Leveraging Text-to-Scene Generation for Language Elicitation and Documentation

Morgan Ulinski

Submitted in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2019

© 2019

Morgan Ulinski

All Rights Reserved

ABSTRACT

Leveraging Text-to-Scene Generation for Language Elicitation and Documentation

Morgan Ulinski

Text-to-scene generation systems take input in the form of a natural language text and output a 3D scene illustrating the meaning of that text. A major benefit of text-to-scene generation is that it allows users to create custom 3D scenes without requiring them to have a background in 3D graphics or knowledge of specialized software packages. This contributes to making text-to-scene useful in scenarios from creative applications to education. The primary goal of this thesis is to explore how we can use text-to-scene generation in a new way: as a tool to facilitate the elicitation and formal documentation of language. In particular, we use text-to-scene generation (a) to assist field linguists studying endangered languages; (b) to provide a cross-linguistic framework for formally modeling spatial language; and (c) to collect language data using crowdsourcing. As a side effect of these goals, we also explore the problem of multilingual text-to-scene generation, that is, systems for generating 3D scenes from languages other than English.

The contributions of this thesis are the following. First, we develop a novel tool suite (the WordsEye Linguistics Tools, or WELT) that uses the WordsEye text-to-scene system to assist field linguists with eliciting and documenting endangered languages. WELT allows linguists to create custom elicitation materials and to document semantics in a formal way. We test WELT with two endangered languages, Nahuatl and Arrernte. Second, we explore the question of how to learn a syntactic parser for WELT. We show that an incremental learning method using a small number of annotated dependency structures can produce reasonably accurate results. We demonstrate that using a parser trained in this way can significantly decrease the time it takes an annotator to label a new sentence with dependency information. Third, we develop a framework that generates 3D scenes from spatial

and graphical semantic primitives. We incorporate this system into the WELT tools for creating custom elicitation materials, allowing users to directly manipulate the underlying semantics of a generated scene. Fourth, we introduce a deep semantic representation of spatial relations and use this to create a new resource, SpatialNet, which formally declares the lexical semantics of spatial relations for a language. We demonstrate how SpatialNet can be used to support multilingual text-to-scene generation. Finally, we show how WordsEye and the semantic resources it provides can be used to facilitate elicitation of language using crowdsourcing.

Table of Contents

List of Figures v

List of Tables viii

List of Abbreviations x

Chapter 1. Introduction 1

1.1 Motivation 2

1.1.1 Endangered Language Preservation 2

1.1.2 Cross-linguistic Modeling of Spatial Language 3

1.2 Tools and Resources 5

1.3 Outline of Thesis 6

Chapter 2. Background on WordsEye and VigNet 10

2.1 Introduction 10

2.2 Related Work 12

2.3 The WordsEye Text-to-Scene System 13

2.4 VigNet 16

2.5 Conclusion 20

Chapter 3. The WordsEye Linguistics Tools (WELT): Using Graphics Generation in Linguistic Fieldwork 21

3.1 Introduction 21

3.2 Related Work 24

3.3 Languages 25
3.3.1 Arrernte 26
3.3.2 Nahuatl 27
3.4 Overview of WELT 28
3.4.1 Modi cations to WordsEye 30
3.5 Grammar Development for Arrernte 33
3.6 Elicitation Tools (WELT English) 34
3.6.1 Cultural Adaptation of VigNet 34
3.6.2 Preparing Scenes and Eliciting Data: Nahuatl Topological Relations . 37
3.7 Documentation Tools (WELT L2) 39
3.7.1 The Lexicon and Ontology Browser 40
3.7.2 Documenting Semantics 41
3.7.3 WELT L2 Text-to-Scene Generation 45
3.8 Future Work: Investigation of Case in Arrernte 48
3.9 Conclusion 51

Chapter 4. Incrementally Learning a Dependency Parser to Support Language Documentation in Field Linguistics 52
4.1 Introduction 52
4.2 Related Work 53
4.3 Corpus 55
4.4 Parsing Experiments and Results 57
4.5 Initial Experiments with Nonlexical Models 60
4.6 Annotation Experiments and Results 62
4.7 Conclusion 67

Chapter 5. 3D-Scene Generation from Semantic Primitives 69
5.1 Introduction 69
5.2 Related Work 71
5.3 Modi cation of WordsEye 72
5.3.1 Adding Semantic Input and Output 72

5.3.2 Spatial and Graphical Primitives 76
5.3.3 Evaluation of Semantic Output 76
5.4 Application to Elicitation in Field Linguistics 78
5.5 Conclusion 82

Chapter 6. SpatialNet: A Declarative Resource for Spatial Relations 84
6.1 Introduction 84
6.2 Related Work 86
6.3 Background on FrameNet and VigNet 87
6.4 Structure of SpatialNet 88
6.4.1 Ontology of Semantic Categories 89
6.4.2 Spatio-graphic Primitives 90
6.4.3 Spatial Frames 92
6.4.4 Spatial Vignettes 95
6.5 Using SpatialNet for Text-to-Scene Generation 99
6.6 Conclusion 100

Chapter 7. Language Elicitation via Crowdsourcing 104
7.1 Introduction 104
7.2 Elicitation of Emotional Language 105
7.2.1 Corpus Collection 105
7.2.2 Emotion Classi cation Experiments 108
7.3 Elicitation of Spatial Language 114
7.3.1 Future Work: Active Learning of Spatial Language 117
7.4 Elicitation of Imaginative Language 120
7.4.1 Crowdsourcing Methodology 120
7.4.2 Evaluating WordsEye: Imaginative and Realistic Sentences 123
7.5 Conclusion 127

Chapter 8. Conclusion 128
8.1 Summary of Contributions 128

8.1.1 Tools and Resources 129
8.2 Future Work 131

Bibliography 133

Appendix A. Custom Arrernte-Related WordsEye Content 149

Appendix B. Nahuatl Topological Relations 155
B.1 WordsEye scenes for Max Planck Topological Relations 155
B.2 Nahuatl elicitations 166

Appendix C. Arrernte Data: Spatial Language and Case 172

Appendix D. Primitive Spatial and Graphical Relations 197
D.1 Spatial relations 198
D.2 Size and shape 199
D.3 Part/whole relations 200
D.4 Object properties 201

List of Figures

2.1 WordsEye System Architecture 13

2.2 A portion of the WordsEye ontology 18

2.3 Two frames augmented with primitive graphical relations 19

3.1 WELT work ow 29

3.2 WELT architecture 31

3.3 WordsEye interface to the WELT elicitation tools 32

3.4 Screenshots of WELT elicitation interfaces 35

3.5 Illustration of Arrernte sentence Ipmenhe-ipmenhele kwatye urinpe-ilemele iteme, \My grandmother is boiling the water." 36

3.6 (Instantiated) vignettes for the woman boils the water 37

3.7 WordsEye scenes using custom 2D gum tree and dingo from our image corpus 38

3.8 Screenshots of the ontology browser 42

3.9 (a) c-structure and (b) f-structure for artwe le goal arrerneme 43

3.10 Creating the left-hand (syntax) side of a syntax-to-semantics rule in WELT . 44

3.11 Creating the right-hand (semantics) side of a syntax-to-semantics rule in WELT 46

3.12 A syntax-to-semantics rule created in WELT. 47

3.13 Selecting among possible f-structures produced by XLE from an ambiguous grammar. 48

3.14 WELT L2 interface for generating a scene from Arrernte text 49

3.15 WordsEye scenes to elicit the \dative of attempt." 50

4.1 Example dependency structures 56
4.2 Screenshot of software used for annotation experiments; the sentence shown here was pre-processed using the baseline method. 63
4.3 Average annotation time for each sentence size, across all annotators 65
4.4 Average annotation time for each annotator, across all sentence sizes 65
4.5 Parser accuracy on each sentence size, for baseline method and for MaltParser (arc-eager) trained on 1, 5, or 25 sentences. 66

5.1 WordsEye system architecture, modi ed to include semantic output and input modules and to separate primitive spatial and graphical relations from other relations in VigNet 73
5.2 Two possible semantic representations for the bottle is in the basket 75
5.3 Using WordsEye backdrops to create scenes that are relevant to endangered language speakers 80
5.4 Screenshots of the new user interface for 3D-scene generation in our elicitation tool 81
5.5 Using semantic output/input to x errors in WordsEye's processing of a gray house with a red door 83

6.1 Examples of new spatial primitives added for SpatialNet 92
6.2 Examples of spatio-graphic primitives: (a) on-top-surface , (b) on-front-surface , and (c) next-to and English/German descriptions. . . 93
6.3 Declarative format for spatial frames 94
6.4 Declarative format for spatial vignettes 95
6.5 Spatial vignettes for di erent meanings of English prepositions 96
6.6 Spatial vignettes for di erent meanings of German prepositions 97
6.7 Pipeline for text-to-scene generation with SpatialNet 98
6.8 Results of morphological and syntactic analysis of German sentences (b) Das Gemälde ist an der Mauer and (c) Das Haus ist an der Mauer. 99
6.9 Final semantic representation for German sentences (b) and (c), used to generate a 3D scene with WordsEye. 101

6.10 A possible sequence of static scenes to represent a dog jumping over a box . . 102

7.1 Completed emotion description HITs 107
7.2 (Top) recall, (middle) precision and (bottom) F-score of the 6-class emotion classi cation problem over the (left) caption dataset and (right) caption+comment dataset 111
7.3 Accuracy of the 6-way emotion classi cation problem 113
7.4 Example of spatial language elicitation HIT 115
7.5 Example of grammaticality veri cation HIT 117
7.6 Imaginative images: situational, iconic, abstract, fantastic 121
7.7 Example of imaginative sentence collection HIT 124
7.8 Examples of the WordsEye evaluation AMT tasks: (a) image comparison task and (b) rating task 126

List of Tables

2.1 Spatial a ordances, represented by the boxes associated with each object . . 17

3.1 Images created for Arrernte WELT 38
3.2 Pictures from the Max Planck Topological Relations Picture Series 39
3.3 Nahuatl elicitations 39
3.4 A mapping from nouns (lexical items) to VigNet semantic concepts 41

4.1 Summary of each language in the corpus. Unless otherwise speci ed, values indicate the count of unique types in the corpus for each category. 57
4.2 Accuracy of each parsing method (to be continued in Table 4.3). 59
4.3 Accuracy of each parsing method (continued from Table 4.2). 60
4.4 Arabic baseline: majority label per fold. 61
4.5 Accuracy of (English) nonlexical model applied to other languages 61
4.6 Time (seconds) for an annotator to label a sentence, for each parsing method and sentence size. 64
4.7 Annotation accuracy for each annotator, across all sentence sizes 67

5.1 Results of evaluation of semantic output 77

6.1 Mapping from German lexical items to VigNet semantic categories, obtained using two di erent German WordNet resources: GermaNet and Open German WordNet. 91

7.1 Distribution of mood labels in crowdsourced emotion dataset 106
7.2 Features selected by mRMR. 112

7.3 Mean and standard deviation of precision, recall, and F-score values for caption dataset and caption+comment dataset 113
7.4 Average confusion matrix for a caption+comments run (66% accuracy) . . . 114
7.5 Spatial descriptions collected with spatial language elicitation HIT 116
7.6 Counts of grammaticality judgments for each input sentence 117
7.7 Categories of words in the lexicon, used to elicit imaginative sentences 122
7.8 Possible combinations of categories for the imaginative sentence construction task 123
7.9 Examples of imaginative sentences collected with AMT 125

List of Abbreviations

1	rst person
3	third person
abl	ablative
acc	accusative
after	`after'-ative
all	allative
as.well	`as well, too, again, still'
asp	aspect
assoc	associative
assoc.mot	associative motion
aver	aversive
but	`now consider this one; on the other hand, by contrast; but'
caus	causative
com	comitative
cont	`do continuously (non-motional)'
dat	dative
do.coming	`do verb action while coming'
do.past	`do verb action while moving past or through a place'
ds	di erent subject
du	dual
eng.tr	marker of English transitive verb
erg	ergative
fact	`it's a fact that'
foc	focus
freq	frequentative
hab	habitual involvement
inch	inchoative
ins	instrumental
loc	locative
nmlz	nominalizer
nom	nominative
npc	non-past completive
npp	non-past progressive
num	number
only	`only, exclusively'
pc	past completive
pl	plural
poss	possessive
prop	proprietive
pst	past

purp	purposive	sembl	semblative
rdp	reduplication	sg	singular
refl	re exive	ss	same subject
rel	relative	tmp.nom	temporal nominal formative
rem.p.hab	remote past habitual	tns	tense

Acknowledgments

First, I would like to thank to my faculty advisor, Julia Hirschberg. I was incredibly lucky to have her as my advisor, and I would not have been able to accomplish as much as I have without her continuous support and encouragement. I am also grateful to Owen Rambow for his guidance and mentorship on many parts of the research that went into this thesis. Thanks also to Claire Cardie, Michael Riley, and John Hale, for their mentorship during my time as an undergraduate, and for providing me with my rst research opportunities in natural language processing and computational linguistics.

Thank you to Mark Dras for giving me my rst opportunity to work on Arrernte, including a trip to Alice Springs in Australia to work directly with Arrernte speakers. I also thank Myfany Turpin for sharing her knowledge and expertise on Arrernte and other Australian languages, both during that trip and after, as we began our work on the WordsEye Linguistics Tools. Thanks also to Daniel Kaufman for two semesters of his language documentation and eld methods class, which provided me with a much stronger background on eld linguistics and endangered languages.

I am grateful to all the members (o cial and honorary) of the Speech Lab for the supportive and welcoming environment I enjoyed during my time at Columbia: Bob Coyne, Fadi Biadsy, Rivka Levitan, Erica Cooper, Daniel Bauer, Arthi Ramachandran, Victor Soto, Sarah Ita Levitan, Nishi Cestero, Rose Sloan, and Zixiaofan (Brenda) Yang.

I am especially grateful to Bob Coyne, without whom this thesis would not have been possible, for his expertise and continual support on all things related to WordsEye. I would also like to thank the many project students who have helped with my work, including: Anusha Balakrishnan, Inna Fetissova, Alexandra Orth, Kevin Kwong, Carly Jackson, Lau-

ren DeLucia, Eden Dolev, and Hannah Ahn.

Finally, I am tremendously grateful to my family for all their encouragement, help, and support over the years. My parents, my sisters, and my husband { none of this would have been possible without you.

Chapter 1

Introduction

Language elicitation is fundamental to many elds of study. Speech pathologists elicit language samples in order to diagnose and treat speech and language problems [Rvachew and Brosseau-Lapre, 2012]. Field linguists elicit language to help preserve endangered languages [Crowley and Thieberger, 2007]. Machine learning scientists elicit language they can use to train classi ers for machine translation [Zaidan and Callison-Burch, 2011], question answering [Chanduet al., 2018], and other tasks [Wanget al., 2012]. Teachers elicit language from students as a regular part of second language instruction [Rosenberg, 2009]. This thesis focuses on eliciting language in the context of both eld linguistics and computational linguistics. Language documentationmay include recording speech and text for archival purposes, annotating or translating language data, writing detailed descriptions of grammars, or creating formal computational models of language [Gipperet al., 2006; Grenoble and Furbee, 2010]. In this thesis, we will address all of these, but our main interest is in what we call formal documentation, by which we mean linguistic grammars that describe a language's abstract system of structures and rules in a way that is both understandable to a linguist reading the grammar and that can also be processed using computational methods. Although the problem of language documentation is often associated primarily with eld linguistics and its relevance to endangered language preservation, we believe that any computational model that is also human-readable is a form of language documentation. This includes many existing computational resources for high-resource languages, including

FrameNet [Ruppenhofer et al., 2016] and WordNet [Fellbaum and Miller, 1998].

The primary goal of this thesis is to explore how we can use text-to-scene generation as a tool to facilitate the elicitation and formal documentation of language. Text-to-scene generation systems take input in the form of a natural language text and output a 3D scene illustrating the meaning of that text. A major bene t of text-to-scene generation is that it allows users to create custom 3D scenes without requiring them to have a background in 3D graphics or knowledge of specialized software packages. This contributes to making text-to-scene useful in scenarios from creative applications to education. In this thesis, we use text-to-scene generation (a) to assist eld linguists studying endangered languages; (b) to provide a cross-linguistic framework for formally modeling spatial language; and (c) to collect language data using crowdsourcing. As a side e ect of these goals, we also explore the problem of multilingual text-to-scene generation, that is, systems for generating 3D scenes from languages other than English.

1.1 Motivation

1.1.1 Endangered Language Preservation

Languages have appeared and disappeared throughout history. Today, however, languages are facing extinction at an unprecedented pace. Over 40 percent of the world's approximate 7,000 languages are at risk of disappearing. When languages die out, we not only lose the ability to study them linguistically, but we also lose access to an invaluable resource for studying the culture, history, and experience of peoples around the world [Endangered Languages Project]. E orts to document languages and develop tools to support these e orts become even more important with the increasing rate of extinction. Bird [2009] emphasizes the particular need to make use of computational linguistics during eldwork. This thesis addresses this need by applying text-to-scene generation to linguistic eldwork and the study of endangered languages. Because 3D scenes naturally lend themselves to the elicitation of spatial and other graphical relations, we focus in particular on methods for the elicitation and documentation of spatial language.

1.1.2 Cross-linguistic Modeling of Spatial Language

Spatial relations have been studied in linguistics for many years. One study for English by Herskovits [1986] catalogs fine-grained distinctions in the interpretation of various prepositions. For example, she distinguishes among the uses of *on* to mean `on the top of a horizontal surface' (the cup is on the table) or `affixed to a vertical surface' (the picture is on the wall). Likewise, Feist and Gentner [1998] describe user perception experiments that show that the shape, function, and animacy of the *Figure* and *Ground* are all factors in the perception of spatial relations as the lexical item *in* or *on*.

The relevance of object properties is language dependent. Bowerman and Choi [2003] describe how Korean linguistically differentiates between the lexical items *nehta*, which indicates putting something in a loose-fitting container (for example, fruit in a bag), versus *kkita*, which indicates putting something in a tight-fitting wrapper (for example, hand in glove). Other languages, including English, do not make this distinction. Levinson [2003] and colleagues have also catalogued profound differences in the ways different languages encode relations between objects in the world. In particular, the Australian language Guugu Yimithirr and the Mayan language Tzeltal use absolute frames of reference to refer to the relative positions of objects. In Guugu Yimithirr, one can locate a chair relative to a table only in terms of cardinal points: saying, for example, that the chair is north of the table. In English such expressions are reserved for geographical contexts (for example, *Seattle is north of Portland*) and are never used for relations at what Levinson terms the \domestic scale". In Guugu Yimithirr one has no choice, and there are no direct translations for English expressions such as *The chair is in front of the table*. These kinds of distinctions, both within a language and between different languages, make the elicitation and documentation of spatial language an interesting and important problem in linguistics and computer science. In particular, tools for eliciting and subsequently documenting this kind of linguistic information should be an important part of endangered language preservation as a whole.

Formal, cross-linguistic models of spatial language are important not only for the purpose of documenting and preserving a diverse set of endangered languages, but also in many natural language processing applications for higher resource languages. Consider the following examples of machine translation. The prepositions are marked in boldface. The English

sentence is a word-for-word gloss of the German sentence except for the preposition.

In our first example, English on is correctly translated to German an:[1]

(1) a. The painting is on the wall.

b. Correct translation: Das Gemälde ist an der Mauer/Wand.

c. Google Translate/Bing Translator (correct): Das Gemälde ist an der Wand.

However, the correct translation changes if we are relating a cat to a wall:

(2) a. The cat is on the wall.

b. Correct translation: Die Katze ist auf der Mauer.

c. Google Translate (incorrect): Die Katze ist an der Wand.

d. Bing Translator (incorrect): Die Katze steht an der Wand.

The problem here is that the English preposition on describes two different spatial configurations: `affixed to', in the case of the painting, and `on top of', in the case of the cat.[2]

Similar problems appear when we translate from German to English. The painting again translates correctly:

(3) a. Das Gemälde ist an der Mauer.

b. Correct translation: The painting is on the wall.

c. Google Translate/Bing Translator (correct): The painting is on the wall.

But when we replace the painting with the house, we no longer obtain the correct translation:

---

[1] Note that English wall should be translated to Wand if it is a wall which has a ceiling attached to it, and Mauer if it is freestanding and does not help create an enclosed three-dimensional space. We ignore this particular issue in this discussion.

[2] We set aside the interpretation in which the cat is affixed to the wall similarly to a clock, which is an extraordinary interpretation and would require additional description in either language.

(4) a. Das Haus ist an der Mauer.

b. Correct translation: The house is at the wall.

c. Google Translate/Bing Translator (incorrect): The house is on the wall.

The problem is again that the German preposition an corresponds to two di erent spatial con gurations, `a xed to' (painting) and `at/near' (house).

In this thesis, we will explore how we can use text-to-scene technology to facilitate the formal modeling of cross-linguistic di erences in the expression of spatial language. These models will be useful in improving machine translation and other multilingual NLP applications.

1.2 Tools and Resources

One of the main innovations of this thesis is that we combine existing tools and resources in di erent ways in order apply them to a new problem. While our focus is on how text-to-scene generation can be applied as a tool to facilitate language elicitation and documentation, our work incorporates many other kinds of tools as well. Some of the tools and resources we use are listed here:

- The WordsEye Text-to-Scene System [Coyne and Sproat, 2001]: to generate custom elicitation materials in Chapter 3; as the foundation of the semantics-to-scene system in Chapter 5; for graphical realization of spatio-graphic primitives in Chapter 6; as the source of images to elicit emotional language in Chapter 7.

- VigNet [Coyne et al., 2011a]: to model lexical semantics in Chapter 3; as the initial source for the graphical primitives and ontology in Chapters 5 and 6; to help elicit imaginative sentences in Chapter 7.

- XFST [Beesley and Karttunen, 2003] and XLE [Crouch et al., 2011]: to create computational models of Arrernte morphology and syntax in Chapter 3.

- FieldWorks Language Explorer [SIL International]: to annotate Arrernte sentences related to spatial language and case in Chapter 3.

- The Max Planck topological relations picture series [Bowerman and Pederson, 1992]: as the model for the topological relations scenes used to elicit Nahuatl in Chapter 3; as a reference for expanding the set of spatio-graphic primitives in Chapter 6.
- The Picture Series for Positional Verbs [Ameka et al., 1999] and the Motion Verb Stimulus Kit [Levinson, 2001]: to obtain sentences for the corpus/treebank introduced in Chapter 4; as a reference for expanding the set of spatio-graphic primitives in Chapter 6.
- Tree Editor (TrEd) [Pajas and Štěpánek, 2008]: to annotate our new corpus with dependency information in Chapter 4; as a component of the UI for the annotation experiments in Chapter 4.
- MSTParser [McDonald et al., 2006] and MaltParser [Nivre et al., 2006]: to train parsers on incrementally increasing amounts of data in Chapter 4.
- FrameNet [Ruppenhofer et al., 2016]: as a prototype for and source of some of our spatial frames in Chapter 6.
- The Princeton WordNet of English [Princeton University], GermaNet [Henrich and Hinrichs, 2010; Hamp and Feldweg, 1997], Open German WordNet [Siegel], and the EuroWordNet Interlingual Index (ILI) [Vossen, 1998]: to create the German lexical mapping in Chapter 6.
- The Stanford CoreNLP Toolkit [Manning et al., 2014]: to perform German and English morphological and syntactic parsing in the SpatialNet text-to-scene pipeline described in Chapter 6.
- Amazon Mechanical Turk [Amazon Mechanical Turk, Inc., 2018]: as the platform for the crowdsourcing examples in Chapter 7.

1.3 Outline of Thesis

Throughout this thesis, we use the WordsEye text-to-scene system [Coyne and Sproat, 2001; Coyne, 2017], which converts English text into 3D scenes representing the meaning of

that text. Our goal is to use 3D scenes created in WordsEye to facilitate the elicitation of language data; subsequently, the text-to-scene system provides a natural semantic grounding for formally documenting a language's lexical semantics. We apply this methodology rst to the problem of eliciting and documenting endangered languages, with the development of the WordsEye Linguistics Tools, or WELT. WELT has two modes of operation, the rst for elicitation and the second for documentation. The rst mode of operation, which we call WELT English, provides tools to assist with the elicitation of language data. We have developed user interfaces that allow users to create custom elicitation materials in the form of 3D scenes and to build and conduct elicitation sessions based around those scenes. We have also created a prototype of the second mode of operation, which we call WELT L2. The purpose of WELT L2 is to provide a way to formally document the semantics of the endangered language. Formal hypotheses speci ed using WELT L2 can be veri ed using a text-to-scene system that takes input in the endangered language, analyzes it based on the formal model, and generates a picture representing the meaning.

We have tested WELT English by creating a set of scenes representing basic topological relations and eliciting descriptions from a native speaker of Nahuatl, an endangered language spoken in Mexico. WELT English has also been used by Anusha Balakrishnan to elicit Ikota, an endangered language spoken along the border of Gabon and the Republic of the Congo. In addition, many of the scenes we used to elicit Nahuatl were created by undergraduate linguistics students who did not have a computational background. These facts demonstrate that although the WELT elicitation tools have not yet been deployed broadly, they are reusable by di erent users and for di erent languages. Our prototype of WELT L2, including converting endangered language text into a 3D scene, works with a limited grammar of Arrernte, an Australian aboriginal language spoken in Alice Springs. The goal of our work on documenting Arrernte with WELT L2 is to show that the success we have had using WELT English with di erent endangered languages can be extended to the documentation tools. In Chapter 3, we describe the WordsEye Linguistics Tools and demonstrate their use with two endangered languages: Nahuatl and Arrernte.

In order to make full use of the WELT work ow, including veri cation of hypotheses via text-to-scene generation, it is necessary for the system to be able to parse input

text into its morphology and syntax in addition to semantics. SIL FieldWorks Language Explorer (FLEx) [SIL International] provides the means to create a morphological parser without knowledge of any particular grammatical formalism, by training the parser in the background as the linguist annotates their language data. The ability to easily document syntax, however, is largely missing from existing documentation tools. In Chapter 4, we explore the question of whether a tool similar to FLEx might potentially be created for syntax. We introduce and describe a new multilingual treebank we have created, consisting of short descriptions of spatial con gurations and motion events, annotated with syntactic dependencies and other linguistic information. We use this treebank to perform experiments in which we train a dependency parser on incrementally increasing small amounts of data. This simulates the scenario of a parser being learned in the background while a linguist gradually annotates more sentences. We also conduct experiments that demonstrate that using a parser trained in this manner improves the performance of human annotators annotating dependency structures. Our experiments show that it is possible to extend FieldWorks' approach to learning a morphological parser into the realm of syntax.

Generating scenes from natural language simpli es the creation of custom 3D content, but there are situations when a natural language interface is not ideal. For example, 3D scenes created from natural language are limited by any errors the system makes processing language of the input text. Even in the absence of system errors, it can be di cult to achieve a precise spatial con guration due to the ambiguity of natural language. In addition, using English to create scenes that will be used for elicitation in another language may introduce unwanted biases and assumptions about the target language. We address these problems by adapting WordsEye into a system that can generate a 3D scene from semantic primitives representing basic spatial relations and graphical properties. This gives users direct access to the underlying semantics of 3D scenes being created instead of relying on black-box conversion from input text to rendered scene. In Chapter 5, we describe the architecture of the system and the incorporation of this functionality into the WELT elicitation tools.

One issue with WELT is that, in order to verify formal models using a text-to-scene work ow, the semantic documentation must ultimately map into VigNet, which, having been developed speci cally for English, is often biased both to the English language and

toward Western culture. Our adaptation of WordsEye to allow scenes to be generated from a language-independent semantic representation enables us to avoid our dependence on the English-speci c components of WordsEye, while still maintaining the bene t of being able to create a text-to-scene system for a language. This adaptation also provides the means to formally document spatial semantics using a language-independent semantic grounding. In Chapter 6, we describe how we use this semantic grounding as part of a new multilingual resource, SpatialNet, which provides a framework for de ning the lexical semantics of spatial relations in a language. SpatialNet uses a declarative format to link linguistic expressions both to semantic frames and to actual spatial con gurations. Because SpatialNet provides a link between surface language and semantic primitives, it can also be used in conjunction with our modi ed WordsEye system to generate 3D scenes from text.

Most of the discussion of language elicitation in this thesis focuses on using text-to-scene generation in the context of endangered languages, with a trained eld linguist eliciting language directly from native speaker informant. However, text-to-scene generation can also be useful to facilitate elicitation of language data in other contexts. In Chapter 7, we show three examples of how WordsEye can be used in conjunction with crowdsourcing to elicit di erent types of language data. First, we use 3D scenes previously created by WordsEye users and published to the WordsEye gallery to create a corpus of descriptions containing emotional language. Second, we show how generating 3D scenes from spatial and graphical primitives can be used with crowdsourcing to elicit targeted descriptions of spatial con gurations. Finally, we use the lexical and semantic information contained in VigNet to obtain a set of descriptions of imaginative scenarios.

In Chapter 8, we conclude by summarizing the contributions of this thesis and discussing future work.

Chapter 2

Background on WordsEye and VigNet

2.1 Introduction

WordsEye [Coyne and Sproat, 2001; Coyne, 2017] is a system for automatically converting (English) natural language text into 3D scenes representing the meaning of that text. To do this, WordsEye relies on VigNet [Coyneet al., 2011a], a knowledge base of lexical, semantic, and graphical information. We will describe VigNet in more detail in Section 2.4.

WordsEye supports language-based control of spatial relations, textures and colors, collections, facial expressions, and poses. It handles simple anaphora and co-reference resolution, allowing for a variety of ways of referring to objects. The system assembles complex scenes from a library of approximately 3,000 3D objects and 10,000 2D images tied to a lexicon of 20,000 nouns. These include a wide variety of common objects (including variations of the same basic type, such as di erent types of doors or chairs) and textures (e.g. wood, grass, granite). WordsEye also supports a range of graphical primitives and properties that are used for spatial relations, (di erent senses of in, on, lateral relations, etc.), spatial properties (absolute and relative sizes and aspect ratios), and surface properties (colors, opacity, re ectivity, etc.). These primitives in conjunction with the objects and semantic knowledge about those objects (such as defaults for size, orientation, and top surface regions) allow

the scene to be composed.

WordsEye is a good choice as a tool for our work for several reasons. First, it is now a reliable web application[1] that has been online since November 2015, with approximately 35,000 registered real-world users. Before the launch of the web application, earlier versions of WordsEye were used for many years on a smaller scale. For example, an earlier online version of WordsEye (then known as Semantic Light) had around 3,000 registered users from 2004{2007. Another version of WordsEye was used in 2010 as an educational tool, helping students develop language skills during a summer literacy enrichment program run by the Harlem Educational Activities Fund [Coyne et al., 2011b]. This history of real-world usage means that WordsEye will be a more stable tool than other text-to-scene systems that have been developed (we will brie y mention some of these in Section 2.2). Second, WordsEye supports a wide variety of spatial relations, graphical properties, and 3D objects. This broad coverage allows for exibility in the kinds of 3D scenes that can be created. In addition, the WordsEye system works by passing input text through a pipeline with distinct modules for di erent stages of language processing and graphical analysis. This allows us to more easily reuse parts of the system while replacing others with new interfaces.

In this thesis, we will discuss using WordsEye in the following ways:

- In Chapter 3, we describe how we are using WordsEye to create a comprehensive tool for eld linguistics, to be used (a) to elicit descriptive language from native speakers of endangered languages and (b) to serve as a test-bed for syntactic/semantic grammars of endangered languages.

- In Chapter 5, we describe using WordsEye to create a new system that can generate a 3D scene from semantic primitives representing basic spatial and graphical relations.

- In Chapter 6, we describe using WordsEye as a realization engine for a deep semantic representation of spatial relations. We also show how this can be applied to support multilingual text-to-scene-generation.

- Finally, in Chapter 7, we show how WordsEye can be used to help with collecting

[1] http://www.wordseye.com

language data using crowdsourcing.

In this chapter, we describe WordsEye and VigNet in some detail, which will help clarify how the di erent parts of the system are used when we discuss them later in this thesis. Section 2.2 discusses related work on text-to-scene generation. Section 2.3 provides an overview of the WordsEye architecture and processing pipeline. Section 2.4 describes VigNet. We conclude in Section 2.5.

2.2 Related Work

Several systems exist for producing graphics from natural language sources. Adorni et al. [1984], Simmons [1975], and the Put system [Clay and Wilhelms, 1996] were early systems used to position objects in a 3D space. Badler et al. [2000] and AVis [O'Kane et al., 2004] used language to control animated characters in a closed virtual environment. CarSim [Dupuy et al., 2001] is a domain-speci c systems to create animations from natural language descriptions of accident reports. Glass [2008] describe a system for transforming text sourced from popular ction into corresponding 3D animations without prior language simpli cation. 3SVD [Zeng et al., 2005] is a 3D scene creation system using story-based descriptions. Parisi et al. [2007] describe an ontology-driven generation of 3D animations for training and maintenance. CONFUCIUS [Ma, 2006] is a multimodal text-to-animation system that generates animations of virtual humans from single sentences containing an action verb. A survey of these and other text-to-graphics systems is given in Hassani and Lee [2016]. In most existing systems, the referenced objects, attributes, and actions are typically relatively small in number or targeted to speci c pre-existing domains. WordsEye was one of the rst text-to-scene systems designed for general use rather than a speci c domain.

Another recent system, introduced in Chang et al. [2014a], focuses on indoor scenes and supports a large number of 3D objects; however, it supports fewer spatial relations and graphical properties than WordsEye. The work described in Chang [2015] also emphasizes incorporating real-world knowledge in order to infer additional information not directly speci ed in the input text. WordsEye, on the other hand, creates a 3D scene that is a more

Figure 2.1: WordsEye System Architecture

precise representation of the meaning conveyed by the input text. In addition, a major focus of the current work on WordsEye is in developing a robust, stable application for real-world users. This makes it a good choice for our research, which requires a text-to-scene system to be a reliable foundation for us to build upon.

2.3 The WordsEye Text-to-Scene System

In this section, we provide an overview of how the WordsEye system works. We explain how WordsEye converts its input from text to semantics to graphical primitives and nally to a rendered 3D scene. This will help clarify how its graphical-semantic primitives, lexical dictionary, knowledge-base, and semantic formalisms are utilized in the tools described in this thesis.

The architecture of WordsEye is shown in Figure 2.1. Input text is converted into a 3D scene by a series of processing stages. Throughout the process, the system relies on VigNet, a knowledge base of lexical, semantic, and graphical information. VigNet includes an ontology of semantic types (including both abstract concepts and speci c 3D objects), a hierarchy of relations, and a set of assertions that represent real-world knowledge by

applying relations to concepts. It also includes links between the concepts/relations and English lexical items. We will describe VigNet in more detail in Section 2.4.

The system operates by first tokenizing each input sentence into lexical items (including modifiers like contractions or possessives) and possible parts-of-speech. It does this by identifying words in its lexicon, which consists of about 30,000 English lexical items. For each lexical item, the dictionary includes information such as verbal inflections, alternative spellings, and singular/plural noun forms. WordsEye tokenizes the text using custom heuristics; tokens are then labeled with (possibly multiple) part of speech tags found in the lexical dictionary. Multi-word prepositions (e.g. *to the left of*) are also identified at this stage. Then text is then parsed into a constituent tree using a custom CYK [Younger, 1967] grammar of approximately 1000 rules. The grammar identifies the head constituent at each level of the parse, allowing the automatic construction of a syntactic dependency tree from the constituency parse. WordsEye's grammar has been optimized to handle descriptions of spatial and graphical configurations. It usually handles low-level lexical-semantic phenomena, such as adverbial modifiers on spatial prepositions, better than off-the-shelf parsers. For example, the English dependency parser in the Stanford CoreNLP Toolkit [Manning et al., 2014] gives incorrect analyses on simple sentences like *the dog jumped 3 feet over the table*, interpreting *3 feet* as a direct object of *jump* (in this case, the noun phrase *3 feet* is acting as an adverbial modifier of the spatial preposition, not as an object of the verb).

After converting the constituent parse into a labeled syntactic dependency structure, the dependency structure is processed for anaphora and other co-reference, which is especially important for depicting multi-sentence input. These resolved structures are converted to lexical-semantic relations using lexical valence patterns and other lexical and semantic information. In this process a lexical item like *on* is converted to a specific graphical primitive relation. For example, in *the picture is on the wall*, *on* is converted to a graphical primitive that puts the object on the front surface of the wall, whereas in *the vase is on the table*, *on* is converted a different primitive that puts the vase on the top surface of the table. The semantic decomposition rules examine the VigNet knowledge base to determine the properties of the objects and pick the correct spatial decomposition. In this case, a wall is known to primarily be a vertical surface, while a table is a horizontal surface.

For higher-level semantic constructs, like verbs and their arguments, the semantics can first be converted into a *vignette*, which is a particular real-world realization of a semantic meaning. For example, the verb *wash* can be realized in many different ways, depending on whether one is washing dishes versus washing one's hair versus washing a car. Vignettes are intermediate graphically-inspired interpretations and representations that are in turn decomposed into low-level graphical primitives. We discuss vignettes in more detail in Section 2.4.

Next, WordsEye chooses objects to satisfy the input semantics. For example, if the word *dog* or concept *dog.n* was specified, it chooses a particular 3D model of a dog to match the given concept. This is primarily a random selection, although certain higher-quality 3D models are given preference over others. After assigning objects, WordsEye performs graphical analysis to compose the scene. The lexical-semantic relations are converted to a graphical semantics, which is a set of graphical constraints representing the position, orientation, size, color, texture, cardinality, and poses of objects in the scene. This graphical semantics can be thought of as a semantic grounding; it is used to construct and render a 3D scene.

Construction of the final graphical semantics involves supplying additional graphical constraints to impose some real-world defaults. For example, scenes are often under-constrained. Normally one expects objects to be on the ground versus floating in the air, but usually the ground is not specifically mentioned in the input. In addition, objects generally do not occupy the same location; for a phrase like *the cat and the dog*, the objects should be next to each other rather than inter-penetrating. WordsEye infers these additional constraints as well as associated objects such as a ground and sky. A similar process occurs with the surface properties of objects (such as colors, textures, and reflectivity) and with light sources. The system then applies the entire set of constraints to position the objects in the scene, and renders the final scene.

A Note on WordsEye External Interfaces. There are two main interfaces to WordsEye which will be used in this thesis: an older desktop application that runs on Mac OS X, and the latest version of WordsEye, which is a web application that can also be accessed

through a web API. We use the desktop application for the development of the WordsEye Linguistics Tools (Chapter 3); we use the web API for generating scenes from semantics primitives (Chapter 5), and as the realization engine for SpatialNet (Chapter 6). For the crowdsourcing experiments in Chapter 7, we used the web application accessed both through a web browser and the API. Although the basic functionality of the two interfaces to WordsEye is the same, there are some di erences. In general, the newer web application provides more functionality, although one notable exception is that the desktop application supports putting characters in di erent poses, while the web application does not.

2.4 VigNet

To interpret input text, WordsEye uses a lexical resource called VigNet [Coyne et al., 2011a]. VigNet is inspired by and based on FrameNet [Baker et al., 1998; Ruppenhofer et al., 2016], a resource for lexical semantics. In FrameNet, lexical items are grouped together in frames according to their shared semantic structure. Every frame contains a number of frame elements (semantic roles) which are participants in this structure. The English FrameNet de nes the mapping between syntax and semantics for a lexical item by rst providing a mapping between the lexical item and frames which it can verbalize, and by providing then for each such frame lists of valence patterns that map syntactic functions to frame elements.

VigNet extends FrameNet in two ways in order to capture graphical semantics, the knowledge needed to generate graphical scenes from language. First, graphical semantics are added to the frames by adding primitive graphical (typically, spatial) relations between the frame element llers. Second, VigNet distinguishes between meanings of words that are distinguished graphically. For example, the speci c objects and spatial relations in the graphical semantics for *cook* depend on the object being cooked and on the culture in which it is being cooked (cooking turkey in Baltimore vs. cooking an egg in Alice Springs), even though at an abstract level *cook an egg in Alice Springs* and *cook a turkey in Baltimore* are perfectly compositional semantically. Frames augmented with graphical semantics are called *vignettes*.

The descriptions of the graphical semantics in vignettes make use of a set of object-

Base	Canopy	Top surface	Side surface
Stem	Cup	Enclosure	Touch-point handle

Table 2.1: Spatial a ordances, represented by the boxes associated with each object, designate regions of those objects used in resolving spatial relations.

centric properties called a ordances [Gibson, 1977; Norman, 1988]. The concept of a ordances has a long history in the study of ergonomics and the psychological interpretation of the environment. WordsEye includes as a ordances any functional or physical property of an object that allows it to participate in actions and relations with other objects. For example, a seat of a chair is used to support a sitter or small object and the interior of a box is used to hold the contents. WordsEye has a rich set of spatial a ordances. Some examples of these are cupped regions for objects to be in, canopies for objects to be under, and top surfaces for objects to be on. See Table 2.1 for more examples of spatial a ordances. A ordances are particularly important for interpreting spatial prepositions, which are often ambiguous; for example apple in the bowl (containment) vs. boat in water (embedding)). To resolve the ambiguity, a constraint is placed on on the ground argument that demands that its ller o ers an interior a ordance.

Information about the 3D objects in WordsEye is organized in VigNet into an ontology. The ontology consists of semantic concepts that are linked together with isa relations. The ontology supports multiple inheritance, allowing a given concept to be a sub-type of more than one concept. For example, a princess.n is a subtype of both female.n and aristocrat.n, and a black-widow.n is a subtype of both spider.n and poisonous-entity.n.

Figure 2.2: A portion of the WordsEye ontology. Semantic types are represented by yellow ovals. Assertions are represented by pink rectangles. The thumbnails represent particular 3D object models.

Types include both 3D objects and more general semantic concepts. For example, a particular 3D model of a dog is a subtype of the general dog.n . Every 3D object has a semantic type and is inserted into the ontology. The 3D parts of 3D objects are also represented as types in the ontology. WordsEye also includes lexicalized concepts (e.g. chair tied to chair.n) in the ontology. If a lexical item has more than one word sense, the di erent word senses are linked to di erent concepts. The semantic concepts in VigNet include the graphical objects available in WordsEye as well as concepts that are not currently supported in WordsEye. While WordsEye might only have a handful of graphical objects for dogs, VigNet has concepts representing all common types of dogs, even if there is no graphical object associated with them.

The ontology includes a knowledge base of assertions that provide more information about semantic concepts. Assertions include sizes of objects and concepts, their parts, their colors, what they typically contain, what a ordances they have, the typical location-

Figure 2.3: Two frames augmented with primitive graphical relations. The high-level semantics of self-motion-from-front.r and self-motion-from-portal are decomposed into di erent sets of objects and primitive graphical relations. Frames are represented by blue octagons, yellow ovals represent semantic constraints, and primitive graphical relations are represented by pink rectangles.

s/habitats of objects, and information about their function. Spatial a ordances and other properties can be applied to both 3D graphical objects and to more general semantic types. For example, the general semantic type cup.n has a cupped region a ordance, since this a ordance is shared by all cups. A particular 3D graphical object of a cup might have a handle a ordance, while another might have a lid a ordance, but these spatial a ordances are not tied to the super-type cup.n . Figure 2.2 shows a small subset of the semantic hierarchy along with examples of assertions.

Figure 2.3 shows an example of two vignettes self-motion-from-front.r and self-motion-from-portal.r . Both are sub-types of self-motion-from.r . The yellow ovals contain semantic constraints on the objects used to instantiate the frame. For example, while the relation self-motion-from-front.r requires only that the source of the motion be a physical-entity.n , self-motion-from-portal.r requires that the source has a door-gate-affordance.n as a part. The primitive graphical relation orientation-away-from.r in the decomposition for the latter then also assigns the argument gnd-part to be the door-gate-affordance in addition to assigning the gnd argument to the source of the motion.

2.5 Conclusion

In this chapter, we have provided an overview of the WordsEye text-to-scene system. We will use WordsEye as a tool for much of the research presented in this thesis, as we explore how text-to-scene generation can be used in new areas and applications. The modularized design of the WordsEye architecture and processing pipeline is a major bene t to the work in this thesis that builds on the WordsEye system. It allows us to easily reuse some parts of the system while replacing others with customized components. For example, the diagram in Chapter 3, Figure 3.2 shows how we modify the WordsEye architecture to be used by the WordsEye Linguistics Tools. Similarly modi ed architecture diagrams are shown in Chapter 5, Figure 5.1, when we discuss modifying the WordsEye pipeline to allow for semantic input and output, and in Chapter 6, Figure 6.7, when we describe using SpatialNet for text-to-scene generation.

Chapter 3

The WordsEye Linguistics Tools (WELT): Using Graphics Generation in Linguistic Fieldwork

3.1 Introduction

Although languages have appeared and disappeared throughout history, today languages are facing extinction at an unprecedented pace. Over 40% of the estimated 7,000 languages in the world are at risk of disappearing. When languages die, we lose access to an invaluable resource for studying the culture, history, and experience of people who spoke them [Endangered Languages Project]. E orts to document languages and develop tools to support these e orts become even more important with the increasing rate of extinction. Bird [2009] emphasizes a particular need to make use of computational linguistics during eldwork.

In this chapter,[1] we describe how we have designed the WordsEye Linguistics Tools, or WELT, [2] to address this issue by helping eld linguists study endangered languages. WELT

[1] Some of the material in this chapter was published in Ulinski et al. [2014a,b], based upon work supported by the National Science Foundation under Grant No. 1160700: \Using Computational Tools to Facilitate Corpus Collection and Language Use in Arrernte (aer)."

[2] In German, Welt means \world".

is a novel tool for the elicitation and documentation of endangered languages. The purpose of WELT is to provide eld linguists with a tool for eliciting endangered language data and formally documenting the semantics of a language. WELT has two modes of operation. The rst mode of operation, which we call WELT English, consists of tools for elicitation. The second mode of operation, which we call WELT L2, consists of tools for documentation. We will demonstrate WELT's use on scenarios involving two endangered languages, Arrernte and Nahuatl.

WELT English provides tools for building and organizing elicitation sessions based on custom 3D scenes. English input automatically generates a picture using WordsEye which can then be used to elicit a description in the target language. The elicitation tools in WELT English provide several advantages over using a set of pre-fabricated static pictures like those commonly used by eld linguists today. Users are not limited to a xed set of pictures but may, instead, create and modify scenes in real time based on the informants' answers. This allows them to create additional follow-up scenes and questions on the y. In addition, since the pictures are 3D scenes, the viewpoint can easily be changed, allowing exploration of linguistic descriptions based on di erent frames of reference. This is particularly useful in eliciting spatial descriptions. Finally, since WordsEye can also be extended to include custom 3D content, the user can customize the images used for elicitation to be maximally relevant to their informants. We have created user interfaces for WELT English which integrate the basic functionality of WordsEye with tools for organizing sets of scenes for elicitation sessions, and other tools for annotating and recording language data during the elicitation session. We have tested these tools by creating a set of scenes representing topological relations and eliciting descriptions from a native speaker of Nahuatl, an endangered language spoken in Mexico. WELT English has also been used by Anusha Balakrishnan to elicit Ikota, an endangered language spoken along the border of Gabon and the Republic of the Congo. This was done as her nal project for a class on eld methods and language documentation at Columbia University. In addition, many of the scenes we used to elicit Nahuatl were created by two undergraduate linguistics students who did not have a computational background. These facts demonstrate that although the WELT elic-

itation tools have not yet been deployed broadly, they are reusable by di erent users and for di erent languages.

WELT L2 provides a means to document the semantics of a language in a formal way. Although there are several existing tools that allow eld linguists to formally document phonetics and morphology, most notably the SIL FieldWorks Language Explorer [SIL International], there is no such tool for the formal documentation of semantics. The WELT documentation tools are designed to address this gap. Formally documenting a language with WELT L2 also creates a text-to-scene system that takes input in the endangered language, analyzes it based on the formal model, and generates a picture representing the meaning. This text-to-scene system will allow linguists to test the theories they develop with native speakers, making changes to grammars and semantics in real time. We have created a prototype of WELT L2 which works with a limited grammar of Arrernte, an Australian aboriginal language spoken in Alice Springs. The goal of our work on documenting Arrernte with WELT L2 is to show that the success we have had with using WELT English with di erent endangered languages can be extended to the documentation tools.

In this chapter, we will describe our development of the two modes of operation, WELT English and WELT L2. We will discuss results from conducting elicitation sessions with a native speaker of Nahuatl, and from documenting the lexical semantics of Arrernte. We will also describe a basic Arrernte text-to-scene system created in WELT. In Section 3.2 we discuss prior work on computational tools for eld linguistics. In Section 3.3, we provide more information about the two endangered languages discussed in this chapter, Arrernte and Nahuatl. In Section 3.4 we present an overview of the WELT system, including the modi cations we made to WordsEye. In Section 3.5, we describe our development of a morphological and syntactic grammar for Arrernte. We describe using WELT English for elicitation in Section 3.6 and describe the WELT L2 tools for language documentation in Section 3.7, including a description of the system for generating 3D scenes from endangered language input. In Section 3.8, we describe our pilot work toward using WELT to study the relationship between case and semantic interpretation of a sentence in Arrernte. We conclude in Section 3.9.

3.2 Related Work

Computational tools for eld linguistics fall into two categories: tools for native speakers to use directly, without substantial linguist intervention, and tools for eld linguists to use. Tools intended for native speakers include the PAWS starter kit [Black and Black, 2009], which uses the answers to a series of guided questions to produce a draft of a grammar. Similarly, Bird and Chiang [2012] describe a simpli ed work ow and supporting MT software that lets native speakers produce useable documentation of their language on their own.

One of the most widely-used toolkits in the latter category is SIL FieldWorks [SIL International], or speci cally, FieldWorks Language Explorer (FLEx). FLEx includes tools for eliciting and recording lexical information, dictionary development, creating interlinear glossed text, analysis of discourse features, and morphological analysis. An important part of FLEx is its \linguist-friendly" morphological parser [Black and Simons, 2008], which uses an underlying model of morphology familiar to linguists, is fully integrated into lexicon development and interlinear text analysis, and produces a human-readable grammar sketch as well as a machine-interpretable parser. The morphological parser is constructed \stealthily" in the background, and can help a linguist by predicting glosses for interlinear texts.

Several computational tools aim to simplify the formal documentation of syntax by eliminating the need to master particular grammar formalisms. First is the PAWS starter kit [Black and Black, 2012], a system that prompts linguists with a series of guided questions about the target language and uses their answers to produce a PC-PATR grammar [McConnel and Black, 2006]. The LinGO Grammar Matrix [Bender et al., 2002] is a similar tool developed for HPSG that uses a type hierarchy to represent cross-linguistic generalizations. The LinGO Grammar Matrix [Bender et al., 2002] facilitates formal modeling of syntax by generating basic HPSG \starter grammars" for languages from the answers to a typological questionnaire. Extending a grammar beyond the prototype, however, does require extensive knowledge of HPSG, making this tool more feasibly used by grammar engineers and computational linguists.

Linguist's Assistant [Beale and Allman, 2011] provides a corpus of semantic representations for linguists to use as a guide for elicitation. After eliciting the language data, a

linguist writes rules translating these semantic representations into surface forms. The result is a description of the language that can be used to generate text from documents that have been converted into the semantic representation. Linguists are encouraged to collect their own elicitations and naturally occurring texts and translate them into the semantic representation.

For semantics, the most common resource for formal documentation across languages is FrameNet [Fillmore et al., 2003]; FrameNets have been developed for many languages, including Spanish, Japanese, and Portuguese. Most start with English FrameNet and adapt it for the new language; a large portion of the frames end up being substantially the same across languages [Baker, 2008]. However, FrameNet is also targeted toward computational linguists. ParSem [Butt et al., 2002] is a collaboration to develop parallel semantic representations across languages, by developing semantic structures based on LFG. Neither of these resources, however, are targeted at helping non-computational linguists formally document a language, as compared to the morphological parser in FLEx or the syntactic documentation in PAWS.

In general, we also lack tools for creating custom elicitation materials. With WELT, we hope to ll some of the gaps in the range of available eld linguistics tools. WELT will enable the creation of custom elicitation material and facilitate the management sessions with an informant. WELT will also enable formal documentation of the semantics of a language without knowledge of speci c computational formalisms. This is similar to the way FLEx allows linguists to create a formal model of morphology while also documenting the lexicon of a language and glossing interlinear texts.

3.3 Languages

We have used WELT with two endangered languages, Arrernte and Nahuatl. This section provides background information about these languages.

3.3.1 Arrernte

Arrernte is an endangered language of the Arandic language group, a group of languages spoken in Central Australia. The main branches of Arrernte can be broadly split into Eastern/Central and Western variants. We will focus on Eastern/Central Arrernte, for reasons of location and resources. Figures for the number of speakers of Arrernte vary quite widely, and are further complicated by questions of degree of uency. The most common gures are 4500-6000 for speakers of the Arandic group as a whole, and about 1800 for speakers of Eastern/Central Arrernte [Aboriginal Art and Culture]. The Arrernte people live in Central Australia, in and around Alice Springs (Mparntwe in Arrernte). Arrernte has a number of characteristics quite unlike English or other Indo-European languages. These include extensive use of morphology, fundamentally free word order (but with word order preferences and restrictions on various sub-parts of the language), lack of a copula verb, and \quasi-in ections" on verbs including a \category of associated motion."

While some aspects of Arrernte are well documented [Wilkins, 1989; Henderson, 1998], others have presented greater challenges. In particular, a number of idiosyncratic lexical and morphological features of the language relating to spatial relations have not previously been given a complete description. These are linguistically interesting because they relate directly to how a language is used by its speakers to describe their perceived reality. In Arrernte, for example, (1) in the category of associated motion there are verb in ections for concepts such as do action x quickly while moving downwards ; (2) indication of location can be much more ne-grained than in English, di erentiating whether a person or thing is at rest, or has been moved, or is being observed; and (3) directions are expressed using cardinal (compass) point terms rather than relative terms. Another topic that needs to be explored in Arrernte is the relationship between case and the semantic interpretation of the sentence; it is possible to signi cantly alter the meaning of a sentence by changing the case on one of the nouns. In a related Arandic language, Kaytetye, adding allative case to the direct object of a sentence with the predicate ltare `shoot' changes the meaning from shooting and hitting the object to ring a shot and not actually hitting it. We will discuss how WELT might be employed to study Arrernte case in Section 3.8.

In terms of linguistic analysis of Eastern Arrernte, there is good coverage of the grammar [Strehlow, 1944; Wilkins, 1989; Green, 1994; Henderson, 1998]. There are also a dictionary [Henderson and Dobson, 1994] and a picture dictionary [Broad, 2008] available.

3.3.2 Nahuatl

Nahuatl belongs to the Aztecan branch of the Uto-Aztecan language family. It is one of the most widely spoken indigenous languages of the Americas, with about 1.5 million speakers, but less than 15% of Nahuatl speakers are monolingual, and Spanish literacy greatly exceeds Nahuatl literacy [Hill and Hill, 1986]. Indigenous populations have become increasingly marginalized in Mexican society; Rolstad [2001] argues that Nahuatl is at serious risk of replacement by Spanish. The Mexican government recognizes 30 distinct varieties of Nahuatl; some varieties have already disappeared and many are severely endangered. The modern varieties show considerable di erences, and not all are mutually intelligible. Some small dialects of Nahuatl still remain virtually undocumented [Grinevald, 2008].

Nahuatl describes spatial relations with relational nouns, which always occur after a noun or a possessive pre x. Most modern dialects of Nahuatl have also incorporated Spanish prepositions that compete with the relational nouns. Other examples of Spanish in uence on the language include simpli cation of the morphology and a tendency toward xed word order. Some varieties of Nahuatl in ect the verb in a way similar to Arrernte associated motion, to show the direction of the verbal action. For example, the pre x on- is used for verbs indicating direction of action away from the speaker [Karttunen, 1992]. The use of these in ections has not been studied much, and it warrants further research.

As far as resources, there are several detailed grammars of Classical Nahuatl published by missionaries, including Horacio Carochi's *Arte de la lengua Mexicana*, and there are several modern textbooks and dictionaries of the larger dialects [Endangered Language Alliance]. There are also some electronic resources, including: an online trilingual dictionary between English, Spanish, and Nahuatl [Wood], a preliminary computational model of Nahuatl morphology in Grammatical Framework (GF) [Ashton, 2013], and a formal morphology in XFST [Maxwell and Amith, 2005]. Amith is also developing the Nahuatl Learning Envi-

ronment, which links lexicon, grammar, and corpus into a multimedia system for research and learning [Amith].

3.4 Overview of WELT

A visual representation of the intended work ow for using WELT is provided in Figure 3.1. At this time, we have used WELT English and WELT L2 only independently of each other, and the work ow as a whole has not been tested in practice. The WELT English tools for scene creation and elicitation are currently useable; the only exception is that adding custom vignettes and custom 2D/3D content requires the assistance of WordsEye developers to make the necessary modi cations to WordsEye and VigNet. We have created a prototype of the WELT L2 tools for modeling and documentation and tested it with a small grammar we have developed for Arrernte. This prototype demonstrates how WELT L2 can potentially be used in eldwork, but the documentation tools will require further development before they can be used more generally. In addition, both WELT English and WELT L2 still rely on the older desktop version of WordsEye. This makes it di cult to release the program to the public for general use, both due to the need to license the 3D objects stored locally on a user's computer and because the desktop application is no longer being supported or updated by the WordsEye developers. In Chapter 5, we will describe some of our work toward updating WELT to use the newest version of WordsEye.

The rst step in the work ow is using WELT English to prepare a set of 3D scenes to be used to elicit targeted descriptions or narratives. An important part of this phase is the cultural adaptation of the graphical semantics used in WordsEye, so that scenes will be relevant to the native speakers a linguist works with. We will discuss cultural adaptation in more detail in Section 3.6.1. Next, the linguist works with an informant to generate language data based on prepared 3D scenes. This can be a dynamic process; as new questions come up, a linguist can easily modify existing scenes or create new ones. WELT English also automatically syncs recorded audio with open scenes and provides an interface for the linguist to write notes, textual descriptions, and glosses. After the elicitation session, the linguist can use WELT English to review the data collected, listen to the audio recorded

Figure 3.1: WELT work ow

for each scene, and revise notes and glosses. The linguist can then create additional scenes to elicit more data or begin the formal documentation of the language. We will discuss creating scenes and eliciting data with WELT in Section 3.6.2, including examples from our elicitation sessions for Nahuatl.

Creating a text-to-scene system with WELT L2 requires formal models of the morphology, syntax, and semantics of a language. The focus of our work on WELT is on modeling the interface between syntax, lexical semantics, and graphical semantics. Therefore, although WELT requires models of morphology and syntax to generate a text-to-scene system, we are relying on third-party tools to build those models. For our pilot work using WELT L2 to model Arrernte, we have used XFST [Karttunen et al., 1997; Beesley and Karttunen, 2003] to model the morphology and XLE [Crouch et al., 2011] to model the syntax in the LFG (lexical-functional grammar) formalism [Kaplan and Bresnan, 1995]. These are mature systems that we believe are su cient for the formal documentation of morphology and syntax. We are also researching other options that would be more accessible to non-computational

linguists. In Chapter 4, for example, we discuss the possibility of automatically creating a syntactic parser for WELT using annotated dependency structures. It is important to note, though, that the modeling done in WELT L2 does not require a perfect syntactic parser, so the syntactic grammars provided as models do not need to be complex (and WELT L2 provides an interface for selecting among ambiguous syntactic structures as part of its text-to-scene pipeline.) We will discuss our grammar development for Arrernte in Section 3.5. WELT L2 also requires the creation of a lexicon for the endangered language that maps lexical items into VigNet concepts; it provides tools for searching the VigNet ontology to nd relevant concepts for the lexicon; we will discuss the lexicon in more detail in Section 3.7.1. WELT L2 also provides user interfaces for modeling the syntax-semantics interface, lexical semantics, and graphical semantics of a language through the creation of syntax-to-semantics rules. We will discuss these in more detail in Section 3.7.2.

Once models of morphology, syntax, and semantics are in place (note that these can be working models, and need not be complete), WELT L2 puts the components together into a text-to-scene system that takes input in the endangered language and uses the formal models to generate pictures. This system can then be used to verify theories with informants and revise grammars if and when an informant indicates that the scene generated for a particular input sentence is not correct. As new questions arise, WELT English can also continue to be used to create elicitation materials and collect linguistic data

3.4.1 Modi cations to WordsEye

The WELT tools described in this chapter use the WordsEye desktop application for Mac OS X. The WordsEye desktop application includes a user interface where the user can type simple sentences that are processed to produce a 3D scene. The user can then modify the text to re ne the scene. In order to use WordsEye with WELT, some modi cations were required, both to the user interface and to the underlying system. The original WordsEye architecture was described in the previous chapter and shown visually in Figure 2.1. The modi ed WordsEye architecture used by WELT is shown in Figure 3.2.

To adapt WordsEye for use with WELT English, we rst made some changes to the user

Figure 3.2: WELT architecture

interface of the WordsEye desktop application. To allow WELT users to focus the attention of an informant on particular items in a scene, we modi ed WordsEye to let individual objects and their parts to be selected and highlighted with a bounding box. We also added the functionality to allow the scene currently open in WordsEye to be accessed by an external application, so that scenes can be saved and reloaded using the WELT user interface. To accomplish this, when WordsEye processes input text in order to generate a scene, it also saves information about the scene to disk, including a JPEG image of the scene, the text used to generate the scene, the position of the camera, and any selected objects that should be indicated with bounding boxes. A screenshot showing the WordsEye interface to the WELT elicitation tools is shown in Figure 3.3. For WELT English, the overall processing pipeline for WordsEye is largely unchanged, taking English text as input and producing a 3D scene representing its meaning. However, in order to allow the cultural adaptation required to create scenes that are maximally relevant to endangered language speakers, we modify VigNet to allow for the addition of custom vignettes and 2D/3D content. For example, for our work on Arrernte, we customize the kick vignette to show Australian football rather than soccer, and change the ground texture from grass to desert dirt.

This new culturally relevant content can also be used in the WELT L2 documentation tools. WELT L2 requires further modi cations of WordsEye in order to support the creation of a text-to-scene system for the endangered language. When a sentence is processed by WordsEye, it goes through three main stages: (1) morphological analysis and syntactic

Figure 3.3: WordsEye interface to the WELT elicitation tools. The \Save Image to Current Scene" and \Save Image to New Scene" buttons are used to transfer information about the scene currently open in WordsEye to WELT.

parsing, (2) semantic analysis, and (3) graphical realization. To produce a text-to-scene system for a new language, WELT must replace the English linguistic processing modules with models for the new language. Since our work on Arrernte uses XFST for morphology and XLE for syntax, our modi cations to WordsEye for the WELT L2 prototype are based on interfacing with these tools. However, it would be fairly simple to adapt our system to substitute other tools in in the future. When text is input into the WELT L2 text-to-scene system, rather than being passed directly into the normal WordsEye pipeline, it is instead passed to an external script that processes it with XLE and the Arrernte grammar. We have added one additional feature to the morphology and syntax module of the WELT L2 text-to-scene system: an interface that allows the user to select an f-structure from multiple options produced by XLE, in case the grammar is ambiguous. This way, it is still possible use the WELT text-to-scene system to verify semantic documentation even if the syntactic documentation is not complete. We will see an example of this in Section 3.7.3. Once

this is done, the f-structure is processed with the syntax-to-semantics rules to produce a lexical-semantic representation that is compatible with WordsEye. This step also requires a lexicon that maps endangered language words into VigNet concepts. The lexical semantic representation is then passed back to WordsEye so that a 3D scene can be generated.

3.5 Grammar Development for Arrernte

We collaborated with researchers at Macquarie University to document Arrernte syntax using LFG [Kaplan and Bresnan, 1995], as part of a project to automatically generate Arrernte text related to Australian football [Lareau et al., 2011; Lareau, 2012]. In LFG, linguistic structure is represented by a parallel, linked combination of a surface-oriented constituent structure (c-structure) and a functional structure (f-structure). The f-structure is a dependency structure that models predicate-argument structure, and a suitable interface to VigNet.[3] The grammar for Arrernte is in two parts, a nite state transducer for the morphology, developed with XFST [Karttunen et al., 1997], and the syntactic grammar developed in XLE [Crouch et al., 2011]. It covers basic sentences and NP structure and a few unusual features of Arrernte: split case pronouns, verbless sentences, associated motion, spatial relationships, and same-subject in ection on the verb [Draset al., 2012].

In deciding to use LFG as the formalism for the grammar, several factors were considered, including linguistic suitability and availability of resources. One of the main attributes of Arrernte is that it is a non-con gurational language (it has a at phrase structure, allowing syntactically discontinuous expressions, and a relatively free word order), and there has already been a substantial amount of work on using LFG to model non-con gurational Australian languages [Simpson, 2007; Nordlinger, 1997; Nordlinger and Bresnan, 2011], making LFG desirable for the Arrernte project in particular. Another point in favor of LFG is that there are already mature and widely used tools for developing grammars. Most

[3] Although it is common with LFG to represent semantics using linear logic and glue semantics [Dalrymple et al., 1993; Dalrymple, 2001], and in fact the Arrernte grammar we use for syntax does have a semantic component that uses glue semantics [Lareauet al., 2012], we do not use that representation. Nor do we need to use LFG's a-structure, since we use an alternative representation for the lexical semantics.

prominent is the Xerox Linguistic Tool (XLE) [XLE Project]. XLE is also the basis for the Parallel Grammar Project (ParGram) [Butt et al., 2002; ParGram / ParSem, 2013], a collaborative effort to develop wide coverage grammars in LFG for multiple languages. These existing grammars and the common standards and tutorials that ParGram provides will also simplify the creation of formal syntactic grammars for WELT L2 as we work toward expanding it for more languages.

3.6 Elicitation Tools (WELT English)

Each elicitation session in WELT is organized around a set of 3D scenes, which are created by inputting English text into WordsEye. The scene that is currently open in the WordsEye application can be saved and added to the WELT session, as indicated in Figure 3.3. Scenes that have been previously added to the session can be re-opened in WordsEye and modified, either overwriting the original scene or saving the changes as a new scene. Each scene can be annotated with with textual descriptions, glosses, and notes. Audio for the session can be recorded, and the recording is automatically saved and synced with timestamps for the scenes open in WELT. The audio can be played back to review any given scene. Scenes can be imported and exported between sessions, so that useful scenes can be reused and data compared. Screenshots of the WELT main elicitation interface and the interface for annotating a scene are included in Figure 3.4.

3.6.1 Cultural Adaptation of VigNet

As we discussed in Chapter 2, semantics in WordsEye are represented with VigNet, a resource which groups similar lexical items, along with graphical relations, into frames called vignettes. We use VigNet to represent the semantics in WELT.

Large parts of VigNet are language- and culture-independent. The low-level graphical relations used to express graphical semantics are based on physics and human anatomy and do not depend on language. However, the graphical semantics for a vignette may be culture-specific, and some new vignettes may need to be added for a culture. In the United States, for example, the sentence *The woman boiled the water* might invoke a scene with a

Figure 3.4: Screenshots of WELT elicitation interfaces. Left: WELT interface for managing a session. Right: WELT interface for annotating a scene.

pot of water on a stove in a kitchen. Among the Arrernte people, it would instead invoke a woman sitting on the ground in front of a kettle on a camp re. Figure 3.5 shows an illustration from the Eastern and Central Arrernte Picture Dictionary [Broad, 2008] of the sentenceIpmenhe-ipmenhele kwatye urinpe-ilemele itemeIMy grandmother is boiling the water." The lexical semantics for the English verbboil and the Arrente verb urinpe-ileme are the same, the relationapply-heat.boil . However, the vignettes map to di erent, culture-typical graphical semantics. Figure 3.6 shows the instantiated vignettes for our example, demonstrating the cultural di erences in the graphical semantics. To handle cultural differences like these, VigNet needs to be extended with new graphical semantics for existing vignettes that need to be modi ed, and new vignettes for scenarios not already covered. Currently, these modi cations to VigNet must be done by the WordsEye developers. Future

Figure 3.5: Illustration of Arrernte sentence Ipmenhe-ipmenhele kwatye urinpe-ilemele iteme, \My grandmother is boiling the water." [Broad, 2008]

work on WELT would allow users to make these changes themselves.

3.6.1.1 Custom WordsEye Objects

Another way to adapt WordsEye to a culture or region is to add relevant 3D objects to the database. WordsEye also supports 2D-cutout images, which is an easy way to add new material without 3D modeling. In the WordsEye desktop application, each new object has to be manually incorporated into WordsEye by the WordsEye developers. In the newer web application version of WordsEye, users can upload custom content directly. In Chapter 5, we will discuss the newer version of WordsEye in more detail, including its application to the WELT elicitation tools.

We have created a corpus of 2D and 3D models for WordsEye that are speci cally relevant to aboriginal speakers of Arrernte, including native Australian plants and animals and culturally relevant objects and gestures. Many of the pictures we created are based on images from IAD Press, used with permission. Most of the original IAD Press images were black-and-white drawings similar to the one in Figure 3.5. We enhanced and cropped these in PhotoShop. Examples of our enhanced images are shown in Table 3.1; the complete collection of enhanced images is provided in Appendix A. Some scenes that use these images are included in Figure 3.7.

Figure 3.6: (Instantiated) vignettes for the woman boils the water. The high-level semantics of apply-heat.boil are decomposed into sets of objects and primitive graphical relations that depend on cultural context.

3.6.2 Preparing Scenes and Eliciting Data: Nahuatl Topological Relations

To evaluate WordsEye's usefulness in the creation of pictures for eliciting spatial language, we created a set of scenes based on the Max Planck topological relations picture series [Bowerman and Pederson, 1992]. Some examples of pictures from the Max Planck topological relation series are shown in Table 3.2. Many of the scenes were created in WELT by two undergraduate students who were pursuing majors in linguistics. These students did not have any background or experience in computational linguistics or computer science, thus demonstrating WELT's usability by non-computational linguists. In creating the scenes, we used the new feature of WordsEye described in Section 3.4.1, that allows users to highlight

(a) Honey Ant (yerrampe) (b) Dingo (akngwelye artnwere) (c) Spinifex (aywerte) (d) Skink (ikwarre)

Table 3.1: Images created for Arrernte WELT (Arrernte translations in parentheses)

Figure 3.7: WordsEye scenes using custom 2D gum tree and dingo from our image corpus

speci c objects (or parts of objects) in a scene. We were able to recreate 40 out of the 71 pictures in the series. One of the main issues that prevented us from creating the full set was that WordsEye does not currently have the objects needed to produce the desired scene. There were also cases where the graphical functionality of WordsEye needs to be enhanced to allow more precise positioning of objects. We used these scenes to elicit descriptions from a native speaker of Nahuatl; some examples of scenes and descriptions are included in Table 3.3. Section B.1 shows the full set of scenes we created along with the input text used to generate the scenes in WordsEye. For scenes that were not reproduced in WordsEye, we also brie y describe the reasons we were unable to do so.[4] Section B.2 shows the Nahuatl descriptions we elicited for each scene.

[4] Note that many of these problems, particularly those involving missing 3D objects, will be xed when WELT has been updated to use the latest version of WordsEye. For example, WordsEye now has a number of 3D objects for bugs, which we were missing in pictures (7) and (52), as well as a 3D object of an open bag with handles, which was needed for pictures (14) and (66). Other issues will be addressed as WordsEye developers continue to add new 3D objects and implement more graphical functionality.

(a) (b) (c) (d)

Table 3.2: Pictures from the Max Planck Topological Relations Picture Series [Bowerman and Pederson, 1992]

(a)	in	tapameti	tìatsakwa	se	kali	(b)	in	amatì	tìakentija	se	kutSara
	the	fence	around	the	house		the	paper	cover	one	spoon

(c)	in	kwawitì	tìapanawi	tìakoja	se	mansana	(d)	in	tsopelik	katsekotok	tìatsintìa	in	tìapitS
	the	stick	pass.thru	in.middle	one	apple		the	candy	sticking	under	the	table

Table 3.3: Nahuatl elicitations

3.7 Documentation Tools (WELT L2)

WELT L2 provides the tools to formally document the semantics of a language. It also uses this documentation to automatically generate a text-to-scene system for the language. The formal documentation allows precise description of the lexical semantics of a language.

Because WELT is centered around the idea of 3D scenes, the formal documentation will tend to focus on the parts of the semantics that can be represented graphically. Note that this can include gurative concepts as well, although the visual representation of these may be culture-speci c. However, users do not necessarily need to be limited by the graphical output; WELT can be used to document other aspects of semantics as well, but it will not be possible to verify these theories using the text-to-scene system. In this section, we will describe the user interface for documenting semantics, as well as a text-to-scene system for Arrernte created with WELT.

In order to create a text-to-scene system for an endangered language, WELT requires the components shown in Figure 3.2. The custom vignettes and 3D objects will largely have been done during the cultural adaptation of VigNet described in Section 3.6.1. In addition to these, the system requires a morphological analyzer, syntactic parser, a lexicon that maps endangered language words into vignette concepts, and a set of syntax-to-semantics rules which map the output of the syntactic parser into vignettes. We discussed our development of morphological and syntactic grammars for Arrernte in Section 3.5. In this section, we will discuss creating the lexicon and the syntax-to-semantics rules.

3.7.1 The Lexicon and Ontology Browser

The lexicon in WELT is a list of word forms mapped to semantic concepts, which allows nouns from the endangered language to be converted into graphical objects. WELT includes a visual interface for searching VigNet's ontology for semantic concepts and browsing through the hierarchy to select a particular category. The ontology browser is used in several parts of WELT, including the tools for creating the lexicon and and the tools for modifying the constraints in syntax to semantics rules. We have created a mapping for the lexical items in the Arrernte grammar; a partial mapping is shown in Table 3.4. As an example, to nd an appropriate concept for the Arrernte word panikane `cup', we can search the ontology browser for cup. Figure 3.8(a) shows the portion of the ontology that results from this search. Semantic categories are displayed one level at a time, so initially only the concepts directly above and below the search term are shown. Selecting another visible

Lexical Item	VigNet Concept
artwe	person.n
panikane	cup.n
angepe	crow.n
akngwelye	dog.n
apwerte	rock-item.n
tipwele	table.n

Table 3.4: A mapping from nouns (lexical items) to VigNet semantic concepts

node in the graph results in its own immediate neighbors being displayed. Figure 3.8(b) shows the result of selecting drinking-cup.n . Here, we have decided to map panikane to cup.n .

3.7.2 Documenting Semantics

The goal of WELT L2 is to provide the means to formally document the semantics of a language and create a text-to-scene system for that language. The formal documentation allows precise description of the lexical semantics of a language. The WELT L2 semantics is represented using VigNet, which has been developed for WordsEye based on English. To use the WordsEye architecture, the system needs to be able to map between the syntax of the endangered language and a representation of semantics compatible with VigNet. Most obviously, the lexical items and valence patterns are di erent for other languages. One instance showing why this is necessary occurs in our example Arrernte sentence. When discussing football in English, one would say that someone kicks a goal or makes a goal In Arrernte, one would say goal arrerneme, which is a light verb construction that translates literally to \put a goal." Although the semantics of both sentences are the same, the entry for put in the English VigNet does not include this meaning, but the Arrernte text-to-scene system needs to account for it. Another issue is the motion-related in ections that can be applied to verbs in both Arrernte and Nahuatl; the valence patterns used by the WELT L2

(a) (b)

Figure 3.8: Screenshots of the ontology browser

text-to-scene system need to handle these in ectional variations, which are not present in English.

To address such instances, the linguist uses WELT L2 to de ne lexical, syntactic, and semantic information. This begins with the creation of a mapping from the lexicon into VigNet concepts, so that the noun strings in the endangered language can be converted to graphical objects. We discussed creating the lexicon in Section 3.7.1. To handle valence patterns, WELT includes an interface for the linguist to specify a set of rules that map from syntax to (lexical) semantics. The interface allows users to develop rules that map the lexical structure of the new language into a high-level semantic representation compatible with VigNet. Since we are modeling Arrernte syntax with LFG, the rules currently take syntactic f-structures as input, but the system could easily be modi ed to accommodate other formalisms. Rules are speci ed by de ning a tree structure for the left-hand (syntax) side and a directed acyclic graph for the right-hand (semantics) side. The left-hand side of a rule consists of a set of conditions on the f-structure elements and the right-hand side is the desired semantic structure.

As an example, we construct a rule to process sentence (5):

(a) (b)

Figure 3.9: (a) c-structure and (b) f-structure for artwe le goal arrerneme

(5) artwe le goal arrerneme
man erg goal put.npp

`The man kicks a goal.'

We begin by creating the left-hand (syntax) side of the rule, by specifying a tree structure for the syntactic constraints. Screenshots demonstrating this process in WELT are shown in Figure 3.10. For this sentence, our Arrernte grammar produces the f-structure in Figure 3.9(b). We could create a rule specifying the speci c lexical itemartwe in the subject position; to generalize somewhat, we create a rule that selects for predicate arrerne with object goal and any subject. First, we specify the syntactic predicate, choosing arrerne from the dropdown menu. Next, we add nodes for each of the syntactic arguments that will be de ned in our rule. These can be speci c lexical items or they can be variables which will be replaced with lexical items when an actual sentence is processed. We then connect these nodes with edges by specifying the start and end nodes and the syntactic grammatical relation between them. The values in the dropdown menus in Figure 3.10 are extracted from the XLE source les for the Arrernte (LFG) grammar.

After we have nished specifying the syntax, we create the right-hand (semantics) side of the rule. Screenshots demonstrating creating the right-hand side of our rule in WELT are shown in Figure 3.11. We begin by searching VigNet for a relevant vignette. In this case, we search forkick and select the vignette we.kick-goal.vg WELT then displays the default semantic constraints for the arguments vignette, which we can edit. We rst modify the agent argument of the vignette, to specify that var-1 should be mapped to this semantic

Figure 3.10: Creating the left-hand (syntax) side of a syntax-to-semantics rule in WELT: selecting a predicate, adding nodes to the syntax tree, and adding edges to specify syntactic arguments.

role. We can also add cultural and geographic customizations at this stage. First, we change the vignette to re ect Australian football. We specify that the target should be footy-goalposts.n and the projectile should be footy-ball.n . We also modify the vignette to re ect the geography of the area, changing background-texture from city.n to butte.n and global-ground-texture from grass-field-substance.n to desert-dirt.n . The values we modi ed are highlighted in Figure 3.11. Custom semantic categories are selected for the vignette arguments using the same ontology browser we used to create the lexicon.

The completed rule is shown in Figure 3.12. Note that var-1 on the left-hand side becomes VigNet(var-1) on the right-hand side; this indicates that the lexical item found in the subject position of the input should be mapped into a semantic concept using the L2 lexicon.

3.7.3 WELT L2 Text-to-Scene Generation

As we discussed in Section 3.4.1, WELT L2 includes a modi ed WordsEye user interface which can generate a 3D scene from endangered language input. Having created a basic Arrernte lexicon and a simple syntax-to-semantics rule, we have the necessary components to use the modi ed WordsEye pipeline. We will now walk through the modules of this pipeline, using sentence (5) from the previous section as our input text, reproduced for convenience here:

(5) artwe le goal arrerneme
man erg goal put.npp

`The man kicks a goal.'

Morphology: The rst step of processing the sentence is to run each word through the morphological analyzer. For our Arrernte project, we process the morphology of the example sentence by running each word through the morphological analyzer in XFST. This transforms the verb arrerneme into `arrerne+NONPAST.' The other tokens in the sentence remain unchanged.

Figure 3.11: Creating the right-hand (semantics) side of a syntax-to-semantics rule in WELT: searching for and selecting a semantic relation (vignette); editing the relation arguments.

Figure 3.12: A syntax-to-semantics rule created in WELT. Syntactic constraints are represented by a tree-structure on the left side of the rule; semantics are represented by the vignette on the right-hand side.

Syntax: The next step is syntactic parsing via our Arrernte grammar, using XLE. Since our LFG grammar is ambiguous, XLE produces several possible f-structures. WELT L2 displays all the possibilities to the user, so that the correct one can be selected. In XLE, ambiguities in the f-structure are indicated by labeling constraints with tags that indicate which contexts the constraints are de ned in. These tags are of the form < a> vs. < a>, < b> vs. < b>, < c> vs. < c>, and so on [Crouch et al., 2011, \Printing Charts"]. WELT displays these tags as pre xes on the node label, and further distinguishes the possible contexts by displaying them in di erent colors. Figure 3.13 shows what this looks like for our sentence. In this case, the red < a> context is the correct one, because artwe should be the subject of our sentence. Selecting this context results in the c-structure and f-structure shown in Figure 3.9. The f-structure is passed on to the semantics module.

Semantics: We now walk through the semantic processing of the sentence, assuming a set of rules consisting solely of the one given in Figure 3.12 and the partial noun mapping from Table 3.4 as our lexicon. The f-structure in Figure 3.9(b) has main predicate arrerne with two arguments; the object is goal. Therefore, it matches the left-hand-side of our rule. The output of the rule speci es the vignette we.kick-goal.vg . To determine the agent, we need to nd the VigNet concept corresponding to var-1 , which occupies the subject position in the f-structure. The subject in our f-structure is artwe, and according to Table 3.4, it maps

Figure 3.13: Selecting among possible f-structures produced by XLE from an ambiguous grammar.

to the VigNet concept person.n . The resulting semantic representation is augmented with the rest of its graphical semantics, taken from the vignette de nition. The WordsEye system then builds the scene from these constraints and renders it in 3D. Screenshots of the WELT L2 WordsEye interface for generating scenes from Arrernte text are shown in Figure 3.14. The top shows the scene generated from our example sentence. The two screenshots on the bottom demonstrate what happens when we substitute other lexical items for artwe in the subject position. On the left, panikane `cup' is not an animate-being.n , so the semantic constraints of the vignette are not met. On the right, akngwelye `dog' is an animate-being.n , so WELT L2 is able to generate the scene.

3.8 Future Work: Investigation of Case in Arrernte

One of our goals for using WELT is to study the relationship between the meaning of a sentence and the case of the nouns in it. The relationship in Arrernte between case and semantic interpretation of a sentence is a topic that could be easily explored with WELT. It is possible to signi cantly alter a sentence's meaning by changing the case on an argument. For example, the sentences in (6) from Wilkins [1989] show that adding dative case to the direct object of the sentence changes the meaning from shooting at and subsequently hitting the kangaroo, to shooting at the kangaroo and not hitting it. Wilkins calls this the \dative of attempt."