

Documenting Endangered Languages with the WordsEye Linguistics Tool

Morgan Ulinski*
mulinski@cs.columbia.edu

Anusha Balakrishnan*
ab3596@columbia.edu

Daniel Bauer*
bauer@cs.columbia.edu

Bob Coyne*
coyne@cs.columbia.edu

Julia Hirschberg*
julia@cs.columbia.edu

Owen Rambow†
rambow@ccls.columbia.edu

*Department of Computer Science
Columbia University
New York, NY, USA

†CCLS

Abstract

In this paper, we describe how field linguists can use the WordsEye Linguistics Tool (WELT) to study endangered languages. WELT is a tool under development for eliciting endangered language data and formally documenting a language, based on WordsEye (Coyne and Sproat, 2001), a text-to-scene generation tool that produces 3D scenes from text input. First, a linguist uses WELT to create elicitation materials and collect language data. Next, he or she uses WELT to formally document the language. Finally, the formal models are used to create a text-to-scene system that takes input in the endangered language and generates a picture representing its meaning.

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 (Alliance for Linguistic Diversity, 2013). Efforts to document languages and develop tools to support these efforts become even more important with the increasing rate of extinction. Bird (2009) emphasizes a particular need to make use of computational linguistics during fieldwork.

To address this issue, we are developing the WordsEye Linguistics Tool, WELT. In one mode of operation, we provide field linguists with tools for building elicitation sessions based on custom 3D scenes. In another, we provide a way to formally document the endangered language. Formal hypotheses can be verified 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.

WELT provides important advantages to field linguists for elicitation over the current practice of using a set of pre-fabricated static pictures. Using WELT the linguist can create and modify scenes in real time, based on informants' responses, creating follow-up questions and scenes to support them. Since the pictures WELT supports are 3D scenes, the viewpoint can easily be changed, allowing exploration of linguistic descriptions based on different frames of reference, as for elicitations of spatial descriptions. Finally, since scenes and objects can easily be added in the field, the linguist can customize the images used for elicitation to be maximally relevant to the current informants.

Creating a text-to-scene system for an endangered language with WELT also has advantages. First, WELT allows documentation of the semantics of a language in a formal way. Linguists can customize the focus of their studies to be as deep or shallow as they wish; however, we believe that a major advantage of documenting a language with WELT is that it enables studies that are much more precise. The fact that a text-to-scene system is created from this documentation will allow linguists to test the theories they develop with native speakers, making changes to grammars and semantics in real time. The resulting text-to-scene system can also be an important tool for language preservation, spreading interest in the language among younger generations of the community and recruiting new speakers.

In this paper, we discuss the WELT toolkit and its intended use, with examples from Arrernte and Nahuatl. In Section 2 we discuss prior work on field linguistics computational tools. In Section 3 we present an overview of the WELT system. We describe using WELT for elicitation in Section 4 and describe the tools for language documentation in Section 5. We conclude in Section 6.

2 Related Work

Computational tools for field linguistics fall into two categories: tools for native speakers to use directly, without substantial linguist intervention, and tools for field 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 simplified workflow 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 FieldWorks, 2014), or specifically, FieldWorks Language Explorer (FLEX). FLEX includes tools for eliciting and recording lexical information, dictionary development, interlinearization of texts, analysis of discourse features, and morphological analysis. An important part of FLEX is its “linguist-friendly” morphological parser (Black and Simons, 2006), 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.

Linguist’s Assistant (Beale, 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.

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. For semantics, the most common resource for formal documentation across languages is FrameNet (Filmore et al., 2003); FrameNets have been de-

veloped for many languages, including Spanish, Japanese, and Portuguese. However, FrameNet is also targeted toward computational linguists.

In general, we also lack tools for creating custom elicitation materials. With WELT, we hope to fill some of the gaps in the range of available field 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 specific 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 Overview of WELT Workflow

In this section, we briefly describe the workflow for using WELT; a visual representation is provided in Figure 1. Since we are still in the early stages of our project, this workflow has not been tested in practice. The tools for scene creation and elicitation are currently useable, although more features will be added in the future. The tools for modeling and documentation are still in development; although some functionality has been implemented, we are still testing it with toy grammars.

First, WELT will be used to prepare a set of 3D scenes to be used to elicit targeted descriptions or narratives. An important part of this phase will be 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 4.1. Next, the linguist will work 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 also automatically syncs recorded audio with open scenes and provides an interface for the linguist to write notes, textual descriptions, and glosses. We will discuss creating scenes and eliciting data with WELT in Section 4.2. After the elicitation session, the linguist can use WELT to review the data collected, listen to the audio recorded 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.

Creating a text-to-scene system with WELT requires formal models of the morphology, syntax,

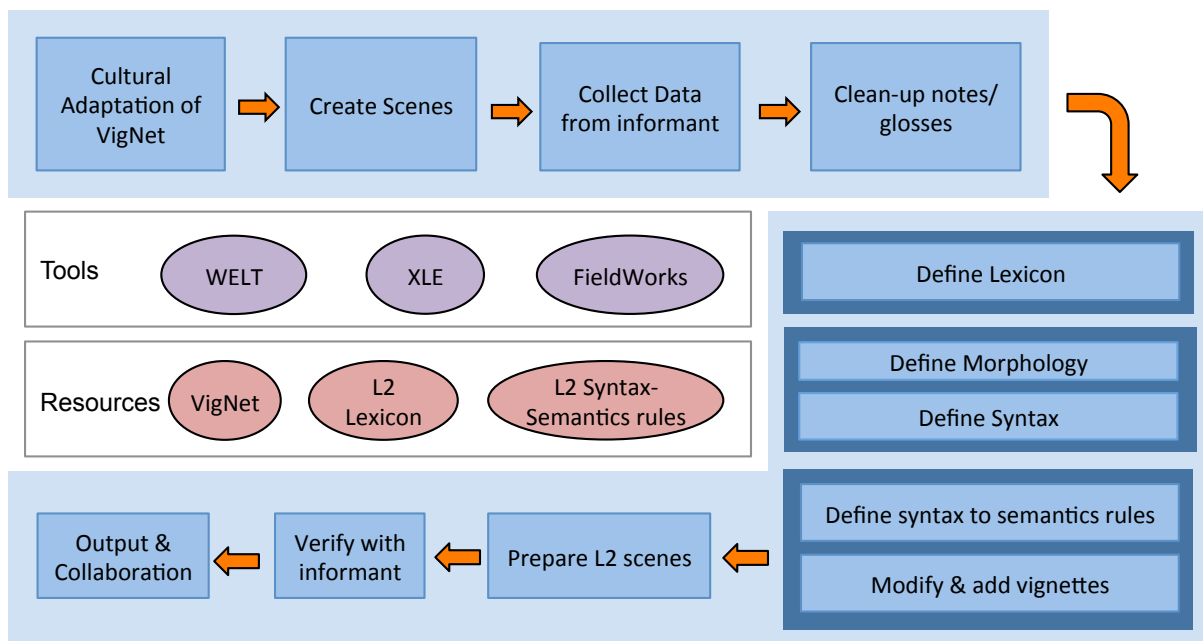


Figure 1: WELT workflow

and semantics of a language. Since the focus of WELT is on semantics, the formalisms used to model morphology and syntax may vary. We are using FieldWorks to document Nahuatl morphology, XFST (Beesley and Karttunen, 2003) to model Arrernte morphology, and XLE (Crouch et al., 2011) to model syntax in the LFG formalism (Kaplan and Bresnan, 1982). We will provide tools to export WELT descriptions and glosses into FLEx format and to export the lexicon created during documentation into FLEx and XLE. WELT will provide user interfaces for modeling the syntax-semantics interface, lexical semantics, and graphical semantics of a language. We will discuss these in more detail in Section 5.3.

Once models of morphology, syntax, and semantics are in place (note that these can be working models, and need not be complete), WELT 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 be used to verify theories with informants and revise grammars. As new questions arise, WELT can also continue to be used to create elicitation materials and collect linguistic data.

Finally, we will create a website for WELT so linguists can share resources such as modified versions of VigNet, 3D scenes, language data collected, and formal grammars. This will allow comparison of analyses across languages, as well as facilitate the documentation of other languages that are similar linguistically or spoken by cul-

turally similar communities. In addition, sharing the resulting text-to-scene systems with a wider audience can generate interest in endangered languages and, if shared with endangered-language-speaking communities, encourage younger members of the community to use the language.

4 Elicitation with WELT

WELT organizes elicitation sessions around a set of 3D scenes, which are created by inputting English text into WordsEye. Scenes can be imported and exported between sessions, so that useful scenes can be reused and data compared. WELT also provides tools for recording audio (which is automatically synced with open scenes), textual descriptions, glosses, and notes during a session. Screenshots are included in Figure 2.

4.1 Cultural Adaptation of VigNet

To interpret input text, WordsEye uses VigNet (Coyne et al., 2011), a lexical resource based on FrameNet (Baker et al., 1998). As in FrameNet, lexical items are grouped in frames according to shared semantic structure. A frame contains a set of frame elements (semantic roles). FrameNet defines the mapping between syntax and semantics for a lexical item with valence patterns that map syntactic functions to frame elements.

VigNet extends FrameNet in order to capture “graphical semantics”, a set of graphical constraints representing the position, orientation, size, color, texture, and poses of objects in the scene,

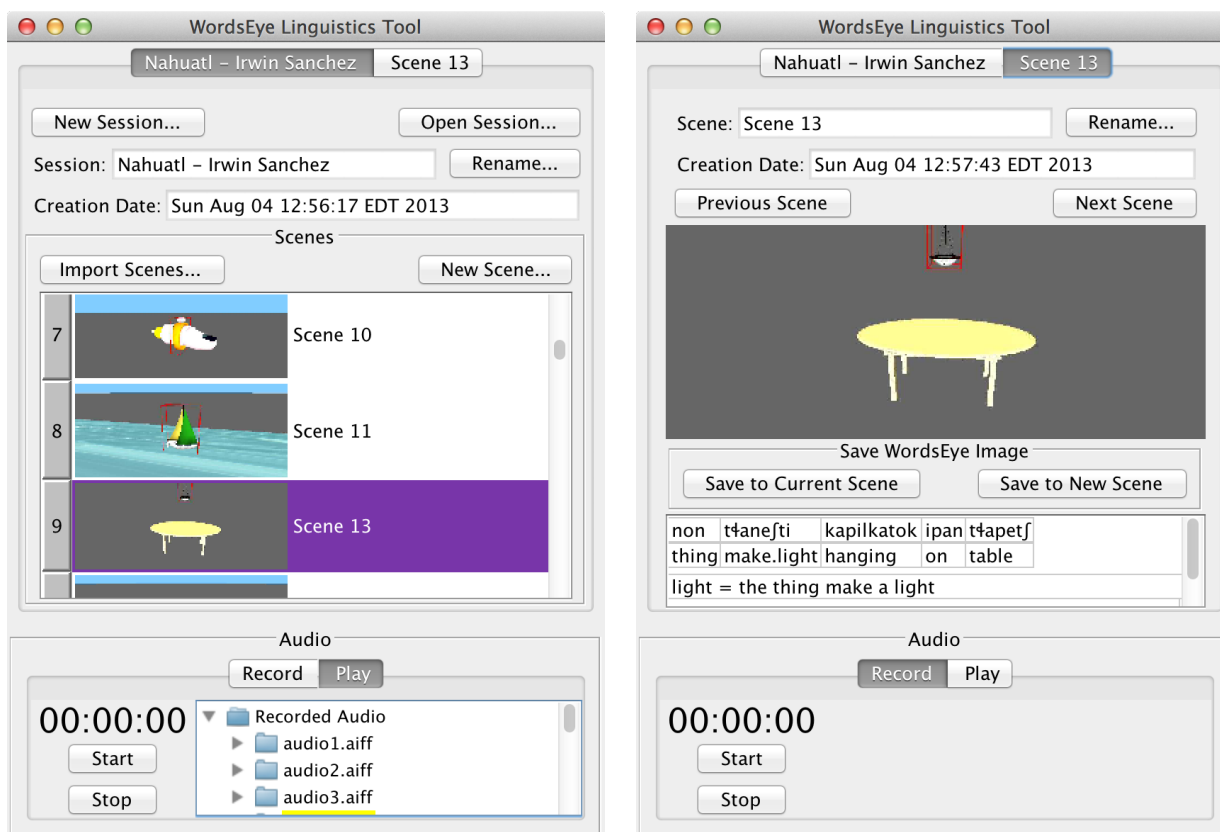


Figure 2: Screenshots of WELT elicitation interfaces

which is used to construct and render a 3D scene. Graphical semantics are added to frames by adding primitive graphical (typically, spatial) relations between frame element fillers. VigNet distinguishes between meanings of words that are distinguished graphically. For example, the specific objects (e.g., implements) 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*.

Vignette Tailoring: Without digressing into a discussion on linguistic relativity, we assume that 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 will need to be added for a culture. In the U.S., for example, the sentence *The woman boiled the water* might invoke a scene with a 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 campfire. Figure 3 shows an illustration from the Eastern and Central Arrernte Picture Dictionary (Broad, 2008) of the sentence *Ipmenhe-ipmenhele kwatye urinpe-ilemele iteme*, “My grandmother is boiling the water.” The lexical semantics for the English verb *boil* and the Arrernte verb *urinpe-ileme* are the same, the relation APPLY-HEAT.BOIL. However, the vignettes map to different, culture-typical graphical semantics. The vignettes for our example are shown in Figure 4.



Figure 3: Illustration from Broad (2008).

To handle cultural differences like these, a linguist will use WELT to extend VigNet with new

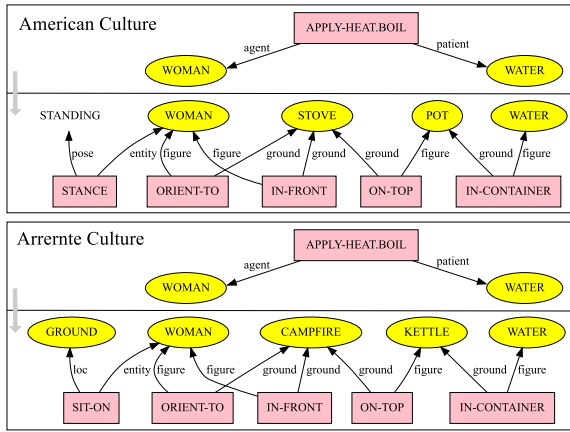


Figure 4: Vignettes for *the woman boils the water*. The high-level semantics of APPLY-HEAT.BOIL are decomposed into sets of objects and primitive graphical relations that depend on cultural context.

graphical semantics for existing vignettes that need to be modified, and new vignettes for scenarios not already covered. We will create interfaces so that VigNet can easily be adapted.

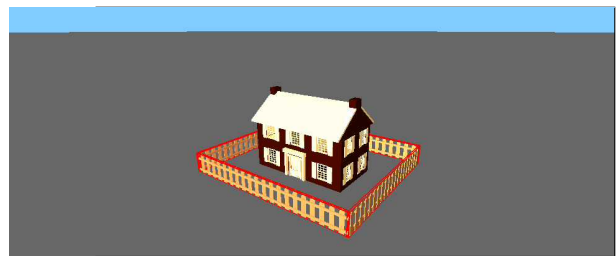
Custom WordsEye Objects: Another way to adapt WordsEye to a culture or region is to add relevant 3D objects to the database. WordsEye also supports 2D-cutout images, which is an easy way to add new material without 3D modeling. We have created a corpus of 2D and 3D models for WordsEye that are specifically 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, which we enhanced and cropped in PhotoShop. Some scenes that use these images are included in Figure 5. Currently, each new object has to be manually incorporated into WordsEye, but we will create tools to allow WELT users to easily add pictures and objects.

New objects will also need to be incorporated into the semantic ontology. VigNet’s 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 SPIDER.N and POISONOUS-ENTITY.N. Concepts are often linked to corresponding lexical items. If a lexical item has more than one word sense, the different word senses would be represented by different concepts. In addition, every graphical object in VigNet is represented by

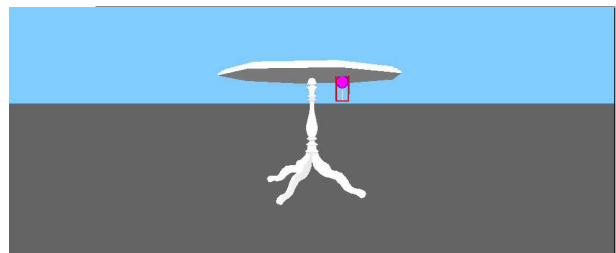
a unique concept. For example, a particular 3D model of a dog would be a linked to the general DOG.N concept by the ISA relation. The semantic concepts in VigNet include the graphical objects available in WordsEye as well as concepts tied to related lexical items. While WordsEye might only have a handful of graphical objects for dogs, VigNet will have concepts representing all common types of dogs, even if there is no graphical object associated with them. We will provide interfaces both for adding new objects and for modifying the semantic concepts in VigNet to reflect the differing lexical semantics of a new language.

4.2 Preparing Scenes and Eliciting Data

The next step in the workflow is the preparation of scenes and elicitation of descriptions. To test creating elicitation materials with WELT, we built a set of scenes based on the Max Planck topological relations picture series (Bowerman and Pederson, 1992). In creating these, we used a feature of WordsEye that allows highlighting specific objects (or parts of objects) in a scene. We used these scenes to elicit descriptions from a native Nahuatl speaker; some examples are included in Figure 6.



(a) in tapametł t̄atsakwa se kali
the fence/wall around the house



(b) in tsopelik katsekotok t̄atsint̄la in t̄apetf
the candy sticking under the table

Figure 6: Nahuatl examples elicited with WELT

One topic we will explore with WELT is the relationship in Arrernte between case and semantic interpretation of a sentence. It is possible to significantly alter a sentence’s meaning by changing the case on an argument. For example, the sentences in (1) from Wilkins (1989) show that adding dative

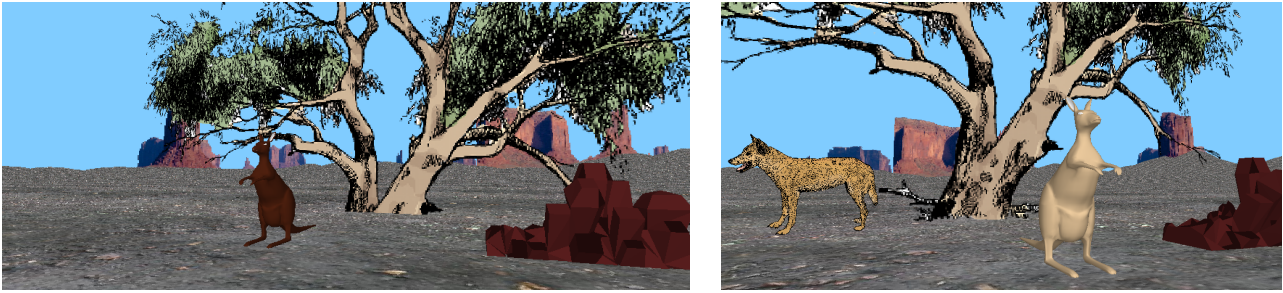


Figure 5: WordsEye scenes using custom 2D gum tree and dingo from our corpus

case to the direct object of the sentence changes the meaning from shooting and hitting the kangaroo to shooting *at* the kangaroo and *not* hitting it. Wilkins calls this the “dative of attempt.”

- (1) a. re aherre tyerre-ke
 he kangaroo shot-pc
 He shot the kangaroo.
- b. re aherre-ke tyerre-ke
 he kangaroo-DAT shot-pc
 He shot at the kangaroo (but missed).

In order to see how this example generalizes, we will create pairs of pictures, one in which the object of the sentence is acted upon, and one in which the object fails to be acted upon. Figure 7 shows a pair of scenes contrasting an Australian football player scoring a goal with a player aiming at the goal but missing the shot. Sentences (2) and (3) are two ways of saying “score a goal” in Arrernte; we want to see if a native Arrernte speaker would use *goal-ke* in place of *goal* in this context.

- (2) artwe le goal arrerne-me
 man ERG goal put-NP
 The man kicks a goal.
- (3) artwe le goal kick-eme-ile-ke
 man ERG goal kick-VF-TV-PST
 The man kicked a goal.

5 Modeling a Language with WELT

WELT includes tools for documenting the semantics of the language. It also uses this documentation to automatically generate a text-to-scene system for the 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 figurative concepts as well, although the visual representation of these may be culture-specific. However, linguists do not need

to be limited by the graphical output; WELT can be used to document other aspects of semantics as well, but linguists will not be able to verify these theories using the text-to-scene system.

To explain the necessary documentation, we briefly describe the underlying architecture of WordsEye, and how we are adapting it to support text-to-scene systems for other languages. The WordsEye system parses each input sentence into a labeled syntactic dependency structure, then converts it into a lexical-semantic structure using lexical valence patterns and other lexical and semantic information. The resulting set of semantic relations is converted to a “graphical semantics”, the knowledge needed to generate graphical scenes from language.

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. The WELT processing pipeline is illustrated in Figure 8, with stages of the pipeline on top and required resources below. In this section, we will discuss creating the lexicon, morphological and syntactic parsers, and syntax-to-semantics rules. The vignettes and 3D objects will largely have been done during cultural adaptation of ViGNet; additional modifications needed to handle the semantics can be defined using the same tools.

5.1 The Lexicon

The lexicon in WELT is a list of word forms mapped to semantic concepts. The process of building the lexicon begins during elicitation. WELT’s elicitation interface includes an option to display each object in the scene individually before progressing to the full scene. When an object is labeled and glossed in this way, the word and the semantic concept represented by the 3D object are immediately added to the lexicon. Word forms glossed in scene descriptions will also be added to the lexicon, but will need to be mapped to semantic concepts later. WELT will provide

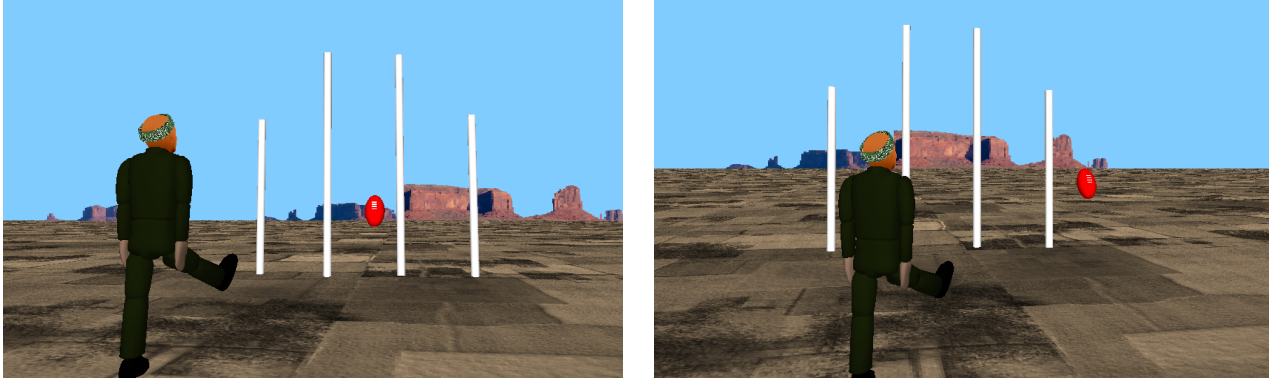


Figure 7: WordsEye scenes to elicit the “dative of attempt.”

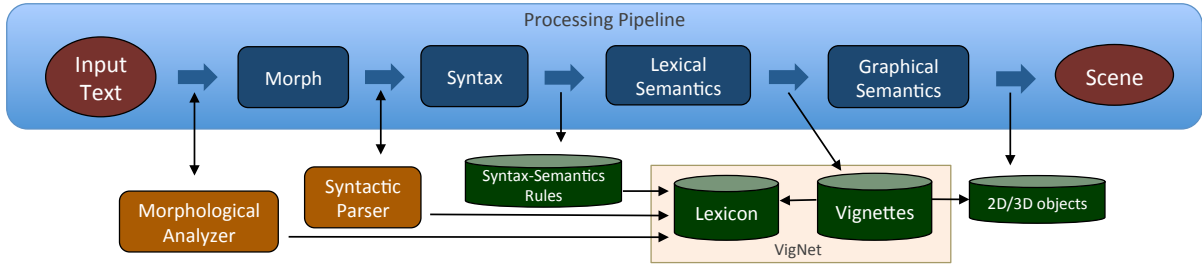


Figure 8: WELT architecture

tools for completing the lexicon by modifying the automatically-added items, adding new lexical items, and mapping each lexical item to a semantic concept in VigNet. Figure 9(a) shows a partial mapping of the nouns in our Arrernte lexicon.

WELT includes a visual interface for searching VigNet’s ontology for semantic concepts and browsing through the hierarchy to select a particular category. Figure 9(b) shows a portion of the ontology that results from searching for *cup*. Here, we have decided to map *panikane* to CUP.N. Semantic categories are displayed one level at a time, so initially only the concepts directly above and below the search term are shown. From there, it is simple to click on relevant concepts and navigate the graph to find an appropriate semantic category. To facilitate the modeling of morphology and syntax, WELT will also export the lexicon into formats compatible with FieldWorks and XLE, so the list of word forms can be used as a starting point.

5.2 Morphology and Syntax

As mentioned earlier, 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 morphology, a very good tool already exists in FLEx, which allows the creation

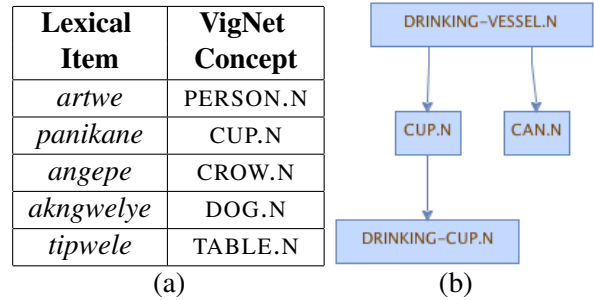


Figure 9: (a) Arrernte lexical items mapped to VigNet concepts; (b) part of the VigNet ontology

of a morphological parser without knowledge of any particular grammatical formalism. For syntax, we are using XLE for our own work while researching other options that would be more accessible to non-computational linguists. It is important to note, though, that the modeling done in WELT does not require a perfect syntactic parser. In fact, one can vastly over-generate syntax and still accurately model semantics. Therefore, the syntactic grammars provided as models do not need to be complex. However, the question of syntax is still an open area of research in our project.

5.3 Semantics

To use the WordsEye architecture, the system needs to be able to map between the formal syntax of the endangered language and a representation of semantics compatible with VigNet. To accomplish

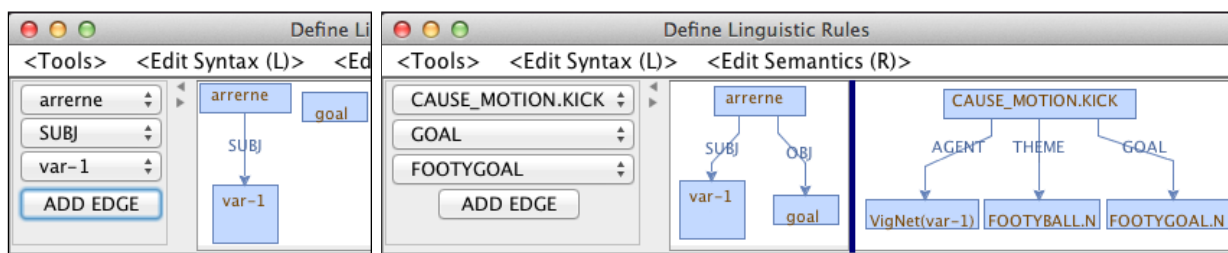


Figure 10: Creating syntax-semantics rules in WELT

this, WELT includes an interface for the linguist to specify a set of rules that map from syntax to (lexical) semantics. Since we are modeling Arrernte syntax with LFG, the rules currently take syntactic f-structures as input, but the system could easily be modified to accommodate other formalisms. 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. Rules are specified by defining a tree structure for the left-hand (syntax) side and a DAG for the right-hand (semantics) side.

As an example, we will construct a rule to process sentence (2) from Section 4.2, *artwe le goal arrerreme*. For this sentence, our Arrernte grammar produces the f-structure in Figure 11. We create a rule that selects for predicate *arrerne* with object *goal* and any subject. Figure 10 shows the construction of this rule in WELT. 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 input is mapped into a semantic concept using the lexicon.

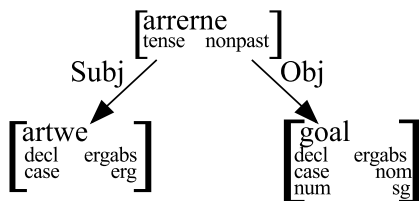


Figure 11: F-structure for sentence 2, Section 4.2.

The rule shown in Figure 10 is a very simple example. Nodes on the left-hand side of the rule can also contain boolean logic, if we wanted to allow the subject to be $[(artwe \text{ ‘man’ OR } arhele \text{ ‘woman’}) \text{ AND NOT } ampe \text{ ‘child’}]$. Rules need not specify lexical items directly but may refer to more general semantic categories. For example, our rule could require a particular semantic category for *VIGNET(var-1)*, such as *ANIMATE-BEING.N*. These categories are chosen through the same ontology browser used to create the lexicon. Finally, to ensure that our sen-

tence can be converted into graphics, we need to make sure that a vignette definition exists for *CAUSE_MOTION.KICK* so that the lexical semantics on the right-hand side of our rule can be augmented with graphical semantics; the vignette definition is given in Figure 12. The WordsEye system will use the graphical constraints in the vignette to build a scene and render it in 3D.

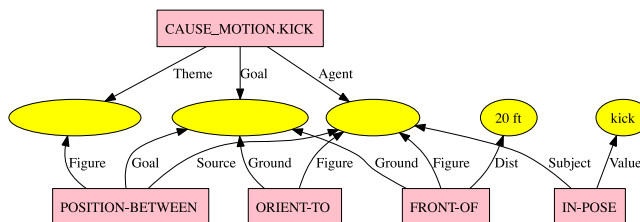


Figure 12: Vignette definition for *CAUSE_MOTION.KICK*

6 Summary

We have described a novel tool under development for linguists working with endangered languages. It will provide a new way to elicit data from informants, an interface for formally documenting the lexical semantics of a language, and allow the creation of a text-to-scene system for any language. In this paper, we have focused specifically on the workflow that a linguist would follow while studying an endangered language with WELT. WELT will provide useful tools for field linguistics and language documentation, from creating elicitation materials, to eliciting data, to formally documenting a language. In addition, the text-to-scene system that results from documenting an endangered language with WELT will be valuable for language preservation, generating interest in the wider world, as well as encouraging younger members of endangered language communities to use the language.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 1160700.

References

- Alliance for Linguistic Diversity. 2013. The Endangered Languages Project. <http://www.endangeredlanguages.com>.
- C. Baker, J. Fillmore, and J. Lowe. 1998. The Berkeley FrameNet project. In *36th Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics (COLING-ACL'98)*, pages 86–90, Montréal.
- Stephen Beale. 2011. Using Linguist's Assistant for Language Description and Translation. In *IJCNLP 2011 System Demonstrations*, pages 5–8.
- Kenneth R. Beesley and Lauri Karttunen. 2003. Finite-State Morphology Homepage. <http://www.fsmbook.com>.
- E. Bender, D. Flickinger, and S. Oepen. 2002. The Grammar Matrix. In J. Carroll, N. Oostdijk, and R. Sutcliffe, editors, *Workshop on Grammar Engineering and Evaluation at the 19th International Conference on Computational Linguistics*, pages 8–14, Taipei, Taiwan.
- S. Bird and D. Chiang. 2012. Machine translation for language preservation. In *COLING 2012: Posters*, pages 125–134, Mumbai, December.
- S. Bird. 2009. Natural language processing and linguistic fieldwork. *Computational Linguistics*, 35(3):469–474.
- Cheryl A Black and H Andrew Black. 2009. PAWS: Parser and Writer for Syntax. In *SIL Forum for Language Fieldwork 2009-002*.
- H.A. Black and G.F. Simons. 2006. The SIL FieldWorks Language Explorer approach to morphological parsing. In *Computational Linguistics for Less-studied Languages: Texas Linguistics Society 10*, Austin, TX, November.
- M. Bowerman and E. Pederson. 1992. Topological relations picture series. In S. Levinson, editor, *Space stimuli kit 1.2*, page 51, Nijmegen. Max Planck Institute for Psycholinguistics.
- N. Broad. 2008. *Eastern and Central Arrernte Picture Dictionary*. IAD Press.
- B. Coyne and R. Sproat. 2001. WordsEye: An automatic text-to-scene conversion system. In *SIGGRAPH*.
- B. Coyne, D. Bauer, and O. Rambow. 2011. Vignet: Grounding language in graphics using frame semantics. In *ACL Workshop on Relational Models of Semantics (RELMS)*, Portland, OR.
- D. Crouch, M. Dalrymple, R. Kaplan, T. King, J. Maxwell, and P. Newman. 2011. XLE Documentation. http://www2.parc.com/isl/groups/nlftt/xle/doc/xle_toc.html.
- C. Fillmore, C. Johnson, and M. Petruck. 2003. Background to FrameNet. In *International Journal of Lexicography*, pages 235–250.
- R.M. Kaplan and J.W. Bresnan. 1982. Lexical-functional grammar: A formal system for grammatical representation. In J.W. Bresnan, editor, *The Mental Representation of Grammatical Relations*. MIT Press, Cambridge, Mass., December.
- SIL FieldWorks. 2014. SIL FieldWorks. <http://fieldworks.sil.org>.
- D. Wilkins. 1989. *Mparntwe Arrernte (Aranda): Studies in the structure and semantics of grammar*. Ph.D. thesis, Australian National University.