)

[002 - Commands](#)

[003 - Simple](#)

[Sentences](#)

[004 - More than](#)

[One](#)

[005 - Adjectives](#)

[006 - And, Or](#)

[007 - Past and](#)

[Future](#)

[008 - More](#)

[Numbers](#)

[009 - Adding Up](#)

[010 - Mine and](#)

[Yours](#)

[Review Cards -](#)

[PDF](#)

[English - ROILA](#)

[Glossary](#)

[ROILA - English](#)

[Glossary](#)

[ROILA Language](#)

[Summary](#)

[How to Learn a](#)

[Language](#)

[About this Course](#)

[TOC](#) | [Next](#)

# Learning ROILA

People talk to people all the time. That's how we learn about the world around us, make friends, get help, and help others. Since there are many different kinds of people around the world, there are also many different languages those people use to communicate with one another. If you want to communicate with someone who uses a different language than yours, one of you has to learn to speak the other's language.

This course will teach you how to speak a completely new language specially invented for talking with robots. After all, what good is a robot if you can't tell it what you want it to do? And if a robot has a problem doing its assigned job, why not let it ask you for help, instead of just running around in circles, bumping into walls? Communication is the key to cooperation, to getting things done. And language is the key to communication.

This new language is called ROILA, which stands for RObot Interaction LAnguage. It is designed scientifically to be easy to learn, easy to pronounce, easy understand, and easy to use for communicating with robots. That makes it perfect for communicating with people, too!

To get started, let's take a look at ROILA's simplified alphabet and pronunciation. Click on this link: [Next: Alphabet](#)

# Lesson 001 - Alphabet

ROILA has five vowels — **a, e, i, o, u**. Pronounce them like this:

**a - aa**, as in hat, fast  
**e - eh**, as in red, fed  
**i - ee**, as in machine  
**o - oah**, as in frost  
**u - uh**, as in but

There are 11 consonants — **b, f, j, k, l, m, n, p, s, t, w**. Pronounce these exactly as you do in English.

**Note:**

There are 16 letters in the ROILA alphabet. ROILA does not use the letters **c, d, g, h, q, r, v, x, y, z**.

Here is a sentence in ROILA — **Pito saki lujusi**. (pee-toh sah-ki luh-juh-see) - which means, word for word, "I have box". ROILA does not use the articles "a", "an", or "the", but you would add those into your English translation. The sample sentence would be, "I have a box", or if you are talking about a particular box, "I have the box."

As you can see from the example, each syllable in ROILA has only one vowel. And each syllable starts with a consonant. That makes ROILA very easy to pronounce, read, and write. You simply say exactly what you see, and write exactly what you hear. In a ROILA spelling bee, everyone wins the first prize!

Stress, or accent, on syllables in a word does not matter in ROILA. Try to say each syllable in a word with the same even emphasis. Remember, this is a language for robots! If you really want to sound a little more human, you could put just a little stress on the first syllable of words, to show that you are starting a new word. That's not really needed, but it might sound a bit more friendly to your listeners.

Now you know everything you need in order to read ROILA out loud and have every ROILA student understand exactly what you are reading.

## 002 - Commands

ROILA is a language designed for telling robots what to do. So, to use ROILA, let's create an imaginary robot, named Lobo, and give it some commands. Here is your first ROILA vocabulary list, with words you can use to tell Lobo what to do:

**kanek** - go  
**koloke** - forward  
**botama** - turn  
**webufo** - left  
**besati** - right  
**nole** - back, backwards

Copy this vocabulary into your notebook, with the page title "002". Writing vocabulary words is the first step to learning the new words. Take a minute to do it now, and you will save yourself hours of work later on!

Here are some basic commands. Commands begin with a verb (the action), followed by an adverb (how to do the action).

**Kanek koloke.** - Go forward.  
**Kanek besati.** - Go right.  
**Kanek webufo.** - Go left.  
**Kanek nole.** - Go back).  
**Botama webufo.** - Turn left.  
**Botama besati.** - Turn right.  
**Botama nole.** - Turn back.

Write these commands in your notebook on the "002" page.

Now, you try some commands! Here is an alphabet grid. Let's put Lobo on letter "A", pointing toward letter "B". Assume that Lobo moves one letter at a time

A B C D E  
F G H I J  
K L M N O  
P Q R S T  
U V W X Y

Where is Lobo after these commands? **Kanek koloke. Botama besati. Kanek koloke. Kanek koloke. Botama webufo. Kanek koloke.**

Click here for the answer.

....

Lobo is on letter M

More commands - **Kanek koloke. Botama webufo. Kanek koloke. Kanek koloke.** Where is Lobo now?

Click here for the answer.

....

Lobo is on letter Y

For practice, get a partner to play the role of Lobo the robot. Give commands to Lobo to move around the room. Lobo will follow your instructions, and move just one step for each **kanek** instruction. Remember that **Kanek webufo** or **Kanek besati** tells Lobo to step sideways, not to turn. When Lobo gets good at following your commands, switch places with your partner so you can be Lobo for a while, to practice hearing and following instructions in ROILA.

## 003 - Simple Sentences

Build sentences in ROILA like you do in English. Sentences start with a noun (the subject), followed by a verb (the action), then if the verb requires it, another noun (the direct object, or what the verb is acting upon). The basic structure looks like this: **subject > verb > object**. Adverbs come after the verb, and before any direct object: **subject > verb > adverb > object**. Capitalize the first letter of the first word in a sentence, just like you do in English.

We need some nouns and verbs to show how this works. Here is your second ROILA vocabulary, with some pronunciation hints. Copy this vocabulary into your notebook under the page title "003":

**jimeja** - quickly  
**kipupi** - slowly  
**buse** - no, not  
**kufim** - to, toward  
**jutof** - like  
**make** - see  
**bama** - you  
**pito** - I

### Note:

**Pito** means both "I" and "me". Use the same word, **pito** for "I like you" (**Pito jutof bama**) and "You like me," (**Bama jutof pito**).

### Note:

Translate **make** as "see", "am seeing ", or "are seeing ", whichever makes the most sense to you. **Make** means all of those variations of "look". Verbs in ROILA do their jobs without the extra little words and spelling changes you see in English. **Pito make bama** means either "I see you ", or "I am seeing you ". After all, both sentences really mean the same thing.

### Reading:

You should be able to figure out what these next ROILA sentences mean. You may use your notebook, with the vocabulary from the previous lesson, to help you understand the sentences.

**Pito make Lobo. Lobo make pito. Lobo kanek kufim pito. Pito jutof bama. Pito kanek jimeja kufim bama. Bama jutof pito. Bama kanek kipupi kufim pito. Lobo botama kipupi. Pito botama jimeja kutim Lobo. Lobo make pito. Lobo buse botama. Lobo kanek nole kipupi. Pito buse kanek. Lobo buse kanek nole. Pito buse make bama.**

Click here for help.

....

I see Lobo. Lobo sees me. Lobo is going toward me. I like you. I am going quickly toward you. You like me. You are going slowly toward me. Lobo turns slowly. I am turning quickly toward Lobo. Lobe sees me. Lobo stops turning. Lobo goes backward slowly. I stop (I no go). Lobo stops going backward. I do not see you

### Practice:

Let's check your understanding. Try to translate these sentences into ROILA. Write your translations in your notebook, with the title "003". Then exchange your translations with another ROILA student so

you can check each other's work.

I am going forward toward you. You are going toward me. You like me. I go left. You go right. Lobo goes backwards quickly. You are not going (You not go). Lobo sees you. Lobo is going toward you. You see Lobo. You do not like Lobo. You go left quickly. Lobo goes right quickly. You go backwards slowly. Lobo goes forward quickly toward you.

How did you do? That wasn't hard at all, was it? Don't worry — we'll make the next lesson harder!

## 004 - More than One

### Vocabulary:

**kilu** - one  
**seju** - two  
**tewajo** - three  
**tuji** - many (plural marker)  
**jinolu** - ball  
**jesime** - step  
**saki** - have

English and many other languages change the form of a word to show plurals, or more than one. "Book" means one, "books" means more than one book. In ROILA, words don't change. To show plurals, ROILA adds the word **tuji** (many). **Jinolu** is one ball, **jinolu tuji** is many balls. If you take one step, it's **jesime**. If you take several steps, it's **jesime tuji**.

### Note:

ROILA does not have words for "a", "an", and "the". You may add those little words into your English translations of ROILA sentences when needed, because your readers will expect to see them in English.

### Examples:

Add **tuji** after a word to show that the word is plural:

**Pito saki jinolu** . - I have a ball.  
**Pito saki jinolu tuji** . - I have balls (ball many).  
**Bama make jinolu** . - You see a ball.  
**Bama make jinolu tuji** . - You see many balls.  
**Lobo kanek jesime tuji** . - Lobo is going many steps (step many).

To describe exactly how many objects there are, add a number before the noun:

**Pito saki jinolu** . - I have a ball.  
**Pito saki kilu jinolu** . - I have one ball.  
**Pito saki seju jinolu tuji** . - I have two balls (two ball many).  
**Lobo kanek koloke tewajo jesime tuji** . - Lobo goes forward three steps (three step many).

### Reading:

**Bama saki kilu jinolu. Pito saki seju jinolu tuji. Lobo saki tewajo jinolu tuji. Lobo kanek koloke jimeja seju jesime tuji. Lobo botama webufo. Lobo kanek nole kilu jesime. Pito make bama. Pito kanek seju jesime tuji kufim bama. Bama kanek jimeja nole seju jesime tuji. Pito kanek webufo tewajo jesime tuji. Bama kanek besati tewajo jesime tuji. Pito kanek jimeja kilu jesime kufim Lobo. Bama kanek kipupi tewajo jesime tuji kufim pito.**

Click here for help.

....

You have one ball. I have two balls. Lobo has three balls. Lobo is going forward quickly two steps. Lobo is turning left. Lobo goes backward on step. I see you. I go two steps



toward you. You go slowly backward two steps. I go left three steps. You go right three steps. I go quickly one step toward Lobo. You go slowly three steps toward me.

**Practice:**

Translate these sentences, in your notebook.

You see many balls. I have two balls. You have three balls. Lobo does not have a ball. I go quickly two steps forward. You go slowly three steps backward. Lobo is going to the right. Lobo is going to the left two steps. Lobo is going forward one step. I am not going backward. I am going forward three steps.

## 005 - Adjectives

### Vocabulary:

**tifeke** - red  
**wipoba** - yellow  
**wekepo** - color  
**kasok** - big  
**kute** - little  
**malula** - size  
**wapisi** - bucket  
**biwu** - what (question word)  
**wopa** - good, okay, right

We use adjectives to describe objects, for instance to tell about the object's color or shape or size. Colors are adjectives, and words like "big" and "little" are adjectives. In ROILA, we use adjectives the same way we use them in English. Put adjectives in front of the nouns they describe in sentences: **Pito make tifeke jinolu** (I see a red ball). **Bama saki kute wapisi** (You have a little bucket).

Numbers describe "how many", so we use them like adjectives, and place them in front of the noun: **Lobo saki seju jinolu tuji** (Lobo has two balls). You can use several adjectives together to describe objects, as you do in English: **Pito make kilu wipoba jinolu** (I see one yellow ball).

### Note about "is":

In English, you might say, "The ball is red." You already know that in ROILA, you don't need to use "the" or "a". To make things really simple, ROILA also does not use "is" for sentences like "The ball is red." So, in ROILA, "The ball is red" becomes simply "Ball red" - **Jinolu tifeke**.

### Asking Questions:

To ask about something, start your question with **biwu** (what). ROILA does not use question marks, so simply end the question with a period (full stop): **Biwu wekepo kute jinolu.** (What color is the little ball? **Biwu malula wapisi.** (What size is the bucket? **Biwu tuji jinulo bama saki.** (What many ball you have - How many balls do you have?)

Turn a statement into a question by adding **biwu** to the beginning of the sentence. **Bama saki jimulo.** (You have a ball.) **Biwu bama saki jinulo.** (Do you have a ball?)

### Reading:

**Pito kanek koloke seju kute jesime tuji. Bama kanek webufo seju kasok jesime tuji. Wopa. Pito make jinulo tuji. Pito jutof kasok jinulo tuji. Pito saki kilu kasok wapisi. Biwu bama saki wapisi. Wopa. Pito saki seju wapisi tuji. Biwu wekepo wapisi tuji. Kute wapisi tifeke. Kasok wapisi wipoba. Biwu wapisi tuji saki jinulo tuji. Wopa. Biwu wekepo jinulo tuji. Kute wapisi saki kilu tifeke jinulo. Kasok wapisi saki tewajo kasok jinulo tuji.**

[Click here for help.](#)

....

I am going forward two little steps. You are going left two big steps. Okay. I see many balls. I like the big balls. I have one big bucket. Do you have a bucket? Yes. I have two

buckets. What color are the buckets? The little bucket is red. The big bucket is yellow. Do the buckets have balls? Yes. What color are the balls? The little bucket has one little red ball. The big bucket has three big yellow balls.

**Practice:**

Translate these sentences, in your notebook.

I see many buckets. Do you have a bucket? Yes. How many buckets do you have? I have three buckets. What color is the big bucket? The big bucket is red. What size are the yellow buckets? The two yellow buckets are little. What color are the balls? I do not have yellow balls. I have three little red balls. Lobo is going forward two little steps. You are going left one big step. You have one yellow bucket. I am going right three big steps. I have three red buckets.

## 006 - And, Or

### Vocabulary:

**pojos** - zero  
**fibi** - four  
**jitat** - five  
**silif** - six  
**kutuju** - blue  
**koleti** - green  
**tobop** - put  
**lamab** - now (at this time)  
**wekapa** - error  
**bemeko** - wrong  
**wolej nawe** - other way  
**sowu** - and  
**buno** - or

**Note:** Use the conjunctions **buno** (or) and **sowu** (and) like you use "or" and "and" in English: **Pito saki jinulo sowu wapisi.** (I have a ball and a basket). **Lobo kanek webufo buno besati.** (Lobo goes left or right).

### Reading:

**Pito saki kutuju wapisi tuji sowu koleti jinolu tuji. Biwu bama saki wapisi tuji. Bama saki kilu tifeke sowu wipoba wapisi. Lobo saki pojós jinulo. Kanek koloke fibi sowu jitat jesime tuji kufim Lobo. Lamab botama besati sowu make kutuju jinulo. Bemeko. Kanek wolej nawe. Biwu bama make jinulo tuji. Wopa. Pito make silif kutuju jinulo tuji sowu kilu tifeke jinulo. Pito make pojós kutuju jinulo. Tobop fibi bunó jitat jinulo tuji. Bama saki seju wekapa tuji. Tobop kutuju jinulo sowu koleti wapisi. Lamab bama saki pojós wekapa.**

Click here for help.

....

I have blue buckets and green balls. Do you have many buckets? You have one red and yellow bucket. Lobo has zero balls. Go forward four or five steps toward Lobo. Now turn right and see a blue ball. Wrong. Go the other way. Do you see many balls? Yes. I see six green balls and one red ball. I see zero blue balls. Put four or five balls. You have two errors. Put a blue ball and a green bucket. Now you have zero errors.

### Practice:

Translate these sentences, in your notebook.

Put one blue ball and one green ball. Wrong. I see an error. You put two blue balls. Now go the other way and put one yellow ball. Good. Lobo has zero errors. Go toward me or toward Lobo. Now go backwards and put a ball. Do you have four buckets or five buckets? No. I have six buckets.

## 007 - Past and Future

### Vocabulary:

**jifi** - past tense (marker)  
**jifo** - future tense (marker)  
**bafop** - into, in  
**nelete** - pick (up), lift  
**jasupa** - put down, drop (v.)  
**lujusi** - box  
**bileki** - carry, bring  
**bobuja** - run  
**fosit** - walk

### Note on Past, Present, Future:

ROILA verbs show action in the present. **Pito fosit** means "I walk" or "I am walking". To show that an action happened in the past, add the past tense marker word **jifi** after the verb: **Pito fosit jifi** - I walked, I was walking. To show that an action will happen in the future, add the future tense marker word **jifo** after the verb: **Pito fosit jifo** - I will walk, I will be walking.

### Reading:

**Nelete lujusi. Pito nelete jifo lujusi. Bileki lujusi kufim pito. Pito bileki lujusi. Pito nelete jifi lujusi sowu jasupa jifi lujusi. Pito bobuja jifi. Lamab pito fosit jijo. Lobo tobop jinolu bafop lujusi. Lobo nelete jifi lujusi sowu bileki jifi lujusi kufim pito. Lobo jasupa jifo lujusi. Pito bobuja jifo sowu bileki jifo lujusi kufim bama. Bama tobop jifo seju jinolu tuji bafop lujusi. Pito saki jifo tewajo jinolu tuji bafop lujusi. Bama fosit jifi kipupi. Biwu bama bobuja jimeja.**

[Click here for help.](#)

....

Pick up the box. I will pick up the box. Carry the box to me. I am carrying the box. I carried the box to you and put the box down. I was running. Now I will walk. Lobo will put a ball into the box. Lobo picked up the box and carried the box to me. Lobo will put down the box. I will run and carry the box to you. You will put two balls into the box. I will have three balls in the box. You were walking slowly. Will you run quickly?

### Practice:

Translate these sentences, in your notebook.

I have two balls in a box. I will carry the box to you. You had three boxes. Now you have four boxes. I will run toward you. I walked slowly. You put down the boxes and picked up a basket. I will put balls into the basket. I carried a little box. Now I will carry a big box.

---

## List of Publications related to this research

---

1. Mubin, O., Shahid, S., van de Sande, E., Krahmer, E.J., Swerts, M.J.G., Bartneck, C. and Feijs, L.M.G (2010). *Using Child-Robot Interaction to Investigate the User Acceptance of Constrained and Artificial Languages*. To Appear in the Proceedings of the 19th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2010, Viareggio, Italy, pp. 588-593. *Comprised in Chapter 5*.
2. Mubin, O., Bartneck, C., & Feijs, L. (2010). *Towards the Design and Evaluation of ROILA: A Speech Recognition Friendly Artificial Language*. In H. Loftsson, E. Rgnvaldsson & S. Helgadtir (Eds.), *Advances in Natural Language Processing, Proceedings of the 7th International Conference on Natural Language Processing (IceTAL 2010, Reykjavik, Iceland)* (Vol. LNAI/LNCS 6233/2010, pp. 250-256). *Comprised in Chapter 3*.
3. Mubin, O., Bartneck, C. and Feijs, L. (2010). *Using Word Spotting to Evaluate ROILA: A Speech Recognition Friendly Artificial Language*. *Proceedings of the 28th International Conference on Human Factors in Computing Systems (CHI2010), Atlanta* pp. 3289-3294. *Comprised in Chapter 3*.
4. Mubin, O., Bartneck, C., & Feijs, L. (2009). *Designing an Artificial Robotic Interaction Language*. In T. Gross, J. Gulliksen, P. Kotz, L. Oestreicher, P. Palanque, R. O. Prates & M. Winckler (Eds.), *Human-Computer Interaction - INTERACT 2009* (Vol. LNCS 5727/2009, pp. 848-851). Berlin: Springer. *Comprised in Chapter 2*.
5. Mubin, O., Bartneck, C., & Feijs, L. (2009). *What you say is not what you get: Arguing for Artificial Languages Instead of Natural Languages in Human Robot Speech Interaction*. *Proceedings of the Spoken Dialogue and Human-Robot Interaction Workshop at IEEE RoMan 2009, Toyama*. *Comprised in Chapter 2 and 3*.
6. Mubin, O., Shahid, S., Bartneck, C., Krahmer, E., Swerts, M., & Feijs, L. (2009). *Using Language Tests and Emotional Expressions to Determine the*

- Learnability of Artificial Languages*. Proceedings of the 27th International Conference on Human Factors in Computing Systems (CHI2009), Boston pp. 4075-4080. *Comprised in Chapter 5*.
7. Suleman Shahid, Emiel Krahmer, Marc Swerts, Omar Mubin: *Who is more expressive during child-robot interaction: Pakistani or Dutch children?* In the ACM Proceedings of HRI 2011: 247-248
  8. Shahid, S., Krahmer, E.J., Swerts, M.J.G., and Mubin, O. (2010). *Child-Robot Interaction during Collaborative Game Play: Effects of Age and Gender on Emotion and Experience*. In the ACM Proceedings of the OzCHI 2010 Conference: 332-335.
  9. Shahid, S., Sande, van de E., Mubin, O., Krahmer, E.J., & Swerts, M.J.G. (2009). *Child-robot interaction: Investigating emotional expressions and social interaction of children in a collaborative game*. In the 2nd International conference on human-robot personal relationships, Tilburg.
  10. Omar Mubin, Abdullah Al Mahmud: *Exploring multimodal robotic interaction through storytelling for Aphasics*. British Computer Society HCI (Vol 2) 2008: 145-146

---

## Summary

---

### **ROILA: Robot Interaction Language**

The number of robots in our society is increasing rapidly. The number of service robots that interact with everyday people already outnumbers industrial robots. The easiest way to communicate with these service robots, such as Roomba or Nao, would be natural speech. However, the limitations prevailing in current speech recognition technology for natural language is a major obstacle behind the unanimous acceptance of speech interaction for robots. Current speech recognition technology is at times not good enough for it to be deployed in natural environments, where the ambience influences its performance. Moreover, state-of-art automatic speech recognition has not advanced far enough for most applications, partly due to the inherent properties of natural languages that make them difficult for a machine to recognize. Examples are ambiguity in context and homophones (words that sound the same but have different meanings). As a consequence of the prior discussed problems at times miscommunication occurs between the user and robot. The mismatch between humans' expectations and the abilities of interactive robots often results in frustration for the user. Palm Inc. faced a similar problem with hand writing recognition for their handheld computers. They invented Graffiti, an artificial alphabet, that was easy to learn and easy for the computer to recognize. Our Robot Interaction Language (ROILA) takes a similar approach by offering a speech recognition friendly artificial language that is easy to learn for humans and easy to understand for robots with an ultimate goal of outperforming natural language in terms of speech recognition accuracy. There exist numerous artificial languages, Esperanto for example; but to the best of our knowledge these artificial languages were not designed to optimize human machine/robot interaction but rather to improve human-human communication.

The design of ROILA was an iterative process having iterations within each step. It started off with a linguistic overview of a pre-selection of existing artificial languages across the dimensions of morphology (grammar) and phonology (the sounds of the language). The artificial languages were also analyzed in comparison to natural languages. The overview resulted in a number of linguistic trends that we would carefully incorporate in the design of ROILA with



the claim that whatever linguistic features are common amongst these existing languages would be easier to learn if they are made part of ROILA. The actual construction of the ROILA language began with the composition of its vocabulary. A genetic algorithm was implemented which generated the best fit vocabulary. In principle, the words of this vocabulary would have the least likelihood of being confused with each other and therefore be easy to recognize for the speech recognizer. Experimental evaluations were conducted on the vocabulary to determine its recognition accuracy. The results of these experiments were used to refine the vocabulary. The third phase of the design was the design of the grammar. Using the questions, options, and criteria (QOC) technique, rational decisions were made regarding the selection of grammatical markings. Recognition accuracy and ease of human learnability were two important criteria. In the end we drafted a simple grammar that did not have irregularities or exceptions in its rules and markings were represented by adding isolated words rather than inflecting existing words of a sentence. As a conclusion to the design phase and also as a proof of concept we designed an initial prototype of ROILA by using the LEGO Mindstorms NXT platform. ROILA was demonstrated in use to instruct a LEGO robot to navigate in its environment, analogous to the principles of the turtle robot.

As a final evaluation of ROILA we conducted a large scale experiment of the language. ROILA was exposed to Dutch high school students who spent three weeks learning and practicing the language. A ROILA curriculum was carefully designed for the students to aid them in their learning both in school and at home. In-school learning was more interactive and hands on as the students tested their ROILA skills by speaking to and playing with LEGO robots. At the end of the curriculum the students attempted a ROILA proficiency test and if successful they were invited to play a complete game with a LEGO robot. Throughout the whole learning process, subjective and objective experiences of the students was measured to determine if indeed ROILA was easy to learn for the students and easy to recognize for the machine. Our results indicate that ROILA was deemed to have a better recognition accuracy than English and that it was preferred more by the students in comparison to English as their language of choice while interacting with LEGO Mindstorms robots.

---

## **Omar Mubin: Curriculum Vitae**

---

Omar Mubin was born on 29-01-1983 in Lahore, Pakistan. He began his formal education by completing a Bachelors of Science (BSc) degree with a major in Computer Science in May 2004 at the Lahore University of Management Sciences (LUMS), Lahore, Pakistan. He then graduated with a Masters of Science (MSc) in Information Technology with a specialization in Interactive Systems Engineering at the Royal Institute of Technology (KTH), Stockholm, Sweden in September 2005. This was followed by the attainment of a Professional Doctorate in Engineering (PDEng) qualification in User System Interaction at the Eindhoven University of Technology (TU/e), The Netherlands in October 2007. As part of the PDEng qualification, Omar Mubin was a visiting researcher at Philips Research Eindhoven from January 2007 to October 2007. From January 2008, he started a PhD at the Department of Industrial Design, Eindhoven University of Technology (TU/e), the Netherlands of which the results are presented in this dissertation. Since April 2011, he is employed as a post doctoral researcher in the CRAFT lab at Ecole Polytechnique Federale de Lausanne (EPFL), Switzerland.



**Jewomo kilu.**

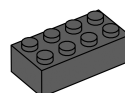
Lobo waboki buse nijofa losa  
bebibe jilufe buno buse jilufe.

**Jewomo seju.**

Lobo waboki nomes jilufe sojan fumene tuji bufo jifi pofan  
losa kenet similu bopup jewomo kilu.

**Jewomo tewajo.**

Lobo waboki pisal jalawe bamas fenob fomu tekanu kenet  
similu bopup jewomo kilu sowu jewomo seju.



**Jewomo kilu.**

Lobo waboki buse nijofa losa  
bebibe jilufe buno buse jilufe.

**Jewomo seju.**

Lobo waboki nomes jilufe sojan fumene tuji bufo jifi pofan  
losa kenet similu bopup jewomo kilu.

**Jewomo tewajo.**

Lobo waboki pisal jalawe bamas fenob fomu tekanu kenet  
similu bopup jewomo kilu sowu jewomo seju.

