Wordcraft: Playing with Sentence Structure

Divya Anand
University of California,
Berkeley, CA
divya.gayathri@gmail.com

Ashley DeSouza
University of California,
Berkeley, CA
ashley.desouza@live.com

Shreyas
University of California,
Berkeley, CA
seekshreyas@gmail.com

Kimiko Ryokai
University of California,
Berkeley, CA
kimiko@ischool.berkeley.edu

Sonali Sharma
University of California,
Berkeley, CA
sonali1718@gmail.com

Marti Hearst
University of California,
Berkeley, CA
hearst@berkeley.edu

Victor Starostenko
University of California,
Berkeley, CA
victor.starostenko@live.com

CHI PLAY 2015, October 03-07, 2015, London, United Kingdom
ACM 978-1-4503-3466-2/15/10.
http://dx.doi.org/10.1145/2793107.2810320

Abstract
We introduce *Wordcraft*, a new interactive tablet application that allows children to explore sentence structures and their meanings. *Wordcraft* uses a constructionist design: children manipulate word cards to build sentences, which come to life in a storybook-like animated world to illustrate meaning. Such visual feedback helps children play with parts of speech and understand how they fit together to form sentences. Preliminary studies suggest that children are able to observe and discuss how different sentence constructs result in different meanings.

Author Keywords
Design; education; games; learning; parts of speech

Introduction
Educational psychologists Neumann and Neumann (2014) found that children are increasingly using touchscreen interfaces, and the finger-operated tactile affordances of tablets may support preschoolers' literacy development. There also exists a need for research into how to scaffold language learning in this new design space.

Wordcraft allows children to build a sentence and get immediate visual feedback. This helps children understand how different parts of speech impact sentence construction. Figure 1 shows an example of a user having to insert a preposition to animate the

sentence, "A pig is hopping with a goat". Preliminary results of the use of *Wordcraft* showed that children discussed aspects of language, or metalinguistic discourse. Thus, *Wordcraft*'s design may improve a child's understanding of the relationship between syntax and meaning. We describe the design rationale and technology, user study, and potential future work.

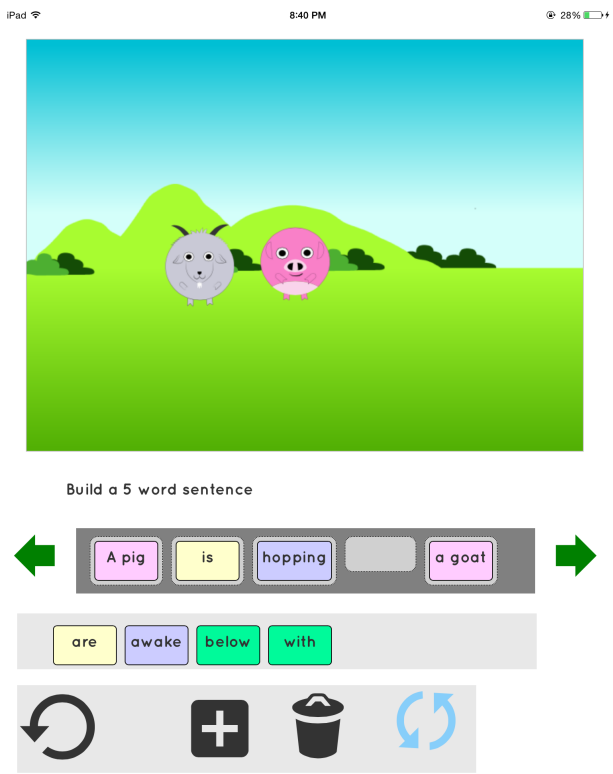

Figure 1: *Wordcraft* user interface showing the farm scene with 2 characters, partially formed sentence, word tray with additional words colored by part of speech, and tool bar. When a sentence is completed, the corresponding action is animated.

Related Work

Wordcraft is partly modeled on Scratch (Resnick et al., 2011), a visual programming language built on ideas from Lego toys (Resnick, 1993). Scratch is a modern descendant of LOGO, designed by Papert (1980) to support constructivist theories of the "child as an epistemologist" who builds knowledge by playing with programmable environments. The design of *Wordcraft* was also influenced by the DuoLingo learning application (von Ahn, 2013). One of Duolingo's games involves displaying a sentence in one language and a list of word cards in another language to be dragged and dropped onto a tray. *Wordcraft* borrows this card-based drag-and-drop interface for sentence formation, but greatly expands on it by displaying the meaning of the sentence in an animated scene on a stage.

There has been a recent increase in NLP (Natural Language Programming) research around generating textual descriptions of information within images (Kulkarni et al., 2013), or conversely, generating a scene illustrating a given text (Coyne and Sproat, 2001), or both (Farhadi et al., 2010). Some very recent work attempts to learn visual features that correspond to semantic phrases, derived from sentence components (Zitnick et al., 2014) and generate images. This work is still too preliminary to build on.

The novelty of *Wordcraft*'s design is in the interactions, the NLP, and the motivation of supporting kids' learning. Most other designs, as outlined by Kam et al. (2009) have either a UI or game play without a focus on a structured language learning component. Our approach of immediate visual feedback allows children

Rule 1 (3 word sentence)
<SN or PN> | be | Verb
A cat | is | walking
The pigs | are | dancing
Rule 2 (5 word sentence)
<SN or PN> | be | Verb | Prep | <SN or PN>
A goat | is | running | near | the fences
The cows | are | shivering | near | a barn
Rule 3 (7 word sentence)
Det | Adj | <SN or PN> |is | Verb | Prep | <SN or PN>
An | excited | dog | is | rolling | next to | the fences
A | fierce | cat | is | walking | near | a fence
The | huge | cow | is | wailing | on top of | the goats
The | happy | pigs | are | hopping | towards | a horse
Nouns: 8 | Verbs: 18| Adjs: 15 | Prep: 12 | Det: 3

Table 1: Wordcraft farm scene grammar and example sentences

to visualize how the use of words can impact sentence construction, thus encouraging metalinguistic learning.

Wordcraft Design and Implementation

The goal was to test the ideas before implementing a full-fledged tool. Therefore, the sentence structure and vocabulary are restricted to simplify prototyping. A farm scene was chosen because of its familiarity to children; similar scenarios should be substitutable. The subsections below describe the interface design, system architecture, NLP component, and visual framework for the animation component.

Interface Overview

The interface consists of 3 areas: the canvas, the sentence builder and the word tray. Sentences are built by dragging words from the word tray to the sentence builder, which has slots defining sentence length. Children are free to move between 3-word, 5-word and 7-word sentences, which correspond to game levels, allowing them to progress, as they're ready. This makes the tool appropriate for different age groups.

When a noun is placed onto a legitimate slot on the sentence builder, an image is rendered on the canvas, and updated as nouns are added or removed. This ensures immediate visual feedback. If an incorrect word is dragged onto a slot, it falls into the word tray to give feedback on grammatically correct sentences.

Words are color-coded to help children develop a model of how parts of speech work. To avoid overwhelming children, only a few words are visible; instead they can scroll through available words. The bottom of the screen shows a tool bar with 3 buttons – Replay (to replay animations), Delete (to remove words) and Refresh (to clear the sentence builder and canvas).

Software Components

Computations are handled client-side, as only a fraction of U.S. parents allow children aged 3-7 access to Internet-enabled mobile devices (Chiong and Shuler, 2010). HTML5 is used for scene rendering. To create an all-device deployable app, a custom grammar rule engine was written using PhoneGap, an open source toolkit that integrates web technologies portably.

Language Components

The NLP components include an English vocabulary drawn from the Dolch word list (Johns, 1970) list of valid sentence structures (Table 1), and a JSON-based data structure that maps objects to visual renderings. In Level 1, children make 3-word sentences in the noun-be-verb format. Level 2 consists of 5-word sentences in the noun-be-verb-preposition-noun format. Level 3 consists of 7-word sentences in the det-noun-adjective-be-verb-preposition-noun format. This simple grammar is in the interest of prototyping. All verbs are shown in continuous present tense to reflect the fact that the completion of a sentence causes an animation to occur in the farm scene.

Visual Components

Constructing the canvas involves three main tasks:

1) Object Rendering: Noun(s) are compound objects created from individually drawn parts (Figure 2).

2) Object Manipulation: Animations are based on verbs and prepositions applied to noun(s) (Figure 3).

3) Object Placement: Object location is determined by current location, prepositions, and other objects.

For the following sequence of moves:

1 'the horse'

2 'the cat'

3 'is'

4 'in front of'

5 'running'

Step 1: The horse image is placed in its default location, center_middle.

Step 2: The cat image appears, offset 2x positions to the left.

Step 3: The word 'is' causes no changes.

Step 4: The cat is moved into the left_back position behind the horse.

Step 5: The sentence is completed and the child is rewarded by seeing the corresponding animation run.

Table 2: Example rendering of a sentence.

Each word has an associated JSON data structure that contains information required to render and animate objects. The current scene is a 'farm', with a 'vanishing plane', a 'ground' plane and a 'sky' plane. Each plane has 9 grid positions, as seen in Figure 4.

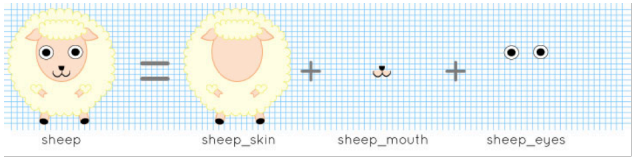

Figure 2: Composition of an object from constituent parts. Emotions are animated by changing the shape of the mouth and the eyes.

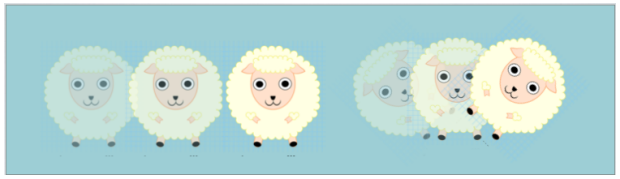

Figure 3: Animations for horizontal movement and rolling

Noun representations contain information about the image for the object's skin (sheep_skin in Figure 2), whether there are one or two images (for singular and plural forms), images for the current mouth and eyes (happy, angry, etc.), dimensions and canvas position of the object and the object's spatial relationship to other objects on the canvas.

Verb representations contain animation instructions: duration, speed (normal, slow, etc.), scale, and animation type (rotate, translateX, translateY). Each verb is restricted to only taking on certain prepositions.

Adjectives' representations indicate body size (normal, small, large), and eye and mouth properties that noun objects can take on.

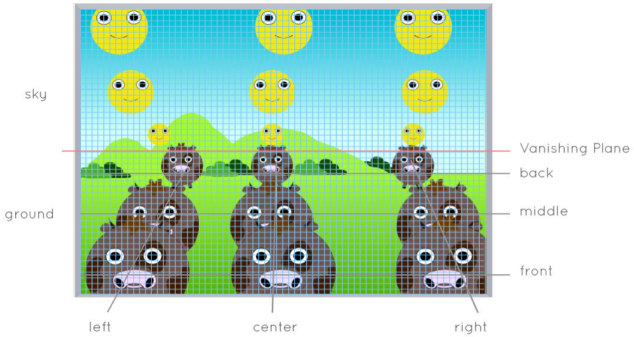

Figure 4: Canvas grid mapping is a vanishing plane separating ground and sky. Within each, 9 grid positions on the x-axis (left, center, right) are co-linear with those on the y-axis (front, middle, back) at increasingly greater distance.

Sentence Processing Algorithm

When a word is moved to its legitimate slot, a data structure containing the word and its properties is augmented as words get added or removed. When a user moves a noun to the sentence area, the object appears in its 'default' state: a happy mouth and happy eyes overlaid on the default skin, placed in the default location. For a plural noun, two copies of the object are rendered slightly displaced from one another. Verbs control the animation as well as change the appearance features. Adjectives change the appearance of nouns.

The verb data structure contains instructions for the prepositions, indicating where to move the subject of the verb relative to the object of the preposition. For example, the verb 'hopping' and the preposition 'with'

Figure 5: Final design of Wordcraft, with a 'five word' design though it contains seven actual words.

used in 'The cat is hopping with the pig.' means that both hop together. By contrast, 'The cat is hopping near the pig.' means that only the cat will hop.

The grammar rules engine is a set of if-else conditions built in JavaScript. Every slot on the sentence builder is mapped to the part of speech it can accept. Dragging words into slots causes the assignment rules to adjust. Table 2 outlines an example of how the interface works when constructing the sentence, 'the cat is running in front of the horse'.

Assessment

We evaluated the app by observing 17 children aged 4-8, across 7 locations over a period of 3 weeks. 7 children played alone with the version in Figure 1; the design was subsequently improved and 10 children played in pairs with the version in Figure 5. Children were asked to read the words (2 nouns and 3 verbs) that appeared when the app loaded. We observed a wide variation in reading ability and thus classified the children into groups based on reading ability, instead of an age-based classification: Beginners (needed help to read), Intermediates (needed help with complex words) and Advanced (could read).

Observations

Beginners

Beginners were most comfortable with 3-word sentences. Playing with a friend motivated them to try 5 and 7-word sentences. The interface helped them understand how alphabets correspond to words. For instance, Sara identified 'p i g' was "pig" after seeing the image. Beginners, who were unaware of parts of speech, were able to identify patterns in the color-coded words. Drew said, "In 3 words [sentences], one of each color makes a sentence".

Intermediates

Intermediates were interested in trying 5 and 7-word sentences, as they found the first level to be "too easy". They incorporated conflicting emotions, like, "the fierce goat is smiling near a fence", and rationalized perceived inconsistencies in the resulting image basis prior understanding. Serena explained, "How can a fierce goat smile? Fierce means it's angry." They also made 'funny' sentences where inanimate objects moved, talked etc. Many intermediates were unaware of the parts of speech, but they could verbalize what the colors meant. Responses ranged from "purple words are actions", to "purple words are verbs".

Advanced

Advanced children also preferred 5 and 7-word sentences. Even though they could read, they were not always aware of meanings. Thus, Ella used 'wailing' in a sentence to discover it means 'crying'. They attempted to make sentences that the app did not support, for instance, using words like 'weeping' (tagged as verbs in the interface) in their adjective forms, and pointed out the mistakes in the interface. They avoided 'funny' sentences, demonstrating an understanding of logic, in addition to grammar. Future work will explore if playing with *Wordcraft* helps beginners and intermediates move to being advanced.

Conclusions and Future Work

Wordcraft is a novel game application that allows children to explore language structure. Initial results suggest this approach, which transfers some of the benefits of children's block toys and takes advantage of

the affordances of software tools, allows children to observe how different sentence constructs result in different meanings. For instance, as they played with the interface, children were able to talk about the difference between "the cow is hopping with the goat" and "the cow is hopping away from the goat."

Children at different reading levels played differently with *Wordcraft*. Beginners experimented with changing one word at a time to see the image change. Intermediates explained why things worked the way they did. Advanced were able to spot inconsistencies. The storybook format encouraged children to talk about what they were seeing. Children also discussed the impacts of using different words, thus engaging in metalinguistic discourses. *Wordcraft*'s visual approach appears to foster language skills by allowing children to discuss the implications of word usage and sentence construction.

Future work should investigate more rigorously if children are able to make the connection between what they observe about the structure and meaning of language in this application to formal notions of syntax, parts of speech, writing and other concepts in school curriculum.

Acknowledgements

We gratefully thank the children and their parents who participated in the studies reported here.

References

1. Cynthia Chiong and Carly Shuler. 2010. Learning: Is there an app for that. Investigations of young children's usage and learning with mobile devices and apps: The Joan Ganz Cooney Center at Sesame Workshop.
2. Bob Coyne and Richard Sproat. 2001. WordsEye: an automatic text-to-scene conversion system. *ACM SIGGRAPH '01*.
3. Ali Farhadi, Mohsen Hejrati, Mohammad Amin Sadeghi, Peter Young, Cyrus Rashtchian, Julia Hockenmaier, and David Forsyth. 2010. Every picture tells a story: Generating sentences from images. *EECV '10*, 15-29.
4. Jerry L. Johns. 1970. The Dolch Basic Word List-Then and Now. *Journal of Lit Res* 3(4), 35-40.
5. Girish Kulkarni, Visruth Premraj, Vicente Ordonez, Sagnik Dhar, Siming Li, Yejin Choi, Alexander C. Berg, Tamara Berg. 2013. Babytalk: Understanding and generating simple image descriptions. *TPAMI '13*, 35(12), 2891-2903.
6. Michelle M. Neumann and David L. Neumann. 2014. Touch screen tablets and emergent literacy. *Early Childhood Edu Journal* 42(4), 231-239.
7. Seymour Papert. 1980. Mindstorms: Children, computers, and powerful ideas. Basic Books, Inc.
8. Mitchel Resnick, John Maloney, Andrés Monroy-Hernández, Natalie Rusk, Evelyn Eastmond, Karen Brennan, et al. 2009. Scratch: programming for all. *Communications of the ACM* 52, 11: 60-67.
9. Swapna Somasundaran and Martin Chodorow. 2014. Automated Measures of Specific Vocabulary Knowledge from Constructed Responses ("Use These Words to Write a Sentence Based on this Picture"). *9th NAACL Workshop on Innovative Use of NLP for Building Educational Applications*, 1-11.
10. Luis von Ahn. 2013. Duolingo: learn a language for free while helping to translate the web. *IUI 2013*, 1-2.
11. Lawrence C. Zitnick, Devi Parikh and Lucy Vanderwende. 2013. Learning the visual interpretation of sentences. *ICCV 2013,* 1681-1688.