Saurus: An Emotionally-Weighted Thesaurus

Jim Gouldstone, Hugo Liu, Henry Lieberman, Hiroshi Ishii

MIT Media Lab
20 Ames St., Bldg E-15
Cambridge, MA 02139 USA
{jimbones, hugo, lieber, ishii}@media.mit.edu

Abstract

For a given word, a thesaurus offers a selection of synonyms. It does not, however, suggest how a user might choose between these synonyms. This paper presents Saurus, an affectively augmented thesaurus. In Saurus, an emotional rating is attached to every entry and synonym. This provides a user with a friendly and logical basis for selecting one synonym amongst many: emotional connotation. Using this same basis, a program can instantly replace many words within a document, enabling the unification of word-choice decisions under a coherent affective tone. Saurus can well support a writer's brainstorming efforts, allowing her to easily explore many affective revisions.

Introduction

When we lookup an entry in a thesaurus, what we know and what we want to know are fairly clear. We already have a sentence structure in mind, and a word we'd like to replace in that structure. We're using a word that fits to find another word that fits. As they are both suitable, it may be said that they have the same structural, or *literal* meaning. If that were all there were to it, we wouldn't require a new word. So beyond the literal meaning, there must be a second consideration: *emotional connotation*.

Emotional connotation is what brings a text to life. Is an author thrilled, miserable, furious, or hateful? Are her characters and settings bubbling and vibrant or twisted and eerie? A well-communicated, unified tone can help to smoothly convey attitude or intent. It can serve to more successfully convince, persuade, or appeal to a particular audience, and can limit written misunderstandings by compensating for missing face-to-face cues.

In this paper we present Saurus, a thesaurus in which these emotional connotations are clear. An affective weight is assigned to synonyms, better constraining the problem of selecting an appropriate word. We also present a simple means of synonym selection, and a program that uses this idea to instantly and automatically select replacements for multiple words within a document, facilitating creative revision by allowing a writer to explore many affective possibilities easily.

Related Work

Though tremendous effort has gone into the generation and organization of synonym sets, (e.g., WordNet -- Fellbaum, 1998) as far as we know there are no systems that assist the user in the task of selecting among synonyms in such a set.

Previous research has demonstrated both the evaluation and generation of affective text. Liu's applications (2003, 2005), for example, have evaluated text and indicated a broad range of emotional tones using color and emoticons. Similarly, Ruberry, et. al. (2005) automatically varied the facial expressions of virtual actors to suit the text they were reading. This is quite relevant, as Saurus must evaluate affect at three points during its operation: database creation, user choice of tone, and interface feedback. In fact, Saurus relies heavily on the affect-sensing methods featured in these works.

Affective text generation has come in several flavors. Hovy's PAULINE (1990) generated affect based upon a speaker's attitude toward a subject. This attitude could be user-specified (speaker views a subject as good, bad or neutral), and manifested itself partially through word choice. Elliott (1998) instead modified the tone of a tale by choice of character behavior via 'story-morphing'. Finally, Callaway's STORYBOOK (2002) generated whole stories with lexical considerations, but with the aim of preventing repetitive prose. In all three cases, input was typically in a coded format (i.e., machine-readable tagged text indicating concepts and relations). Saurus differs from these in that it is a more of a natural language *modifier* than a generator. The input and output are natural language, the affective results are governed exclusively by lexical (synonym) choice, and the user can specify one of a continuous and broad range of emotional tones.

```
love                ( 0.571,  0.203,  0.305)
----
NOUN:
  1. Deep and ardent affection
    adoration       ( 0.447,  0.014,  0.123)
    devotion        ( 0.492,  0.070,  0.255)
    worship         ( 0.302,  0.030,  0.106)
  2. The condition of being closely tied to
another by affection or faith
    affection       ( 0.589,  0.202,  0.214)
    attachment      ( 0.347,  0.015,  0.132)
    devotion        ( 0.492,  0.070,  0.255)
    fondness        ( 0.425,  0.091,  0.183)
    liking          ( 0.338,  0.024,  0.147)
    loyalty         ( 0.413,  0.063,  0.172)
  4. A strong, enthusiastic liking for
something
    love affair
    passion         ( 0.473,  0.318,  0.194)
    romance         ( 0.387,  0.102,  0.159)
  5. An intimate sexual relationship between
two people
    affair          ( 0.305,  0.021,  0.136)
    amour
    love affair
    romance         ( 0.387,  0.102,  0.159)
  6. The passionate affection and desire felt
by lovers for each other
    amorousness
    fancy           ( 0.336,  0.021,  0.137)
    passion         ( 0.473,  0.318,  0.194)
    romance         ( 0.387,  0.102,  0.159)
VERB:
  1. To feel deep, devoted love for
    adore            ( 0.375,  0.073,  0.170)
    worship          ( 0.302,  0.030,  0.106)
  2. To like or enjoy enthusiastically, often
excessively
    adore            ( 0.375,  0.073,  0.170)
    delight          ( 0.551,  0.073,  0.140)
```

Figure 1. Database entry for the word "love". Each term is accompanied by an emotional rating (pleasure, arousal, dominance) ranging between –1.000 and +1.000 .

Implementation

The system is comprised of three parts: the database, the user interface, and the program. The user makes requests via the interface. The program then interprets those requests, and drawing upon information found in the database, can structure and return an appropriate result, again via the interface.

The Database

The heart of the system is a modified copy of Roget's II Thesaurus.[R] The emotional weight of each entry and synonym is appended as a PAD triplet (Mehrabian, 1995). PAD format describes an affective tone using three independent numerical axes: pleasure (P), arousal (A), and dominance (D). Each axis may vary between –1.0 and +1.0. For example, a PAD triplet with a negative D value would suggest submissiveness, and a positive A would suggest excitement. This representation allows for simple programmatic comparisons between tones, as affective similarity is reduced to Cartesian distance. For this reason, PAD was favored over systems featuring dependent axes, such as Ekman's six (1993), which would have required more complex comparisons.

The actual rating for each word is generated using spreading activation. A small, hand-crafted list of weighted words (ANEW - Affective norms for English words) (Bradley, 1999) is expanded using relations found in a large knowledge corpus (OMCS - Open Mind Commonsense) (Singh, 2002). We begin with 1,000+ ANEW word ratings determined by psychological focus groups. Using the statements found in the OMCS corpus, these rated words are linked to thousands of terms and synonyms found in the thesaurus. Simply, if two words appear in the same OMCS statement, they are assumed to be linked. These associations are then used to suggest weights for as yet unrated words. Specifically, the PAD value of an initially unrated term is determined by averaging the values of all the ANEW words with which it is associated. This method has been used with some success by Liu for textual affect sensing, and in this case successfully generates weights for about 70% of the thesaurus' 35,000 synonyms.

Note that WordNet was considered as an alternative to OMCS for making word associations. As the relations in WordNet are largely denotative, such associations would have likely reinforced similarities between synonyms, resulting in a very small spread of emotional weights within a given synonym group.

The augmented thesaurus is stored as twenty-six flat text files. Entries beginning with the same letter share the same file. The format is intentionally easy for a human to skim, and allowed us to confirm that the weighting system made intuitive sense before development continued. Figure 1 reproduces the entry for the word "love". The weightings largely agree with expectation; "affection" is a more pleasurable synonym than "attachment", for example, and "passion" is by far the most arousing of all the synonyms.

The User Interface

The interaction consists of three main steps, shown in Fig. 2. First, the user enters some text she wishes to alter. This can range in size from a single word to a full document. Second, rather than enter a PAD value directly, the user enters a guide phrase with the intended tone. Typing "sexy" is far simpler than visualizing and entering its PAD triplet (0.507, 0.339, 0.283). Furthermore, suggesting a tone by example should keep the system accessible to users unfamiliar with PAD. Third, the program applies the suggested tone to the original text, and displays the result. The system confirms the tone of each step with a PAD indicator.

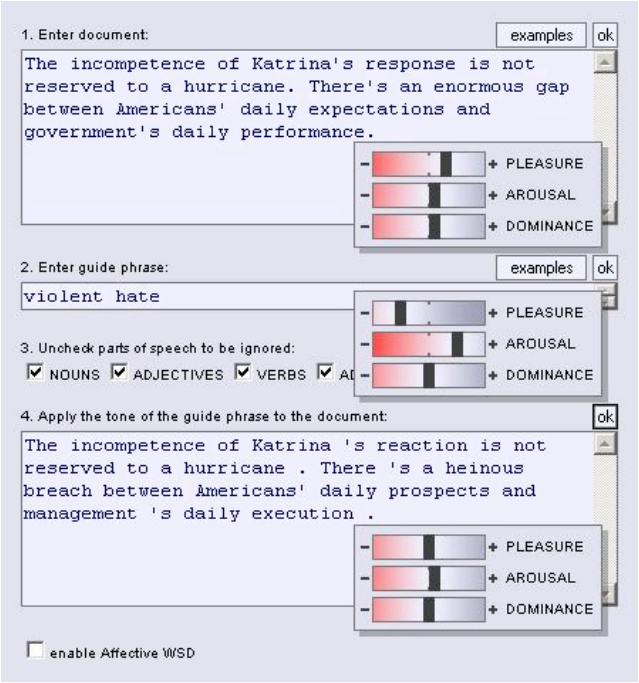

Figure 2. The emotional tone of the harsh guide phrase, "violent hate", is applied to a quotation, resulting in a more negative and inflammatory choice of words.

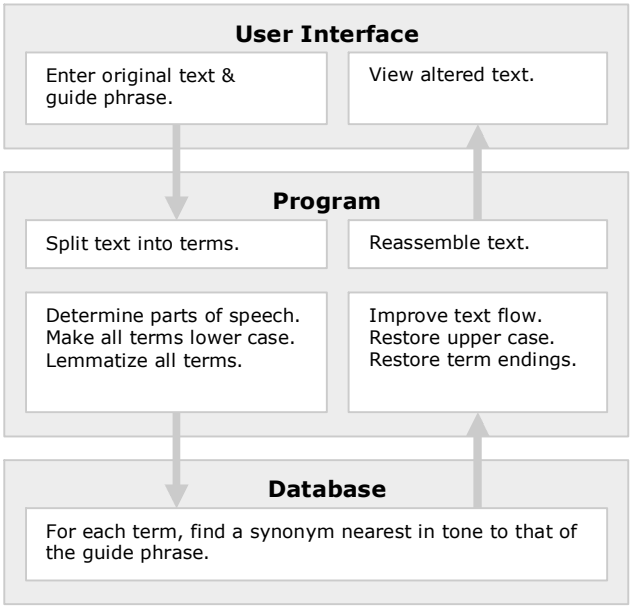

Figure 3. Overview of the system.

The Program

The program performs three major functions: disassembly, replacement, and reassembly (Fig. 3). Input text is broken down into single terms, which are in turn reduced to forms that may be found within the thesaurus. *Compare* is in the database; *comparing*, *compared*, and *compares* are not. So all words are reduced to lower case, lemmatized, and their parts of speech are identified. Next, the program attempts to identify the sense of the words from the context. Each word is then replaced with the synonym nearest to the target tone. Endings appropriate to part of speech are restored, and basic flow is improved. (e.g., During replacement, "an automobile" may have become "an car". This is corrected.) Finally, capitalization is restored and the new terms reassembled. The result is a similarly structured body of text, but with individual words swapped out.

Evaluation

Our initial experiments and experiences are reported here.

Single Term Performance

For single terms, Saurus yields very promising results. For example, when the positive guide phrase “angel” is applied to the term *to anger*, Saurus returns *to bristle*, a gentle-sounding term that still conveys the mechanical meaning. On the other hand, the guide phrase “devil” returns *to burn*, a harsher-sounding synonym. “Raging hell fire” yields an even more negative and arousing result: *to explode*. Finally, a neutral phrase like “potatoes and beans” results in a less arousing term with softer connotations: *to foam*. In this single-term case, we see results that agree with affective expectations.

Document Performance

Saurus produces interesting results for larger bodies of text as well. In a speech made in Sept. 2005, U.S. Senator John Kerry criticized the government’s response to Hurricane Katrina’s devastation of New Orleans:

> The incompetence of Katrina's response is not reserved to a hurricane. There's an enormous gap between Americans' daily expectations and government's daily performance.

Had the speechwriter used Saurus to modify the tone with the guide phrase “violent hate”, Senator Kerry might have spoken:

> The incompetence of Katrina's reaction is not reserved to a hurricane. There's a heinous breach between Americans' daily prospects and management's daily execution.

Several of the words have been replaced with those that have more negative connotations. Most notably, the "enormous gap" has become a "heinous breach". Also, it is interesting to note that the negative tone of some replacement words comes not from their present usage (*execution* meaning *operation*) but from another of their senses (*execution* meaning *capital punishment*).

Note that the program is evaluating the *tone* of the guide phrase, not the *meaning*. Strong phrases are used here because they yield livelier results. The speechwriter could have employed the phrase "everlasting love" to instead soften the criticism:

> The incapability of Katrina's answer is not reserved to a hurricane. There's a heroic incongruity between Americans' daily expectancies and governance's daily performance.

Here, "heroic incongruity" still means "enormous gap", but the term heroic has many positive connotations. "Incompetence" which implies an almost willful ineptitude has become "incapability", which instead conveys a kind of blameless inability.

Word Sense Disambiguation

Though the above results are encouraging, the system can produce nonsensical results. Most commonly, this is due to the program choosing the incorrect meaning for a particular word. The following is a quote from *The Washington Post* regarding Iraqi elections:

> Polls opened at 7 am Thursday for nationwide elections in Iraq, with a large explosion in central Baghdad moments later serving as a reminder of recurrent violence. (Washington Post, Dec.15, 2005)

After modification by Saurus, it reads:

> Heads extended at 7 am Thursday for nationwide choices in Iraq, with a substantial outbreak in key Baghdad imports afterward serving as a remembrance of cyclic coercion.

This example clearly illustrates the program's word sense disambiguation difficulties. Part of speech locks have been added to the interface in response to this. The user can uncheck NOUN, VERB, ADVERB, and/or ADJECTIVE in the interface, specifying that the program leave words of that type unchanged. The assumption was that the "how" elements (modifiers: adverbs and adjectives) might be more fault-tolerant than the "what" elements (nouns and verbs) of a sentence. Here is the above example again, but with verbs and nouns locked:

> Polls opened at 7 am Thursday for nationwide elections in Iraq, with a substantial explosion in key Baghdad moments afterward serving as a reminder of cyclic violence.

Unexpected Results

Finally, some word sense mix-ups can be entertaining. They are most effective when multiple sequential words have their mechanical meaning drastically altered, but the sentence continues to make sense. A grandfather searching for his grandchild might say, *"Where's my pretty little girl, eh?"* but Saurus muddles the senses and asks, *"Where's my fairly inconsequential girl, eh?"*

Interface Performance

The interface works well for smaller bodies of text (4-5 lines). The PAD indicators help to quantify and compare tones, and the equally sized input and output text areas make it easy to spatially match a word with its replacement. Larger documents are not so easily evaluated, however. As number of words grows, the PAD value for a body always approaches a single neutral value. (Thus, guide phrases are best kept short.) Also, changes in word length alter how the text wraps, making it difficult to spatially pair a word with its original. A more consistent word layout should ameliorate this.

Applications

There are many uses for a system that can quickly and automatically alter written tone. Of course, professional and amateur writers can use it to experiment with phrasing, whether describing a scene, or composing character dialog. There are also several less obvious uses.

Political Spin – Public figures can instantly comment on their own mistakes or successes in favorable tone, and respond to opponents in a more critical tone.

Video Game Characters – Computer controlled characters can speak what is essentially the same dialog in a range of different tones, depending upon their attitude towards the player.

Children's Stories – Saurus could be modified to examine every word in a document and exaggerate its tone. Wicked witches become *diabolical*, dark villains *sinister*, and mighty heroes *gargantuan*, making the story larger than life. (Conversely, a document intended for legal purposes could be made to sound more neutral.)

Three-stage Poetry - An experimental poet could compose the mechanical and emotional meanings of a poem in two separate works, structure in one, and feeling in another. They could then be overlapped, line by line, to produce a complete result.

Reader Preferences - A *netizen* reading from a particularly aggressive forum or slanted news source could subtly alter the tone to something more agreeable.

Foreign Language - Another use could be in language learning. Subtleties of strength of meaning are hard for non-native speakers to pick up (e.g., Which connotes the strongest praise in French: aimable, amicale, sympathique, gentil?).

Conclusion

It is possible to programmatically classify words beyond simple literal meaning. Saurus has demonstrated that it is also possible to use an augmented thesaurus to ease user synonym selection through the use of an affective guide phrase. Furthermore, this method can be used to automatically select and replace multiple synonyms, thus changing the tone of an entire document while preserving its basic literal meaning. Such a system can greatly assist writers, allowing them to quickly generate and explore many affective revisions.

Further Work

Though Saurus has demonstrated some interesting ideas, it needs more work before it can be counted on to reliably produce readable documents. Nearly all of the work that needs to be done involves word sense disambiguation. Any solid gains in Saurus' WSD ability will be immediately apparent in its output. Concurrently, effort should be made to make the system more fault-tolerant, or fail-soft. The system has a fail-soft aspect in that the user is never left guessing what has gone wrong. It is usually quite clear that the wrong sense has been used. Though this transparency should make a user more forgiving, more reliable performance would be welcome.

References

Bradley, M.M., & Lang, P.J. (1999). *Affective norms for English words (ANEW)*. Gainesville, FL. The NIMH Center for the Study of Emotion and Attention, University of Florida.

Callaway, C. and Lester, J. *Narrative Prose Generation*. Artificial Intelligence 139(2):213—252, 2002.

Ekman, P. *Facial expression of emotion*. American Psychologist, 48, 384--392. 1993.

Knickmeyer, Ellen. *Polls Open in Iraq As Danger Remains*, Washington Post Foreign Service. Dec.15, 2005; Page A20

Elliott, C., Brzezinski, J., Sheth, S., Salvatoriello, R., *Story-morphing in the affective reasoning paradigm: generating stories semi-automatically for use with "emotionally intelligent" multimedia agents*, Proceedings of the second international conference on Autonomous agents, p.181-188, May 10-13, 1998, Minneapolis, Minnesota, United States

Fellbaum C (Ed): '*WordNet: An Electronic Lexical Database*', MIT Press (1998).

Hovy E., *Pragmatics and natural language generation*, Artificial Intelligence d3(2), pp 153-197, 1990.

Kerry, John. *Speech at Brown University*. Sept. 19, 2005. http://www.johnkerry.com/pressroom/speeches/spc_2005_09_19.html

Liu, Hugo (2004). *MontyLingua: An end-to-end natural language processor with common sense*. Available at: web.media.mit.edu/~hugo/montylingua.

Liu, H., Lieberman, H., Selker, T. *A model of textual affect sensing using real-world knowledge*, Proceedings of the 8th international conference on Intelligent user interfaces, January 12-15, 2003, Miami, Florida, USA

Liu, H. and Maes, P. (2005), *Proceedings of the Eighteenth International Florida Artificial Intelligence Research Society Conference (AI in Music & Art Special Track),* 74-79. Menlo Park, Calif.: AAAI Press.

Liu, H., Selker, T., Lieberman, H. (2003) *Visualizing the Affective Structure of a Text Document.* Proceedings of CHI, 2003, pp. 740-741

Mehrabian, A.: 1995, *Framework for a comprehensive description and measurement of emotional states*, Genetic, Social, and General Psychology Monographs, 121: 339-361.

Roget's II: The New Thesaurus, 3rd ed. Boston: Houghton Mifflin, 1995. www.bartleby.com/62/.

Ruberry, M., Owsley, S., Shamma, D.A., Hammond, K., Budzik, J., *Affective Behaviors for Theatrical Agents*, Workshop on Affective Interactions (2005) international conference on Intelligent user interfaces.

Singh, P., The public acquisition of commonsense knowledge. Proceedings of AAAI Spring Symposium. Palo Alto, CA, AAAI. 2002.