Sentences vs. Phrases: Syntactic Complexity in Multimedia Information Retrieval

Sharon Flank
eMotion, Inc.
2600 Park Tower Dr., Ste. 600, Vienna, VA 22180
sharon.flank@emotion.com

Abstract

In experiments on a natural language information retrieval system that retrieves images based on textual captions, we show that syntactic complexity actually aids retrieval. We compare two types of captioned images, those characterized with full sentences in English, and those characterized by lists of words and phrases. The full-sentence captions show a 15% increase in retrieval accuracy over the word-list captions. We conclude that the syntactic complexity may be of use in fact because it decreases semantic ambiguity: the word-list captions may be syntactically simple, but they are semantically confusingly complex.

1 Introduction

In this paper, we describe experiments conducted on an image retrieval system, PictureQuest, which uses text captions to characterize images. The text captions are of two types. Optimally, they consist of a prose description of the image, generally two to three sentences, with perhaps three or four additional words or phrases that describe emotional or non-literal image content, e.g.

Two little girls play with blocks. The younger girl, wearing a blue shirt, laughs and prepares to knock over the tower that the older girl has constructed. The older girl, dressed in a red shirt, winces in anticipation.

Siblings, cooperation, rivalry

Some of the captions in PictureQuest are not as well-behaved. They may contain legacy data or data shared with a keyword-retrieval system. They are optimized for exact-match retrieval, and, as such, consist of lists of words or, at best, a few short phrases mixed in with long lists of words. The same image might appear with the following caption:

girl, girls, little girl, little girls, block, blocks, play, playing, plays, blue, red, shirt, tower, knock, over, construct, construction, siblings, cooperation, rivalry

PictureQuest relies on several natural language processing techniques to enhance retrieval accuracy. It contains a part-of-speech tagger, morphological analyzer, noun phrase pattern matcher, semantic expansion based on WordNet, and special processing for names and locations. These have been tuned to perform most effectively on caption text of the first type, i.e. sentences. The following chart illustrates how these linguistic processes operate – or fail to operate – on syntactic units.

Process	How it operates on an intact phrase	What happens when the phrase is reduced to a word list
Tagger	dog-N herding-V sheep-N	dog-N,V; herding-N,V; sheep-N
Morphology	dog herd-ING sheep	(same)
NP Patterns	*small child* wearing a hat *green swirls*	small, child, wearing, hat green, swirls (modifiers de-coupled from head nouns)
Semantic Expansion (WordNet-based)	cat jumping into the air: cat-N (7 senses) jumping-V (13 senses) air-N (13 senses)	cat, jumping, air cat-N,V (9 senses) jumping-N,V,Adj (16 senses) air-N,V,Adj (20 senses)
Names	George Bush, Al Gore	George, Bush, Al, Gore (matches *bush, gore*)
Locations	Arlington, Virginia New England	Arlington, Virginia (matches other *Arlingtons* in other states) New, England (matches *England, new*)

2 Complexity Measures

2.1 Competing Complexity Measures

How do we determine what syntactic complexity is? Does it relate to depth? Nesting? Various definitions have been used in the various research communities: Alzheimer's research, normal and abnormal child language acquisition, speech and hearing, English teaching, second language teaching and acquisition, and theoretical linguistics of various persuasions (see, e.g., MacDonald 1997; Rosen 1974; Bar-Hillel et al. 1967). Fortunately, for the purposes of our investigation, we are dealing with broad distinctions that would foster agreement even among those with different definitions of complexity. For the captioned data, in one case, the data are in full sentences. The average sentence length is approximately ten words, and the average number of sentences is between two and three. In the other case, the data are either in lists of single words, or in lists of single words with a few two-word or three-word phrases included, but with no sentences whatsoever. Regardless of the exact measure of syntactic complexity used, it is clear that sentences are syntactically more complex than word lists or even phrase lists.

2.2 Query Complexity

The standard query length for Web applications is between two and three words, and our experience with PictureQuest confirms that observation. In comparisons with other text-based image retrieval applications, including keyword systems, query complexity is important: one-word queries work equally well on keyword systems and on linguistically-enhanced natural language processing systems. The difference comes with longer queries, and in particular with syntactic phrases. (Boolean three-word queries, e.g. *A and B*; *A or B*, do not show much difference.) The more complex queries (and, in fact, the queries that show PictureQuest off to best advantage) consist either of a noun phrase or are of the form *NP V-ing NP*. The table below summarizes the differences in query complexity for natural language information retrieval as compared to keyword-only information retrieval.

Query	NLIR vs. Keywords
one word, e.g. *elephant*	Both are equally good
Boolean, e.g. *rhino or rhinoceros*	Both are equally good, assuming they both recognize the meaning of the Boolean operator
NP V-ing NP, e.g. *girl leading a horse*	NLIR shows some improvement
noun phrase, e.g. *black woman in a white hat*	NLIR shows major improvement; keyword retrieval scrambles modifiers randomly

2.3 Semantic Complexity

Semantic complexity is more difficult to evaluate, but we can make certain observations. Leaving noun phrases intact makes a text more semantically complex than deconstructing those noun phrases: *rubber baby buggy bumpers* is more semantically complex than a simple list of nouns and attributes, since there are various modification ambiguities in the longer version that are not present once it has been reduced to *rubber, baby buggy, bumpers* (or *rubber, baby, buggy, bumpers*, for that matter).

As for the names of people and locations, one could argue that the intact syntactic units (*Al Gore; George Bush; Arlington, Virginia; New England*) are semantically simpler, since they resolve ambiguity and eliminate the spurious readings *gore, bush, Arlington [Massachusetts], new England.* Nonetheless, we would argue that they are syntactically more complex when intact.
The PictureQuest system uses a WordNet-based semantic net to expand the caption data. To some extent, the syntactic measures (part-of-speech tagging, noun phrase pattern matching, name and location identification) serve to constrain the semantic expansion, since they eliminate some possible semantic expansions based on syntactic factors. One could interpret the word-list captions, then, not as syntactically less complex, but rather as semantically less constrained, therefore more ambiguous and thus more complex. This view would, perhaps, restore the more intuitive notion that complexity should lead to worse rather than better results.

3 Experiments

While the sentence captions are syntactically more complex, by almost any measure, they contain more information than the legacy word list captions. Specifically, the part-of-speech tagger and the noun phrase pattern matcher are essentially useless with the word lists, since they rely on syntactic patterns that are not present. We therefore hypothesized that our retrieval accuracy would be lower with the legacy word list captions than with the sentence captions.

We performed two sets of experiments, one with legacy word list captions and the other with sentence captions. Fortunately, the corpus can be easily divided, since it is possible to select image providers with either full sentence or word list captions, and limit the search to those providers. In order to ensure that we did not introduce a bias because of the quality of captioning for a particular provider, we aggregated scores from at least three providers in each test.

Because the collection is large and live, and includes ranked results, we selected a modified version of precision at 20 rather than a manual gold standard precision/recall test. We chose this evaluation path for the following reasons:

- Ranking image relevance was difficult for humans
- The collection was large and live, i.e. changing daily
- The modified measure more accurately reflected user evaluations

We performed experiments initially with manual ranking, and found that it was impossible to get reliable cross-coder judgements for ranked results. That is, we could get humans to assess whether an image should or should not have been included, but the rankings did not yield agreement. Complicating the problem was the fact that we had a large collection (400,000+ images), and creating a test subset meant that most queries would generate almost no relevant results. Finally, we wanted to focus more on precision than on recall, because our work with users had made it clear that precision was far more important in this application.

To evaluate precision at 20 for this collection, we used the crossing measure introduced in Flank 1998. The crossing measure (in which any image ranked above another, better-matching image counts as an error) is both finer-grained and better suited to a ranking application in which user evaluations are not binary. We calibrated the crossing measure (on a subset of the queries) as follows:

Measure	%Correct
Precision at 20 Images for All Terms	53
Precision at 5 Images for All Terms	59
Precision at 20 Images for Any Term	100
Crossing Measure at 20 Images	91

That is, we calculated the precision "for all terms" as a binary measure with respect to a query, and scored an error if any terms in the query were not matched. For the "any term" precision measure, we scored an error only if the image failed to match any term in the query in such a way that a user would consider it a partial match.

Thus, for example, for an "all terms" match, *tall glass of beer* succeeded only when the images showed (and captions mentioned) all three terms *tall*, *glass*, and *beer*, or their synonyms. For an "any-term" match, *tall* or *glass* or *beer* or a direct synonym would need to be present (but not, say, *glasses*). (For two of the test queries, fewer than 20 images were retrieved, so the measure is, more precisely, R-precision: precision at *the number of documents retrieved* or at *20* or *5*, whichever is less.

4 Results

We found a statistically significant difference in retrieval quality between the syntactically simple word list captions and the syntactically complex sentence captions. The word list captions scored 74.6% on our crossing measure, while the sentence captions scored 89.5%.

We performed one test comparing one-word and two-word queries on sentence versus word list captions. The sentence captions showed little difference: 82.7% on the one-word queries, and 80% on the two-word queries. The word-list captions, however, were dramatically worse on two-word queries (70.5%) than on one-word queries (89.7%).

	Simple Word List Captions	Complex Sentence Captions
Overall	74.6%	89.5%
1-word	89.7%	82.7%
2-word	70.5%	80%

5 Conclusion

Our experiments indicate that, in an information retrieval system tuned to recognize and reward matches using syntactic information, syntactic complexity yields better results than syntactically

mixed-up "word salad." One can interpret these results from a *semantic* complexity standpoint, since the syntactically simple captions all include considerably more semantic ambiguity, unconstrained as they are from a syntactic standpoint. This observation leads us to an additional conclusion about the relationship between syntactic and semantic complexity: in this instance, at least, the relationship is inverse rather than direct. The word-list captions are syntactically simple but, *as a result*, since syntactic factors are not available to limit ambiguity, semantically more complex than the same information presented in a more syntactically complex fashion, i.e. in sentences.

6 References

Bar-Hillel, Y., A. Kasher and E. Shamir 1967. "Measures of Syntactic Complexity," in *Machine Translation*, A.D. Booth, ed. Amsterdam: North-Holland, pp. 29-50.

Flank, Sharon, 1998. "A Layered Approach to NLP-Based Information Retrieval," in *Proceedings of COLING-ACL, 36th Annual Meeting of the Association for Computational Linguistics*, Montreal, Canada, 10-14 August 1998.

MacDonald, M.C. 1997. *Language and Cognitive Processes: Special Issue on Lexical Representations and Sentence Processing*, 12, pp. 121-399.

Rosen, B.K. 1974. "Syntactic Complexity," in *Information and Control* 24, pp. 305-335.