Computing Frege's Principle of Compositionality: An Experiment in Computational Semantics[1]

Kathy Linton and Neal Bouffard

Presented to: Hans-George Ruprecht
LALS 3506 - Section A - Semantics

1 Carleton University Cognitive Science Technical Report 2004-07
http://www.carleton.ca/ics/techReports.html

Introduction and Background

> *"The speaker ... defends his impression (which even in real life may be illusory) by attributing to his conversational partner all sorts of background knowledge, insights and reasoning ability. But again, these are the* ***speaker's*** *contribution to the conversation."* (Weizenbaum, 1966)

In 1966, Joseph Weizenbaum produced a paper about a rather simple computer program that he'd written. The paper introduced the world to ELIZA, a program which was meant to understand natural language responses from its users, showing that computers would be capable of the exact sort of understanding people insisted that they were not. ELIZA was a simple construct, little more than a system of pre-written (or canned) responses with a few rules for keyword matching from the user-input. By following these simple rules, however, ELIZA became an excellent mimic of a Rogerian-style psychoanalyst, posing new, salient, questions from each user-response, delving deeper into the nature of the user's issues. ELIZA seemed at least a partial solution to the challenge the famous Turing Test, in that "she" was able to fool at least some of the people some of the time (whereas Turing posited that a true Artificial Intelligence would be reached when users could not, in conversing with a unit, determine whether they were interacting with a machine or a human being, in effect being able to fool all of the people all of the time). In fact, ELIZA was so effective that, in some experiments, test subjects actually felt better in having spoken to her.

Of course, ELIZA could not understand what was being said to her. She had no mechanisms toward comprehension, her internal states were not in any way affected by user input other than in the selection of her own responses. ELIZA's ability to trick people into believing her to be human also faded over multiple sessions or prolonged exposure, as the patterns in her responses became more and more obvious. One terminal flaw is that ELIZA would often pose questions which were unlikely to generate a keyword response, and then react badly to the keyword-free response (imagine someone asking you a question which you answer directly and then being asked what you meant by your response). In effect, ELIZA was the very prototypical example of Searle's Chinese Room: a system which matched input to output using only rules and had no understanding or even context as to what it was that she was doing, just like the person in Searle's conceptual room. The limitations of this approach became clear as later experiments and attempts to make ELIZA more complex did not in any way increase ELIZA's abilities to understand (or, more accurately, to make people feel understood). In fact, over 35 years after ELIZA's introduction to the world, her initial success was still her strongest, despite increases in the understanding of both computation and cognition (although, to be fair, these must also be balanced against an increase in user-sophistication and familiarity with technology). In fact, ELIZA and ELIZA-like programs are now considered to be little more than parlour-tricks in which people delight far more in confusing the program than "properly" approaching the experiment.

Still, when considering a group project for this course, it was ELIZA which inspired us. Perhaps due to our mutual background in Cognitive Science, the thought of attempting to pick up where ELIZA left off was a very appealing one. Therefore, after some research on the subject of ELIZA and some consideration of the issue and its relevance to a combined linguistics and philosophy course in Semantics as well as the problems that we would encounter (including, most importantly, time constraints), we decided to invert the focus of a classic ELIZA-like program. Instead of attempting to sound intelligent in response to user input without comprehension as ELIZA did, we sought to build a system that is capable of understanding the input it is given, without necessarily being able to respond intelligently. After all, simply repeating Weizenbaum's experiment with quasi-intelligent canned (or semi-canned) responses, even if there was an updating the responses to appear more modern and less rigid, would not truly be adding anything to the available body of knowledge on the topic of the computation of semantic knowledge (and, as a consideration, this is a ***Semantics*** course, after all). In addition, the source-code for several versions of ELIZA are easily accessible online (in fact, there are even some running, online instances of ELIZA which can be found, complete with their source code, such as the one at http://www.manifestation.com/neurotoys/eliza.php3) demonstrating how simple building an ELIZA-like construct is to modern programmers. Ultimately, we decided that if the need to have our program respond intelligently were to arise, such a thing could be added later.

As such, we began to look at the problem of understanding sentences, attempting to determine a method by which a computer could be made to comprehend standard speech utterances. In performing our research, we were faced with the truly massive scope of the problem we were describing. Essentially, we were looking at creating a full-effect Natural Language Processor - a task far beyond our own abilities (in fact, one which has not yet been accomplished by anyone, anywhere). As such, we began to consider limiting the scope of our experiment. Rather than attempting to understand the whole of the English language, we decided to create something more akin to an expert system (a term used in Artificial Cognition research to refer to a system with a specific type of knowledge) and narrowed our focus to the task that

Weizenbaum claimed generated the best successes for ELIZA: that of dealing with human emotion. Specifically, we sought to create a program capable of asking a person how they were feeling and from real-world utterance type responses (be they single-word, sentential or fragmentary) to these questions and generating an understanding of their responses within this limited context.

Theoretical Concepts

> *"[The universe] cannot be read until we have learnt the language and become familiar with the characters in which it is written."* (Gottlob Frege, as cited on the School of Mathematics and Statistics of the University of St. Andrews website @ http://www-gap.dcs.st-and.ac.uk/~history/Quotations/Frege.html)

Obviously, the simplest approach to such an undertaking would be to process only single-word responses. Single-word user input could then be compared against the contents of a lexicon, drawing inferential strengths and context from the values stored with each word, giving the response a relatively easily measured meaning, assuming the existence of such a lexicon. Fortunately, a lexicon of emotional states (and therefore, emotional words) can, in fact, be built based on the fact that, ultimately, there a limited number of types of emotional primitives (Solso, 1998). Thus, by limiting ourselves to defining the lexical meanings based on these emotional primitives, we create what is essentially a multi-dimensional theoretical space in which all possible emotional states exist (an emotion-space, to coin a simple term), where the strengths of each word along the various axises generate its coordinates within that space. As such, the inferential strength and lexical meaning of each word within the emotion-space can be expressed as a vector. This translation, of course, works both ways, such that we can look at any single point within the space and express it as the combination of the six axises.

Unfortunately, single-word responses, while quite common in conversation, are not the whole set of utterances. Therefore, a means to parse and draw meanings from sentential and fragmentary forms was necessary, and one which could be implemented a computer. Since the cognitive aspects of parsing ambiguous sentences is still not fully understood and believed to be at least in part performed by the parallel processing systems of the mind (Solso, 1998) and the existing background and contextual knowledge we posses (Saeed, 1997), attempting to duplicate the methodology that human listeners use was, at the time of the experiment, impossible. (And likely to remain so for quite some time, despite attempts to build vast cognitive inference engines at such institutions as MIT, or the more successful CYC by Cycorp) Fortunately, one of the simplest principles of semantics presented itself as a viable and relatively easily implemented system for analysing a sentence's meaning: Frege's Principle of Compositionality.

Frege's Principle of Compositionality (sometimes simply referred to as Frege's Principle) states that "the sense if a complex is compounded out of the senses of the constituents". (Dummet, 1981, cited by Pelletier, 2001). As such, the meaning of a complex such as a sentence or sentential fragment can be created by combining or concatenating the meanings of the constituent words and expressions that make up the sentence. This concept is key to many modern theories of linguistics which give serious attention to semantics, including the MG (Montague Grammar), GPSG (Generalized Phrase Structure Grammar), CG (Categorial Grammar) and LTAG (Lexicalized Tree-Adjoining Grammar) (Janssen, 2001). One reason that this principle is widely embraced within the Linguistics community (beyond the obvious fact that it seems, intuitively, to make a great deal of sense) is the simplicity of formalism that it entails. Compositionality allows sentence-meanings to be expressed as some type of algebraic expression of the meanings of the words within, whether a simple addition in its simplest form to a far more complex lexical-mathematical construct as would be found in the Montague Grammar.

Thus, theoretically, creating the ability to draw sentential meanings for the experiment became a matter of treating each complex response as a rapid series of single-word responses, where each word in the response is compared to the lexicon and the lexical values for each are generated and added together as would be a series of multi-dimensional vectors, with the overall (emotional) meaning of the complex utterance then being the location within the emotion-space that is the result of the vector addition. Because we know the address of the point in space where we end up, we can apply a single inferential strength (which is to say, create a single, simple meaning-vector) to the entire utterance by generating the vector that brings us to that point in space (which may or may not be identical to some other single-word utterance's meaning vector).

Methodology

"We shall do a much better programming job, provided we approach the task with a full appreciation of its tremendous difficulty, provided that we respect the intrinsic limitations of the human mind and approach the task as very humble programmers." (Alan Turing, as cited on the Quotations for CS1 website @ http://www-2.cs.cmu.edu/~pattis/quotations.html)

"When a programming language is created that allows programmers to program in simple English, it will be discovered that programmers cannot speak English." (Anonymous)

Having established a theoretical background for the experiment, the next step was to convert the theories into reality (or, at least, virtual reality). Research and theorizing was one thing, the time was then to create an application which would implement our version of Frege's Principle. In homage to Weizenbaum's work with ELIZA in the 1960's, we named our program "Lizzie".

Before programming could begin on Lizzie, however, the first (and most important step) was to build her Lexicon. While there are numerous online corpa available, including actual computational and categorical lexicons, (one of the most remarkable and easily available being WordNet, build at Princeton University, homepage: http://www.cogsci.princeton.edu/~wn/) our desire for a limited subset of English, as well as our methodology for generating emotional meaning-vectors made using these quite impractical. As a secondary consideration, also, was the nature of the test subjects who would be performing the experiment. It seemed most appropriate that our lexicon should be generated, and in fact, given its strengths by the very people who would be our most likely test subjects: our fellow students. To this end, a questionnaire was circulated to roughly eighty (80) people in which we sought to learn what words for each type of emotional state are in "common" use. (A copy of this questionnaire can be found in Appendix

A, along with some samples of the responses with the names blanked out). Unfortunately, the response rate was less than half. The responses from this questionnaire were then compiled, with the words categorized and sorted into a series of eight new questionnaires in which we asked people to assign lexical strengths to a subset of the words and terms gathered from the first questionnaire. (We used a subset because, ultimately, we believed there were too many different terms to ask people to rate each and every one). These new questionnaires were distributed to the same eighty people in an attempt to gather rankings, thinking that we would compile the results and calculate mathematical means for each axis strength for each term and use these values. (A copy of each of these questionnaires can be found in Appendix B, along with some samples of the responses with the names blanked out). Unfortunately, the response to this questionnaire was less than twenty percent, which meant that our data from the second questionnaire was quite open to selection bias, doubly so due to the non-uniform return rate of the sub-forms (one of them, literally, had zero responses). The effects of this selection bias were even more apparent when we considered some of the results which were given to us. (This particular issue, that of questionable survey results, will be addressed again in the "Problems" section of this document). As such, a secondary method of generating lexical strengths was required. In this, we gathered a small group of peers (a total of six individuals) in order to collaborate and generate a ranking along each axis of the words which. In each case, we sought to place 'stronger' words ahead of the 'weaker' ones, bridging in towards neutrality from each extreme. We then used these hierarchies to implement numerical scores along each axis for each word, and then placed these scores into the lexicon - each score representing one address component of the total lexical vector.

Unfortunately, it quickly became quite clear that this lexicon was insufficient to handle complex utterances. There existed a class of words which could not be given lexical strengths on their own, and yet were playing an influential role in the total meaning of the sentence. As an example, consider the word *extremely*. This word cannot be given a meaning vector within the our emotion-space of meaning on its own; after all is *extremely* happy or sad? Angry or relaxed? Nervous or calm? Yet, ignoring the word as neutral would not allow Lizzie to generate a proper sentential meaning from user-input. Thus, Lizzie gained a secondary, smaller lexicon of modifier words and their own strengths. In this case, however, the strengths of the modifiers represented not a point in the emotion-space, but rather a scalar multiplier that would be applied to the lexical meaning-vector of the following term. (Fortunately, the extreme order-dependancy of the English language makes this possible: in almost every case, modifiers *precede* that which they modify, which must be a part of the same utterance. For instance: *very happy* vs. **happy very*). In the case of multiple modifiers, the values were multiplied to each other until the modifier-chain was applied to an emotional lexical term. Thus, *very, very happy* would be as much stronger, proportionally and not absolutely, than *very happy* as *very happy* is from *happy*.

Lizzie's purpose, then, was to take user input from a dialogue, and then attempt to evaluate a person's day based on the emotional responses that they gave. She would question the user, and then based on their responses, attempt to determine the final point within emotion-space where the person would be, overall, for the day, and then collapse this six-axis vector address into a single scalar value, from -100 to +100. The mathematical transformation for this effect is a somewhat arbitrary one, ultimately, derived largely because it "felt right" after a degree of self-experimentation, in terms of which of the vector dimensions would have the strongest effects on an overall report of the status of one's day. The user would then be asked to rate their own day on the same scale (-100 to 100), and then Lizzie would record silently the self-reported values against her own calculated values.

With the lexicons generated (Appendix D contains the full listing of Lizzie's primary (emotional) lexicon, while Appendix E represents her secondary (modifier) lexicon), the actual coding of Lizzie was likely the easiest part of the process. Working from the conversational script that we had developed, we created Lizzie in Java, choosing to generate a modular structure to the program, creating classes for the Lexicon (rather, the union of the two lexicons), Emotion (Graded) and Modifier Words, and the User Interface as a whole. (Appendix C contains a full-source listing of all of each of the component classes which constructed Lizzie) It was at this point that the Alpha testing phase began. The Alpha Testing process was carried out by running the program on ourselves, as well as a few of our closest and most trusted friends. In these cases we would compare Lizzie's results to those of the self-reported person and attempt to determine where and how we, and Lizzie, were going wrong.

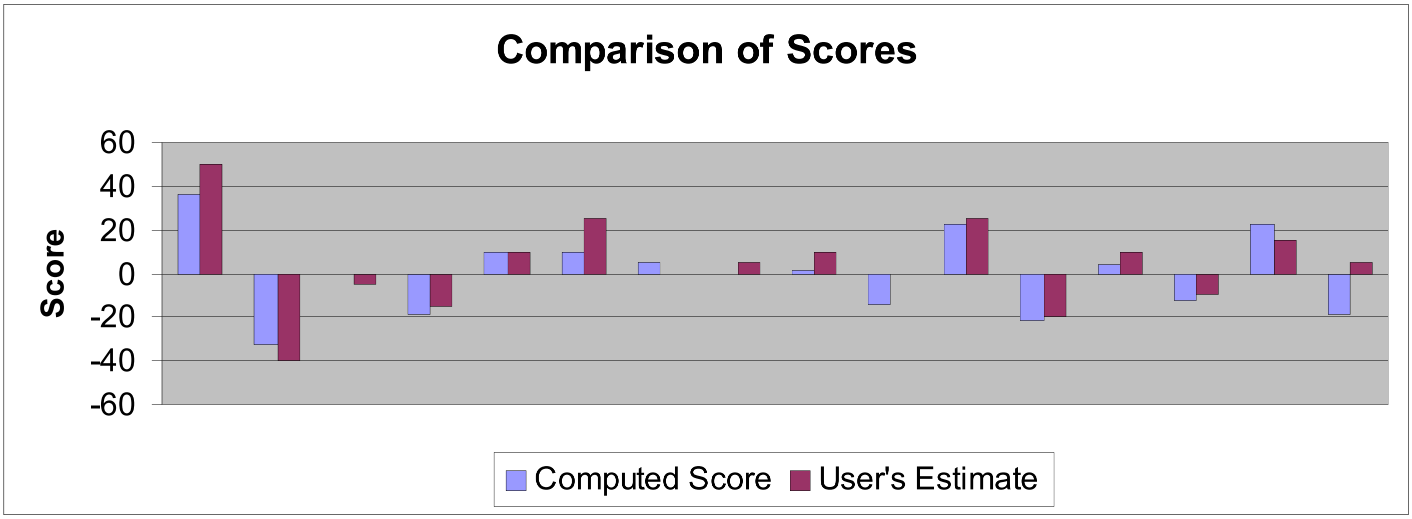
Alpha Testing revealed a few flaws in Lizzie, notably that her dialogue was not sufficiently encouraging to users to have them express their emotions. In addition, there was the matter of temporal effects - emotions not currently affecting a person could still influence the way they considered their day. This is true both in terms of delayed effects, wherein someone who was feeling fine now may still claim a negative day based on the strength of a very upsetting morning, and anticipatory effects, wherein someone looking forward to (or dreading) a later event will report an effect on their day without expressing any of the anticipated emotions in the dialogue. Lizzie was rewritten in an attempt to handle these problems, her functionality remaining largely unchanged, with only modifications to her dialogue being implemented. We then proceeded on to Beta Testing, asking the same testing group as our Alpha Testers.

Beta Testing revealed no new flaws in Lizzie, although it did indicate that some of the improvements that we'd made had overcorrected certain problems (the impact of temporal effects were being too strongly considered, as an example, treating them with the same strength as current emotions), while others had not increased Lizzie's accuracy enough. After a few minor adjustments to what we still considered to be trouble spots, and a few tests, we were prepared to begin the experiment.

The experiment process itself is quite simple. A user is placed before a console which is ready to begin running Lizzie with just a keypress. Lizzie then begins her dialogue, first instructing the user in how to use the system, and then asking questions about the user's emotional state at present, and then with slight forays into the past and future. Lizzie does make use of a few feedback mechanisms to the user, expressing either pleasure or sympathy based on responses which merit them (and neutral responses only to those which do not). To finish the experiment, Lizzie makes a text-based declaration of the person's day (based on the collapsed vector), and then asks the user to evaluate their day numerically. At this point, Lizzie thanks the user and exits. However, during this entire process, Lizzie is recording the questions she's asked and the user responses to them (in order to assist us in determining what, if anything, she is doing wrong). At the end of the logfile is Lizzie's score, the user's self-reported score and the difference, all values which are used to calculate our result statistics.

Results

"Every good mathematician is at least half a philosopher, and every good philosopher is at least half a mathematician." (Gottlob Frege, as cited on the School of Mathematics and Statistics of the University of St. Andrews website @ http://www-gap.dcs.st-and.ac.uk/~history/Quotations/Frege.html)



<Hey Kathy! Insert the Results Here!>

Problems

"Sometimes I seem to see a difficulty, but then again, I don't see it." (Gottlob Frege, as cited on ThinkExist.Com @ http://www.thinkexist.com/English/Author/x/Author_4431_1.html)

"If you're not failing every now and then, it's a sign that you're not doing anything very innovative." (Woody Allen, as cited on the Quotations for CS1 website @ http://www-2.cs.cmu.edu/~pattis/quotations.html)

In examining the results of the experiments, it became clear than even the changes made during the Alpha and the Beta testing were not quite enough to make Lizzie perfectly accurate. (Of course, considering the fact that human beings often do not understand each other, we do not believe that anyone or anything could be perfectly accurate). Throughout the various testing phases, and the final experiment phase itself, a few issues presented themselves, which we will now describe and discuss in detail.

The first issue we encountered during our testing phases was that of *underreporting*. As children, we are often told to "suck it up" (or other delightful adages along the same vein, such as "I'll give you something to cry about" or "Don't be such a baby.") when we're upset or feeling hurt, and generally, displays of negative emotions (such as crying) are frowned upon in public. There is, essentially, a societal pressure to be happy, or at least, to act happy, even when we are not. This effect manifested itself during tests with Lizzie. Simply put, some people who were having negative days downplayed their feelings when asked how they were feeling, and yet, assigned a large (negative) mathematical value to their day, overall. In reading over the transcripts of the experiment sessions, it is hard to fault Lizzie in her "misdiagnosis" of these individuals - the responses appear, even to a human being, as indicative of only a mildly negative situation. Ultimately, as this problem had manifested during Alpha and Beta testing, we attempted to rework the dialogue to reduce this effect, but it never quite went away.

The next issue, or problem, that we encountered we refer to as the *Drama Queen* phenomenon. The term *Drama Queen* is somewhat of a misnomer as they appear in both genders, however, it is the expression we began using when gathering results, partially because it is already a lexicalized expression, and partially because the first such person we encountered was, indeed, female. In fact, *Drama Queens* have plagued the project at all stages where others were involved. First, in the gathering of data, when attempting to evaluate the relative strengths of words, questionnaires came back where *every* term was rated extremely strongly. In one extreme case, all words were given maximum ratings, or slightly off maximum ratings (50s and 45s), despite the relative "weakness" of some of them. (Try as we might, we could not understand how someone could give the word *nice* a lexical strength of 45 out of 50 on **any** axis when terms such as *amazing* and *incredible* existed to be evaluated) Results such as this, when combined with our lack of response, made the second-pass of questionnaires totally worthless and forced us to implement a different methodology in assigning lexical strengths to words, as described in the Methodology section of this document.

We were not free from this phenomenon once the data gathering had been done, however. In experimental conversations with Lizzie, the *Drama Queen* effect still existed, manifesting itself as the opposite of the *underreporting* problem. In this case, people would use lexically strong words (such as *ecstatic*, *amazing*, *terrible*, *horrible*) and strong modifier terms (*very, extremely*) and then, after these vastly powerful emotional responses, rate their days as somewhat average (scoring perhaps +/-40 or so, out of a possible +/-100). Again, it is hard to fault Lizzie for a misinterpretation of this type, for human evaluation of the transcripts tended to agree more with Lizzie than the self-reported values.

The issue of *Temporal Displacement* discussed in the Methodology section remained a problem. Ultimately, we found that we had no means to determine how strongly to rate past and future events. Some people's earlier emotional states or anticipatory effects had a very powerful bearing on their self-evaluation, while others seemed to be all but totally unfazed by anything outside their present feelings. In human-to-human conversation, this would be somewhat clear as we would notice the emphasis or lack thereof when certain questions were answered, however, Lizzie is incapable of such a thing. Lizzie's current methodology to treat the past and future combined as a single-strength modification is quite artificial, yet represented the best compromise we could determine at the time. In comparison, human methodology when dealing with a scripted dialogue *without* the benefit of the unspoken, such as intonation, body language, facial expressions is just as limited as Lizzie. This was proven to us when we were reading over the session transcripts and evaluating the responses ourselves.

Ultimately, Lizzie's largest limitations lay within her ability to understand complex utterances. While using an extremely strong interpretation of Frege's Principle to attempt to understand a sentence, Lizzie is subject to limitations which would not plague a human with the same interaction. First amongst these is that Lizzie lacks the contextual information that exists outside of the verbal (or, in this case, typed) channel. Whereas in human-to-human conversations, even ones performed purely textually, there is a great deal of non-verbalized information which is accessed and used, drawn from situational context, inference and entailments (Saeed, 1997), Lizzie is incapable of the feats that would grant her access to this information. Other systems have been built which have attempted to make use of entailment/context information, but as yet, despite holding vast sets of knowledge primitives (literally, millions of small instructions in some cases), these systems cannot (at the present time) perform the simple tasks of drawing information and conclusions from input that a human can.

Second, Lizzie suffers from an extremely limited lexicon. While she understands many emotional words, granting them definitions within the emotion-space that we have created, these words and modifiers represent the entirety of her vocabulary. As such, it is quite possible that extremely meaningful responses (or portions thereof) are ignored as Lizzie does not have the ability to understand the words used (or worse, the user has mistyped). Ultimately, we are not adhering to Frege's Principle in its entirety, as we are ignoring non-emotional and non-modifier words in parsing our meanings from complex utterances, however, the limited domain of understanding we seek does much to balance out this issue. An additional concern about Lizzie's lexicon, however, is that it was built based on the response to questionnaires of Canadian university level students. Therefore, it seems quite likely that the words that she knows, and their rankings, would prove to be less than ideal if the experiment were to be performed on subjects of a different social class, generation or even region.

Finally, Lizzie suffers from a total lack of syntactic understanding. Lizzie cannot identify NPs or VPs within a sentence, and instead derives the overall sentential meaning based on a lexical analysis of each word in the utterance. Again, this is an application of Frege's Compositionality Principle, however, the fact is that Lizzie does not understand things like actions and events, which certainly have an impact on the sort of thing she seeks to discuss. Presently, Lizzie's dialogue is worded in such a way to attempt to only gain personal emotional responses, and this has been largely successful. This is, however, an artificial limitation at best. The fact that Lizzie lacks an understanding of NPs, for instance, means that she is incapable, for instance, of realizing the difference between "I was extremely mad." and "My boyfriend was extremely mad." While attempting to enforce some simple parsing rules could be implemented (looking for the "I", for instance), the fact of the matter is that Lizzie was meant to understand all "standard" utterances and not only formal, correct sentences, which such a construct would change.

Discussion

> *"Computers do not solve problems; computers carry out solutions, specified by people, to problems."* (D.D. Spencer, as cited on the Quotations for CS1 website @ http://www-2.cs.cmu.edu/~pattis/quotations.html)
>
> *"Well, we* ***are*** *wizards. We're supposed to meddle with things we don't understand. If we hung around waiting 'till we understood things we'd never get anything done."* (Ridcully the Wizard, in the novel Interesting Times by Terry Pratchett)

From the outset, Lizzie was intended to be a limited system. As discussed earlier, the idea of creating a fully functional Natural Language Processor, while appealing, was one which was never seriously considered. Simply put, it was far too lofty a goal for the confines of an undergraduate term project. However, having performed the experiment has granted us insights into the areas of computational semantics, human-computer interaction and language-processing AI. In addition, while some of the limitations from which Lizzie suffered were anticipated (in fact, some were even there by design), there were also some results which were somewhat surprising, such as the issues regarding human nature in cases such as underreporting. We will now address the issues of these limitations in our final discussion, attempting to present alternative methodologies or enhancements that could be applied to Lizzie in order to bring her closer to that ideal goal of a Natural Language Processor.

The first point that we seek to address is the remaining effects of misreported results, whether as a result of underreporting or the *Drama Queen* phenomenon, which are, essentially opposite sides of the same coin. The problem, distilled, is that people are using words which are either lexically weaker or stronger than they ought to use based on their emotional status. The solution to this problem is somewhat easy to see, if somewhat difficult to implement: having

Lizzie learn who people are (based on their names, for instance) and have her derive a scaling factor based on how far "off" she is from their status on prior experiments. For instance, Lizzie could realize that, when dealing with a *Drama Queen*, all of the words used should be evaluated as if their strength was only 2/3s of what it would be when dealing with a normal person. We did not implement this solution due to time constraints, both in the generating of the derivation math (which would begin to look like the backpropegation algorithms uses in Neural Network/Connectionist models) and in the need to perform multiple experiments with the same subject. In addition, it seems likely that a single scale factor may be insufficient, requiring instead a scale-factor for each of the emotional axises, or perhaps even a shift of the lexical strengths of each word within the lexicon for each user. Ideally, these issues could be solved in much the way that Speech Recognition software is now implemented: by having a user-configuration period before the user begins the experiment in earnest in which Lizzie gets to know their "voice', as it were. By applying similar principles to that of a Connectionism model, Lizzie could ask questions of the user and compare her expected response to the user's response and fine tune her own meanings against these results. Using such an algorithm to modify a "reasonable" base lexicon would mean that users would not need to rate each word in the lexicon (a most daunting task) in order to fine tune the results.

The next point to consider, then, flows naturally from that, which is Lizzie's limited lexicon. Ultimately, the only solution to that problem is to increase the lexicon by adding new terms. Several possibilities to automate this process come to mind. The first is to anchor her existing lexicon and add to it the contents of an electronic dictionary, such as the Webster's Unabridged (which is available on CD-ROM). Once the file format of the words was established, a simple script could extract all words within that source which have definitional similarities to the words in our Lexicon, granting us a much larger host of words to use (though these words would then need to be rated along the six axises). One solution to the need to plot these new words in our virtual emotion-space would be to have Lizzie process the text definition as an utterance. While this would not generate precise or perfect meaning-vectors for the new terms, it would create a starting point, which, if combined with the learning faculty discussed earlier, could naturally implement these new words somewhat painlessly. A second possibility would be to make use of an available lexicon such as WordNet, discussed earlier in the paper, drawing the terms out by the categories in which they are listed. The generation of meaning-vectors in this case would require creating some sort of translator from the lexical form of the words within the WordNet lexicon and software into the six-axis system which we are using.

The next issue to consider is syntactic knowledge. Lizzie's ability to understand complex utterances is wholly based on Frege's Principle of Compositionality, in which we say that "the sense if a complex is compounded out of the senses of the constituents". (Dummet, 1981, cited by Pelletier, 2001). Depending on one's reading (or possibly translation) of Frege's Principle, however, this can be considered a rather strong, possibly even extreme, interpretation. A more widely accepted form of Frege's principle is "The meaning of a compound expression is a function of the meaning of its parts and of the syntactic rule by which they are combined." (Janssen, 2001) The inclusion of syntax to meaning of compounds means that Lizzie's abilities to understand will always be impaired if she cannot be made to understand the syntactic structure of the utterances that she is given. To be honest, this limitation was expected as we had encountered both interpretations of Frege's Principle during the research for our project, however, time constraints and ease-of-programming required that the first, "stronger" interpretation be applied. Also, we thought that the limited structure of the dialogue would make this a practical application, which we still believe is the case. However, looking toward making Lizzie into something powerful and flexible instead of the rigid and somewhat brittle construct that she is at the time of this writing will require that she be capable of understanding syntax. The additional costs and complexities in both computation and implementation (the very issues which precluded such an implementation in this experiment) would, in addition to being able to handle referential events as discussed in the section on problems, also create the potential for understanding complex lexicalized expressions (such as *kicked the bucket*) based on their position as an NP, VP or PP within the utterances' framework. Lizzie would also then, assuming a vastly expanded lexicon (likely, based on maintaining her current framework, a tertiary (non-emotional) lexicon), be able to handle life events in addition to emotional states, such that a statement such as *"My mother died"* would have a strong emotional impact (and Lizzie would rate accordingly), and not simply have their power based on the presence of the word *died* as would be the case if we were to expand Lizzie's lexicon without looking at the combinational (syntactic) rules used to generate the sentence. Therefore, *"My mother died"* would be considerably stronger than *"My goldfish died"* or the even weaker *"My walkman's batteries died."* (Although, to be fair, having the batteries die on one's walkman unexpectedly really is *quite* frustrating, and some people may be *very* attached to their goldfish and less so their mothers.) Combining Lizzie with the syntax parser generated by Joshua Tacoma and Michael Dufton (as discussed in Dufton & Tacoma, 2003, unpublished) would be a leap forward in this regard, even with the lingering issues of sentential (form) ambiguity and the difficulties in processing incomplete utterances that exist within that parser. Unfortunately, doing so within the time constraints of a single term proved wholly impossible (especially considering the fact that said parser was being created at the same time as Lizzie herself), yet it remains a very strong consideration as one looks toward possible futures for Lizzie.

A final consideration when dealing with Lizzie comes from looking at the current state of the art in terms of existing language processing and inference engines. Cycorp, who have built CYC, considered at present to be the leading general knowledge base and inference engine, have released an open-source toolkit making use of the same technologies. OpenCyc (website: www.opencyc.org) is a powerful and flexible toolkit, which contains a great deal of, essentially, common sense information in an attempt to mimic to some extent the interconnectivity of human knowledge. Had we discovered OpenCyc at an earlier stage in the project, we may have actually attempted to perform Lizzie's implementation within the OpenCyc framework rather than Java. OpenCyc has a degree of Natural Language Processing, although primarily in the ability to generate Natural Language from its own principles expressed in Cyc-L (Cyc-Language) rather than understand Natural Language input. Additionally, OpenCyc is far more strongly related to syntax than semantics, drawing its primary English lexicon from WordNet, and essentially attempting to determine meaning from structure and reference to other internalized facts. As such, OpenCyc as a general system does not make use of rules such as Frege's Compositionality, either strongly or weakly, and its (WordNet) lexicon holds no concept of things such as emotional strengths, which were necessary for Lizzie. Yet, none of these things pre-existed in Java, either, prior to our creation of Lizzie, and Java offers little in terms of the tools for future expansion that OpenCyc provides. Ultimately, while using OpenCyc would not have helped us in our ability to create Lizzie, it would also put us in a position where the same sorts of increases of functionality we describe earlier in this section are more easily attainable. Thus, we could have Lizzie prove not this simple and rigid form of Frege's Principle, but a more robust and rich one.

- Neal Bouffard & Kathy Hoos

References:

Dever, J. (1999), "Compositionality As Methodology" in Linguistics and Philosophy 22 : 311-326. Amsterdam: Kluwer Academic Publishers.

Dufton, M. & Tacoma, J. (2003). As yet unpublished, completed for the Carleton University Course LALS 3506 in the Winter Term of 2003.

Fromkin, V. (1997), An Introduction to Language: First Canadian Edition. Toronto: Harcourt Brace Canada.

Frege, G. (1948). "*On Sense and Reference"* The Philosophical Review 57: 207-230. In the public domain.

Janssen, T.M.V. (2001), "Frege, Contextuality and Compositionality" in Journal of Logic, Language and Information 10 : 115-136. Amsterdam: Kluwer Academic Publishers.

Pagin, P. (1997), "Is Compositionality Compatible with Holism?" in Mind and Language, Vol. 12 No. 1. Oxford: Blackwell Publishers.

Pelletier, J. (2001), "Did Frege Believe Frege's Principle?" in Journal of Logic, Language and Information 10 : 87-114. Amsterdam: Kluwer Academic Publishers.

Saeed, J. (1997), Semantics. Oxford: Blackwell Publishers.

Solso, R. (1998). Cognitive Psychology (Fifth Edition). Toronto: Allyn and Bacon.

Weizenbaum, J. (1966). *ELIZA – A Computer Program For the Study of Natural Language Communication Between Man and Machine*. Retrieved March 14, 2003 from http://i5.nyu.edu/~mm64/x52.9265/january1966.html.