

MCDS INDPNT STUDY

11633 B
Meaning Bee

Gina Huang
Industrial Design, 2016

Weikun Liang
Information Systems & HCI, 2016

Advisors:
Anthony Tomasic
Alex Rudnicky

Problem

Origin Story

Anthony was at home quizzing his daughter about the week's spelling words. She was getting them all correct. Then Anthony asked what a certain word meant, and she responded "I was asked to learn to spell these words." Anthony thought that learning the meaning was much more important than learning the spelling. Historically, spelling correctly distinguished the literate from the illiterate, but no longer - and, in any case, it is the meanings of words that give richness to life, not the spelling. The national spelling bee is a fine exercise in memorization, but it needs to be complemented with a meaning bee that tests student's understanding of "word senses" of words.

Problem Statement

Create a "meaning-bee," where participants are tested on the meanings of words instead of their spellings. The meaning bee would be centered around word senses, which are defined as one line in the dictionary. Players would be shown a word and try to generate as many correct word senses as possible. In this way, players would be encouraged to know a diversity of word senses for each word, to demonstrate mastery of the word's meaning. The goal is not to have players memorize definitions from the dictionary, but rather to represent a word's meaning in a thorough and utilitarian manner.

Approach and Process:

Literature Research

We began with a literature survey and researched similar games or competitions that already existed. We took a more in-depth look into the spelling bee and geography bee in particular, to get a sense of the format and flow of these kinds of competitions. Several scenarios were generated as potential formats, applications, and environments for the meaning bee game.

The format our meaning bee game will mostly be similar to that of the spelling bee and the geography bee in that participants will be asked questions and they will provide answers to the question. The main aspect that is different is the format that answers will be given. In the geography bee, answers are given orally and most of the time, the judge provides options for the participants. In addition, in many games similar to the meaning bee, participants are often given a word and several definitions to choose from, or a definition and several words to choose from. The way our meaning bee distinguishes itself from all similar existing games is that choices are not given. The participant will have to write the answer from scratch. We believe that this way, participants will learn more from the game. In addition, future work of this game can include ways of giving participants more points if they write a more explanatory definition.

Scenarios

Scenario 1:

Players are given a word and they are asked to select all the definition that correspond to the given word. For example, the player is given the question below.

Define the word “construct” (choose all that apply).

- a. To build or form by putting together parts
- b. An image, idea, or theory, especially a complex one formed from a number of simpler elements.
- c. To arrange objects into a certain order

The answer in this case would be a and b. These choices represent two different word senses of the word “construct”. Over time, as the player answers the words correctly, the difficulty level of the words also increases. There can also be a time limit on how long the player have to answer the question. The ideal platform would be on a tablet or a phone.

Scenario 2:

This scenario takes place at a meaning bee competition. The judge asks a question about the meaning of a word and each player needs to write as many definitions as possible for that word. The number of points each contestant get is determined by how many definitions of that word the contestant provides. There will be a penalty if a wrong definition is provided. For each round, the screen will display the word and contestants will have to write down definitions of that word. After everyone has written down their solutions, the screen will display all possible definitions to the word and the judge can then determine the score of the contestant based on that. The questions will increase in difficulty (words will be harder or there will be more definitions for a word).

Scenario 3:

Children are given a definition and they are given options on what the word is. They are asked to select all the words that correspond to the definition. For example, the question below is asked.

What is “to damage something so badly that it cannot be repaired”? (choose all that apply)

- a. destroy
- b. demolish
- c. cripple

The answer in this case would be a and b since they both mean the definition.

Experiment

As an experiment, we ran a pilot study to see people's reactions towards the game and what kind of definitions we got. We tested our idea on a total of 11 people, in small groups of 2-3 people.

Our goal of the experiment was to see if the game makes sense and if people can play it and enjoy it.

This first prototype of the game took the form of a powerpoint slideshow. In the experiment, Weikun or Gina acted as the moderator and flipped through the slides according to the following procedure.

Procedure

1. Each person brings their laptop. We created a google form that each person will open on their laptops.
2. Explanation of the rules and basic concept.
 - a. **Overview:** We're about to play a game that works kind of like a spelling bee, except you will be asked to demonstrate your mastery of word meanings instead of spelling.
 - b. **Time limit:** 60 seconds to write down as many different definitions of the word that you can think of.
 - c. After time is up, each player reads out the definitions that they came up with, and the judge (Weikun or Gina) distributes points.
 - d. **Scoring:** +1 for correct definition, -1 for incorrect definition, 0 for all common definition - if everyone comes up with a common definition, nobody will receive points for that definition. This is to encourage them to come up with as many unique definitions as possible
 - e. **Two rounds - easy and hard:** The first round is comprised of common words that have many different definitions (easy). The second round has less common words and tests simply your knowledge of that word (hard).
3. Introduce with the example word. Ask if there are any questions
4. Start the game! After each word, distribute points and say who is in the lead

After the experiment, we compiled all the data into a spreadsheet and the four of us scored each of the players' responses.

Insights from Experiment

- Importance of setting an example: Many players gave definitions that we thought were too vague and generic, and some players gave more lengthy definitions while others opted for fragmented words or phrases. Having a set response format that all players can follow makes judging a lot easier and more fair.

- Hard versus easy: The “easy” words (i.e, words chosen because they had a greater volume of word senses) were more compelling. Most players could not define the “hard” words that we chose.
- Fun: Players did have fun playing the game, and further development shows promise for an engaging and competitive experience.
- Scoring: The major problem point became how to score the players’ responses and give them that result during the game. We determined that having feedback as to which word senses were awarded points gave players opportunity to learn and develop their responses.

The input and format of the game was clear--players would be presented with a word, and then have to type out as many word senses that they could think of within a set time period. Scoring became the next topic of focus.

Scoring

We brainstormed different ways to approach scoring, and how we could use technology to automate the process.

- word analysis and comparison to a computer’s definition database
- semantic spaces
 - project words into a semantic space: for ex, “stream” is close to “river” in meaning (if you’re trying to define “bank”)
 - measure the distance between the input and accepted definition to create a correct/incorrect judgement
- creating an accepted definition from the answers
 - continuously revising accepted definitions through active player feedback
 - crowdsourcing the scoring - people can vote and decide on which definitions are better, much like how we judged the pilot test responses

To automate scoring, it is necessary for the computer to have a database of definitions to compare each player’s word sense. However, just using a dictionary would not offer an adequate comparison because of how different a usable word sense can be from the dry, academic word senses offered in a dictionary.

The idea of a player alternating between defining a word and judging other players’ responses became the primary method of scoring. With this initial set of definitions, an algorithm can be developed that will compare the player’s word sense to a database of previous word senses for a particular word, to help automate and streamline the scoring process in the future.

Interface Design

From there, we began to visualize the interface of the game through some initial wireframing. We created the wireframes for the single player mode, cooperative mode, and competitive

mode, but eventually decided to stick with a simple simple player mode of the game to start with.

After finalizing the wireframes, we went through several iterations of the interface design. We mainly focused on designing the interface for the judging process and the accounts page. Initially, for the judging page, we displayed each definition and had people mark it as “correct” or “incorrect”. After discussing with Anthony and Alex, we decided to create buckets for people to drag definitions to, so that we can find out how people phrase the same definition differently. Our buckets included each correct definition, “incorrect”, “unsure”, and “correct definition but not in list”. Our next iteration mainly focused on the user interaction when the user drags and drops the definition. Our first version of the dragging and dropping interaction included dragging the user’s definition on top of the correct definition, removing the user’s definition, and adding a “1” next to the correct definition. This allows the user to click the “1” beside the list of correct definitions and see which definition was dragged there. However, after some discussion, we felt that we should simplify the design to make the interaction more intuitive. We came up with several alternatives, including just matching the definitions with arrows and having boxes for each definition and dragging it into the boxes. In the end, we decided to use arrows to match the definitions and rearranging it based on position so the arrows look cleaner.

Another main design that we added is the objections page, where the user can object to the judging result if the user thinks that the results were unfair. We included an objection button beside each of the result and also a list of pending and processed objections in the accounts page.

The next step is to implement these initial designs into a working prototype. The design that we came up with represents the simplest and most basic version of the game, and we hope that it serves as a springboard for more design and development to come out of.

Design Decisions:

Free Response

For the format of the questions, we have decided to have the players write out their definitions instead of giving them multiple choices. We feel that this will make the game more interesting and the player would also be able to learn more from the game.

Judging

Since the definition for each player can vary tremendously, we have decided to use the original definitions from a dictionary and continuously revise the definitions through active player feedback. For each round, a player will have to write definitions of a certain word and immediately judge other player’s definitions after. For the judging portion of the game, there

will be four categories of responses - "correct dictionary definitions", "incorrect", "unsure", and "correct but not in dictionary definition". The player will need to drag each definition onto one of these categories, matching the player's definition with each of the dictionary definition if the definition is correct. Using the player's responses from judging, we can compile a database and build a smart computer model based on the responses.

Level System and Points

We decided that the game needed some incentive for defining and judging words correctly. There are two point systems - playing and judging, and a player receives points when they do either. Leveling up requires the player to reach a set amount of points from both playing and judging.

Initially, 1 point was given for correct word senses, -1 point was given for incorrect, and 0 was given for "I don't know." Similarly, if a player judged a word with the majority (so their judgement was correct), they would receive a point, and otherwise they would lose a point. We decided to eliminate the deduction of points completely to focus on a system of positive reinforcement. Players only receive points for responses in which a majority of people have judged it to be correct. Judges receive points for responses where they have judged with the majority.

Cooperative and Competitive Modes

We have explored the possibility of integrating "Cooperative Mode" and "Competitive Mode" into Meaning Bee. In the cooperative mode, each player will write as many definitions as possible and the scoring will be based on the intersection of correct definitions. In the competitive mode, scoring will be based on number of unique definitions that the other player has not written.

The multiplayer modes of the game makes it more interesting but it has its disadvantages. For the cooperative mode, since players are only judged based on the intersection of definitions, they will be less willing to write definitions that are less common. For the competitive mode, since players are judged based on unique definitions, they will be less willing to write more common definitions. The goal of our game is to have players write as many definitions as possible and not only the common ones or the less common ones. As a result, we have decided to build a simple one player game.

Future Work:

Speech Recognition

Since the input of definitions require typing long texts, there is possibility of incorporating speech recognition into the input method. In this way, Meaning Bee can be expanded onto more platforms, like tablet or smartphone.

Automated Scoring

After receiving all the crowdsourced data, can this be put into a machine that judges each response?

Game Mechanics

Since we were designing for the first version of the game, the game mechanics are very basic. There is a lot of room for expansion to create more compelling gameplay.

There is the possibility of creating a more complex scoring system that distributes points based on a qualitative analysis of each player's response. We considered using the idea of "semantic spaces" to measure the semantic distance between key words and phrases in the player's response and an accepted definition.

The level system also needs to be developed more - for now, the required number of points are arbitrary because we needed more data from actual gameplay and trials. After a working prototype is created, the data can be used to make a challenging yet approachable level system.

Increasing difficulty or creating some kind of added challenge as you continue to play - this would create more incentive to level up and participate in long term play. Ranking words depending on the number of senses and the rareness of the word could be a part of this.

Dealing with objections - creating the system for people to object to their scores and how these objections will be handled.

Evaluation of the Game

As the overall intent of the game is to be a learning tool, something that needs to be explored is how to gauge the educational value of the game. One way to measure this quantifiably is to see if players can increase the amount of word senses that they come up with for a given word. This would involve re-circulating words that players have already defined as a way to re-test them on their mastery of the word. Words would be re-testing according to the correctness of the word sense that the player gave as well as the volume—if the player gave a lot of incorrect responses to a word, or gave very few word senses for a word that had a lot, that could be criteria that would necessitate a re-test. This could give an unfair advantage to

players that are being re-tested on a word, so some kind of system would need to be put into place to distribute points accordingly.

Another important area comes from the words that the player does not know and chooses not to define. Since players are shown the definition of the word after it is played, there is the opportunity for each player to learn the definition of a word they don't know. However, simply re-testing these words with the hopes that the player will eventually learn might not be the ideal situation. A way for the user to review the words that they don't know and access those definitions outside of the game might be a better way to help players learn.

Again, developing a way to measure the quality of a word sense may also be another important indicator of whether or not players are learning. It's easy to measure quantity, and coming up with the qualitative measurement also ties into a more sophisticated scoring system that could distinguish the game further.

Overall Reflection:

Both of us learned a lot about game design, and that was probably the most difficult portion of the project. Thinking about how the game would flow smoothly and the experience of the player felt like the most important part, but also the most abstract. Neither of us expected the scoring system to be so complicated, and it was difficult to conceptualize the process in the perspective of the player and also the underlying software. There was a point where we realized that we were proposing many hypothetical situations for how the game could work, and really the only thing to do next was to implement the game before moving forward.

Design-wise, the visual style was pretty straightforward. Anthony and Alex always gave us feedback on how to change the wireframes we made based on insights that we hadn't considered. Designing the interactions was probably the most difficult, again because it was hard to conceptualize and visualize the user interactions.

Overall, it was rewarding to be part of a team to get this project started, and it's exciting that the meaning bee will be implemented into a real game.