Abstract

We introduce an annotation scheme and corpus study to investigate the use of base and target components of analogies in tutorial dialogues. We present the development of the scheme and test its final form on a corpus of one-to-one tutorial dialogues on computer science, for which we achieve over 0.77 multi-rater inter-annotator agreement. We then annotate data from the same corpus to investigate the use of semantic wave structures from Legitimation Code Theory in tutoring, and we find a regular adherence to semantic wave structures in explanations which use analogies. We further identified different semantic wave shapes and show their distributions. We conclude that semantic waves and the novel characterisation of analogical explanations in tutorial dialogues reported in this investigation can be useful tools for both the analysis of human tutorial dialogue and future implementation of tutorial dialogue systems.

1 Introduction

We present an empirical study of analogy stemming from the goal of building a spoken dialogue system for computer science tutoring capable of explaining concepts using analogies. While there has been work investigating the use of analogy in tutoring, it is currently insufficiently detailed to build a system with the ability to generate analogies in an interactively natural way; in fact, in general there is an insufficient understanding of how people interact using analogies. In this paper we focus on the sequential unfolding of analogies by tutors on an utterance-by-utterance incremental basis. The paper investigates how tutors go up and down the level of abstraction during their explanations— a structure known as semantic waves (Maton, 2013) – with the motivation that discovering how this is done sequentially over the dialogue could eventually be transferred to an artificial tutoring agent.

2 Background

2.1 Analogies in Explanations

People continuously search, create and evaluate explanations (Keil, 2006; Thagard, 1989) and our explanatory capacity is similar to our ability to reason analogically (Hummel et al., 2014). Analogy is habitually interpreted as a cognitive process involving a target domain and a base (or ‘source’) domain, the former being the one that is being explicated and the latter functioning as a different but structurally similar domain used to communicate the concept (Gentner, 1983; Gick and Holyoak, 1983). An example of the base and target components of an analogy in an utterance can be seen in (1) where the base is underlined and the target is in bold.

(1) um the stack is a lot like a Lego set, okay?
concepts to students (Holyoak et al., 2001). Reasoning with analogies is conceptualised as mapping a single source to a single target (Hummel et al., 2014).

2.2 Tutorial Dialogues

Human one-to-one tutoring has been shown to be a very effective form of instruction (Chi et al., 2001) and is considered one of the most effective methods of helping students to learn. However, there are a number of variables which could either improve or impede learning gains during a tutoring session, including the domain, tutor, tutee and session structure features (Hacker et al., 2009). Under the category of structure-related variables, a number of pedagogical strategies have been tested empirically for their efficacy, including direct procedural instructions, direct declarative instructions, positive feedback, negative feedback, worked-out examples and analogies (Di Eugenio et al., 2013, 2009; Alizadeh et al., 2015). Evidence on the pedagogical efficacy of using analogical explanations shows that their presence in combination with specific dialogue acts correlate positively with learning gain (Alizadeh et al., 2015). However to our knowledge a statistical study on the use of base and target components of analogies within human one-to-one tutorial dialogues showing the structural nuances of how tutors unfold analogies over time has not yet been researched.

2.3 Semantic Waves

The annotation scheme and corpus study we present here aim to uncover patterns of base and target component utilisation in the tutoring dialogues in terms of their adherence to the the structure of semantic waves. This concept is part of the Legitimation Code Theory (LCT) (Maton, 2013) and provides an explanatory framework of what constitutes an effective explanation (Waite et al., 2019). According to semantic waves, the complexity of meanings fluctuates in terms of semantic density and semantic gravity. Semantic density is a continuum that ranges from the use of common words utilised with their ordinary meaning at the lowest density to the use of specialized brief terms or symbols at the highest density. Semantic gravity contrasts between abstract concepts and real world examples (Maton, 2011). For a learning episode to adhere to the semantic waves construct, it should start with high density and low gravity, descend to low density and high gravity and ascend back to the initial state (Curzon et al., 2018), as illustrated in the diagram in Fig 1. In Fig. 1 initially the concept of algorithms is presented abstractly as precise sequences or steps, then comparison to instructions or recipes is made to unpack the meaning as the semantic density is reduced and a more concrete or simpler base concept is used, followed by a repacking of meanings to go back to the abstract and complex meaning originally presented. In the analogical explanations we study here, we consider the base component to be used at the trough of the wave, and the target component to be at the two peaks. The annotation scheme and corpus study described below seek to answer the question as to whether analogical explanations adhere to the semantic wave structure in tutoring dialogues empirically, and therefore whether the theoretical construct is useful for modelling human-human tutoring dialogue and for designing tutorial dialogues systems.

3 Analogy Annotation Scheme

3.1 Corpora

We used 3 different corpora for developing an analogy annotation scheme, which are the British National Corpus (BNC1994), the Basic Electricity and Electronics Corpus (BEEC) and Computer Science Tutorial Dialogues (CSTD). We selected subcorpora from these three sources which have the following characteristics:

---

1From the National Centre of Computing Pedagogy Quick Read ‘Improving Explanations and learning activities in computing using semantic waves’ https://raspberrypi-education.s3-eu-west-1.amazonaws.com/Quick+Reads/Pedagogy+Quick+Read+6+-+Semantic+Waves.pdf
**British National Corpus (BNC1994)**  We used one career orientation dialogue of 700 utterances from the British National Corpus (Burnard, 2000), obtained with the SCoRE tool (Purver, 2001).

**Basic Electricity and Electronics Corpus (BEEC)**  We used a dialogue excerpt of 15 utterances and one entire dialogue of 292 utterances of tutorial dialogue of the BEEC corpus (Litman et al., 2009) also obtained with SCoRE.

**Computer Science Tutorial Dialogues (CSTD)**  The largest and principal corpus we use is that of tutorial dialogues on computer science data structures collected in the late 2000’s (Di Eugenio et al., 2009) consisting of 54 one-to-one tutoring sessions on the topics of linked lists, stacks and binary search trees. The corpus contains a total of 35,609 utterances annotated with tags signaling beginning and ending of the pedagogical strategies of feedback, worked-out example and analogy. We created a subcorpus of the utterances within all the analogical episodes plus a context of five utterances before the beginning and after the end of the episodes. Our subcorpus contains a total of 3,887 utterances—the size of our subcorpus relative to the size of the whole corpus is as in Table 1.

<table>
<thead>
<tr>
<th>Utterances within analogical episodes</th>
<th>2,528</th>
<th>7.10%</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ context</td>
<td>1,359</td>
<td>3.81%</td>
</tr>
<tr>
<td>Utterances outside analogical episodes</td>
<td>31,722</td>
<td>89.09%</td>
</tr>
<tr>
<td>Total</td>
<td>35,609</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

Table 1: Analogical episodes comprising the CSTD sub-corpus as percentage of whole corpus.

**3.2 Base and Target Annotation Scheme Development**

The development of our annotation scheme included the participation of six researchers, two of whom are the first two authors of this paper. Five of them participated in the annotation exercises. The following paragraphs explain the settings and results of each iteration.

**3.2.1 Iteration 1**

For the first iteration, two non-native English speakers annotated the BEEC subcorpus described in Section 3.1. The annotators were instructed to mark each utterance as including only the base (B), only the target (T), both (BT) or none of the analogy components. Before the annotation exercise, the annotators were provided with an annotated example of the career orientation dialogue from the BNC1994 subcorpus, which was 700 utterances long, of which 25 were annotated with B, T or BT. Each annotator then executed a practice run with an excerpt of a BEEC dialogue of 15 utterances with real-time feedback from the main author. During the interactive practice, the disagreements were discussed with a twofold purpose; identify any annotation rule which elicited disagreements and creating new rules if needed. Every modification and new rules were made explicit in an updated version of the manual. After the practice sessions, utilising the updated version of the annotation scheme manual and in individual sessions, they annotated the other BEEC dialogue comprised of 292 utterances.

The two-way Cohen’s Kappa (Siegel and Castellan, 1988) inter-annotator agreement results are as in Table 2 where G is the Gold Standard we assume, which are the annotations by the first author, compared against annotators A1 and A2. The right-hand column also gives the Fleiss’ Kappa multi-rater agreement of all three annotators reaching a moderate agreement level of 0.544.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>B, T, BT, N</td>
<td>0.644</td>
<td>0.304</td>
<td>0.544</td>
</tr>
</tbody>
</table>

Table 2: Kappa inter-annotator agreement on Iteration 1.

Table 3 shows a tutoring dialogue in which the three annotators agree about all the utterances containing the analogy component of type base. In this case, the tutor explains electrical potential energy (the target domain) by referring to a ball tossed in the air (the base domain).

Table 4 shows a dialogue excerpt in which the two annotators agree about all the utterances containing the analogy component of type target. In this case, the tutor explains conservation of energy.

Finally, Table 5 shows a tutoring dialogue in which the two annotators agree about all the utterances containing both analogy components of type base and target while the third annotator marks the last two of these as only being base.

In addition to the potential ambiguity between both (BT) and base (B), one of the most frequent sources of disagreement in this first iteration was due to not considering anaphoric references to base
Think of a ball tossed into the air. At first the upward force caused by your hand throwing it causes it to move up. But eventually it stops - gravity causes it to slow down until it stops. Then it falls down.

Table 3: All three annotators in Iteration 1 agree on base annotation. P = participant.

Again, energy would be conserved. You just have to think what that energy was converted into. Some of it would be converted into heat because of the friction, etc.

Table 4: Two annotators agree on target.

You’re right that kinetic energy was zero, but at the maximum height, when the ball stops, the height makes it possible for it to start moving again. Now it’s going to start moving in the opposite direction. So that height, since it will make it possible for the ball to move, is a form of energy. It’s the total energy that is conserved, not the kinetic energy, since the velocity of the ball is not constant.

Table 5: Two annotators agree on 4 utterances while the other annotator disagrees with them.

Table 6: Excerpt of disagreement in Iteration 1

3.2.2 Iteration 2

For the second iteration, two monolingual native English speakers were recruited as annotators, with the purpose of increasing the inter-coder agreement. The set-up was adjusted such that annotators had to decide whether each utterance contained a base (B) or not as a binary decision, and also whether the utterance contained a target (T) or not. The annotators coded the same BEEC dialogue of 292 utterances used in iteration 1 and received the same coding rules, with the addition of the rule that considers anaphora. They were provided with the same annotated example which was provided as per the previous iteration and also executed the practice annotation with the main author giving live feedback on their decisions. This session allowed for discussion and clarification of the rules in the provided manual. As in iteration 1, all changes were registered in an updated version of the annotation manual. The results from iteration 2 on the two labels are as in Table 7 (again with Cohen’s Kappa for the pair-wise agreement and Fleiss’ Kappa for the three-way multi-rater agreement).

While very high agreement is reached on the base component, there was large disagreement on identifying target utterances, particularly the agree-
ment between the gold standard annotation (first author) and annotator A2.

### 3.3 Final Annotation Scheme

For the third and final iteration, we used the CSTD corpus for both; the practice and the disagreement and language interpretation experiment, the main reason being the fact that the CSTD was the corpus we wanted to do the study of the semantic waves on. Another change in this iteration was the substitution of one of the two monolingual native English speakers.

The disagreements with A1 and A2 from iteration 2 were discussed and the manual updated accordingly. The annotators coded a new dialogue using the final version of the annotation manual based on these insights. The definitions and examples given to annotators for annotating base and target components is as in Fig 2. An expanded version of the instructions are as in the Appendix A.

The annotators were provided with an annotated example of 6 analogical episodes from the CSTD corpus consisting of 193 utterances. The annotators executed a practice annotation exercise with another selection of 3 analogical episodes and a total of 116 utterances of the same subcorpus and received feedback from the main author with the purpose of clarifying their questions when they judged the annotation rules did not fit in a particular case. Once this practice was executed, to test the agreement the annotators annotated a new selection of 5 analogical episodes for a total 188 utterances of the CSTD corpus. The final inter-annotator agreement Kappa results are as in Table 8, where it can be seen a strong overall agreement with all three annotators is reached for base and target at over 77%.

### 4 Corpus Study on Analogical Episodes in Tutorial Dialogues

With the appropriate level of agreement reached, the main author annotated the entire subcorpus of the CSTD analogical episodes of 3,887 utterances with the scheme described above (each utterance with the two binary decisions for base and target presence) which contains all the sections annotated as analogical episodes and, additionally, the single analogies (Di Eugenio et al., 2009) which consists of individual utterances, and in both cases the 5-utterance context window either side of the analogical utterances.

#### 4.1 Descriptive statistics of all episodes

The histogram in Fig. 3 shows the distribution of lengths of the analogical episodes in number of utterances.

Table 9 and Fig. 4 show the distribution of utterances labelled as containing only base, only target, both or no analogy components, derived from the two binary labels B and T. Note that 33% of utterances contained no analogical components and mainly communicative grounding types of dialogue acts. For the analysis that follows, and consistently with the concept of question under discussion (Ginzburg, 2012), we assume those to have a B, T or BT component still under discussion based on the most recent label in the dialogue.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>B</th>
<th>T</th>
<th>BT</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1269</td>
<td>1302</td>
<td>859</td>
<td>457</td>
<td>3887</td>
<td></td>
</tr>
<tr>
<td>32.65%</td>
<td>33.50%</td>
<td>22.10%</td>
<td>11.76%</td>
<td>100%</td>
<td></td>
</tr>
</tbody>
</table>
Analogy definition and examples:

An analogy is a linguistic device that uses a specific concept (a base) to transfer information or meaning to another concept (a target). Tutors use analogies to explain concepts. The following examples contain a section of a tutoring dialogue which includes an analogy. The base concept is formatted with underlined characters and the target concept is formatted with bold characters.

<table>
<thead>
<tr>
<th>Utterance Examples</th>
<th>Base</th>
<th>Target</th>
<th>B</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>um the <strong>stack</strong> is a lot like a Lego set, okay?</td>
<td>lego set</td>
<td>stack</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>uh a <strong>binary search tree</strong> is +/\ one way of looking at it is like a family tree.</td>
<td>binary search tree</td>
<td>family tree</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>the comparison I like to make with <strong>linked lists</strong> is a movie line.</td>
<td>movie line</td>
<td>linked lists</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Figure 2: Annotation definitions and examples for Base and Target for Annotators

**Vis-a-vis** the complexity of meaning in semantic waves, we map the concepts of linked lists, binary search trees and stacks, our target analogical components, to the notion of low gravity (i.e. abstract concept), and the references to people queing at movie theaters, family members, restaurant trays, sheets of paper and other tangible examples, which are our base analogical components, to the notion of high gravity (i.e. concrete concept) in semantic waves.

We would expect analogical episodes to begin with the target component of an analogy, descend in terms of semantic gravity to the base component, and then ascend again to the target concept at the end of the episode– see Section 2 for details. Here we define a semantic wave as any descent in terms of semantic gravity between utterances and rise again, so, for instance beginning with a utterance with both base and target (BT), descending to base (B) and then back to BT is still counted as a wave. Dialogues excerpts exemplifying the semantic waves can be seen in Tables 10, 11 and 12.

To test this, we automatically searched for semantic waves in the 138 analogical episodes of our CSTD subcorpus and we found that 129 (93.47%) of them contained at least one, which supports the idea tutors use the semantic wave in analogical episodes. We also found that a mean of 2.1 waves are used in every analogical episode in a tutorial dialogue explanation.

64 of the 138 episodes had at least two consecutive semantic waves. The distribution over the number of waves per episode can be seen in Fig. 5. One episode contained 16 consecutive waves.

We additionally found that there were seven
main types of waves which represent different patterns of base and target components. We take from the surfing domain the names of the types of waves, which vary from strong to weak. The strongest is the point break wave as shown in Fig. 6, and in our sequence model it represents starting the analogical episode with a target component, descending to the base component at some point during the
episode, and finishing with the target again. Two sub-types of point break wave were also observed, namely the point break descending and point break ascending—see Figs. 7 and 8. The next type of wave is moderate in intensity and is called a reef break, which can be ascending (Fig. 9), descending (Fig. 10) and standard (Fig. 11). Finally, the weakest type of wave was observed, the beach break wave (Fig. 12).

In total the 291 waves existing across the 138 episodes were distributed by type as is in Figure 13. Example dialogue excerpts showing a point break, reef break (standard) and beach break wave can be seen in Tables 10, 11 and 12.

5 Discussion and Conclusion

We have presented, to our knowledge, the first annotation scheme and corpus study which investigates how the base and target components of analogies are deployed by human tutors during their explanations. We used the annotation scheme to verify whether analogical explanations follow the structure of semantic waves, whereby they begin from the target component, descend to the base component and return to the target. 93.47% of the episodes contain the structure of a semantic wave and, 74% of the episodes used a series of semantic waves consecutively. We showed there are a variety of different wave types used and we define some shapes to understand these different structures. We claim this novel characterisation of analogical explanations in tutoring dialogues to be a formalisation that could be used as a tool for both; the design of human tutorial dialogue pedagogical strategies and intelligent tutoring systems.
<table>
<thead>
<tr>
<th>Tutor</th>
<th>*uh a binary tree is kind of like mother and father and xxx a family tree.</th>
<th>BT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student</td>
<td>no that’s not bad *uh that’s bad.</td>
<td>N</td>
</tr>
<tr>
<td>Tutor</td>
<td>it’s +/- because families can have more than two kids.</td>
<td>B</td>
</tr>
<tr>
<td>Tutor</td>
<td>so here what it means is that binary is that each node can have two trees, two children.</td>
<td>BT</td>
</tr>
</tbody>
</table>

Table 12: Beach Break Wave Dialogue Example

In future, we intend to further investigate the semantic wave structure in analogies in a more fine-grained level, to analyse the mapping between the base and target domains which happens dynamically during the semantic wave explanation and to incorporate and test these explanatory models in a spoken dialogue system.

Acknowledgements

Del-Bosque-Trevino is partially supported by EP-SRC and AHRC Centre for Doctoral Training in Media and Arts Technology (EP/L01632X/1) and CONACYT (National Council of Science and Technology of Mexico). Purver is partially supported by the EPSRC under grant EP/S033564/1, and by the European Union’s Horizon 2020 programme under grant agreements 769661 (SAAM, Supporting Active Ageing through Multimodal coaching) and 825153 (EMBEDDIA, Cross-Lingual Embeddings for Less-Represented Languages in European News Media). The results of this publication reflect only the authors’ views and the Commission is not responsible for any use that may be made of the information it contains. Thanks to Thomas Kaplan, Louise Bryce, Shamila Nasreen, Morteza Rohanian and Jack Ratcliffe for their participation in the development of the annotation scheme and annotations.

References


Diane J. Litman, Johanna D. Moore, Myroslava Dzikovska, and Elaine Farrow. 2009. Using natural


A Annotation Instructions

Purpose of the annotation exercise The main purpose of this annotation protocol is to identify interactive patterns of explanations analogies in human-human tutoring conversations. This study will be conducted on a dataset consisting of 54 one-to-one human basic computer science tutoring dialogues collected in the late 2000’s. The dialogues topics are limited to three basic computer science (CS) data structures, which are: stacks, linked lists and binary search trees

Analogy definition and examples An analogy is a linguistic device that uses a specific concept (a base*) to transfer information or meaning to another concept (a target*). Tutors use analogies to explain concepts. The following examples contain a section of a tutoring dialogue which includes an analogy. The base concept is formatted with underlined characters and the target concept is formatted with bold characters.

<table>
<thead>
<tr>
<th>Utterance Examples</th>
<th>Base</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>um the stack is a lot like a Lego set, okay?</td>
<td>stack</td>
<td>lego set</td>
</tr>
<tr>
<td>uh a binary search tree is +// one way of looking at it is like a family tree.</td>
<td>binary search tree</td>
<td>family tree</td>
</tr>
<tr>
<td>the comparison I like to make with linked lists is a movie line.</td>
<td>linked lists</td>
<td>movie line.</td>
</tr>
</tbody>
</table>

Analogy bases and analogy targets are alluded indirectly by the use of coreference. Analogies which are alluded by the use of coreference should also be annotated.

CHILDES notation The research group which annotated the 54 dialogues used the CHILDES notation. The following table contains the symbols you might encounter during your analysis. Use this table as a reference when doing your annotations.

<table>
<thead>
<tr>
<th>Example</th>
<th>Marker</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>uh# you know what a linked list is?</td>
<td>#</td>
<td>pause between words</td>
</tr>
<tr>
<td>it’s a +// it’s a concept, not a language thing.</td>
<td>+//</td>
<td>self interruption</td>
</tr>
<tr>
<td>so all you’re given is this header, that why h@l is here.</td>
<td>h@l</td>
<td>the letter h</td>
</tr>
<tr>
<td>&lt;/*uh, and they want us&gt;///&lt; oh O_K_.</td>
<td>&lt;&gt;</td>
<td>angle brackets group words marked by the following symbol, in this case, retracing with correction</td>
</tr>
<tr>
<td>&lt;/*uh, and they want us&gt;///&lt; oh O_K_.</td>
<td>///</td>
<td>retracing with correction</td>
</tr>
<tr>
<td>then you’re losing all your xxx.</td>
<td>xxx</td>
<td>unintelligible speech</td>
</tr>
<tr>
<td>yeah, but yeah, then you know +...</td>
<td>+...</td>
<td>The trailing off or incompletion marker (plus sign followed by three periods) is the terminator for an incomplete, but not interrupted, utterance.</td>
</tr>
<tr>
<td>&lt;/the second one wants&gt;///&lt; so that was an insertion, the second one is a deletion.</td>
<td>///</td>
<td>retracing with reformulation</td>
</tr>
<tr>
<td>(be)cause the xxx.</td>
<td>(be)</td>
<td>noncompletion of a word</td>
</tr>
<tr>
<td>you got to start here at the root, just like in &lt;a binary&gt; /// in a linked list you have to start at the first node.</td>
<td>///</td>
<td>retracing with correction</td>
</tr>
<tr>
<td>++ right.</td>
<td>++</td>
<td>other completion</td>
</tr>
<tr>
<td>star+wars</td>
<td>+</td>
<td>compound or rote form marker</td>
</tr>
</tbody>
</table>
The three cases of coreference that we should be able to spot and annotate are called anaphora, cataphora and coreferring noun phrases.

**Continuation and preambles** Some utterances contain few words (4 or less) and are continuations of the previous utterance of the same speaker, or a preamble of the following utterance. Assign the same annotations that you gave to that speaker’s previous or subsequent utterance.

**Session management** Session management utterances should not be considered or annotated.

**Utterance Example (Metacognition)**
let’s start off  
let me just grab a clean sheet of paper

**Metacognition** Utterances which are observations or reflections of the mental processes that occur during the tutoring session should not be considered or annotated.

**Utterance Example (Session Management)**
OK, now you are on to something  
I’m so happy that you understand now  
It’s good that you recognize that your answer doesn’t look right.  
That shows you’re thinking!  
I like to see that

**Additional considerations**
- An utterance can be allocated as base and target at the same time.
- Rely only on the text.
- Within the linked lists analogies, tutors and students sometimes refer to letters as if they were people (e.g. b@l is looking to c@l). If this is the case, annotate as BASE. Referencing letters does not determine that speakers are talking about the TARGET
- Annotate as TARGET when there is an explicit reference to concepts within the TARGET domain.