Chapter 3

Where Attitudinal Expressions Get their Attitude

Jussi Karlgren, Gunnar Eriksson, and Kristofer Franzén
Swedish Institute of Computer Science
Box 1263
164 29 Kista
Sweden
Email: {jussi | guer | franzen}@sics.se

Abstract

A number of attitudinal expressions are identified and analyzed using dependency based syntactic analysis. A claim is made that attitudinal loading of lexical items is dynamic rather than lexical and that attitudinal loading of individual lexical items is acquired through their use in attitudinally loaded structures.

Keywords: Attitude extraction, attitudinal expressions, dynamic attitudinal loading, emotion, perspective.

1. Research Questions to Motivate the Study of Attitudinal Expressions

The new field of attitude extraction is motivated primarily as an application area – as an area which will provide innovation to future generations of information access services. The study of attitude, expression, emotion, and perspective in text and discourse is interesting for other reasons as well. The link between form of expression, topical content, and referential processing is exceptionally clearly present in the examples discussed in this volume. In this paper we explore text to find attitudinal expressions, to find what characterizes such expressions, and to eventually understand them in the sense of being able to note what attitudes are held by whom as regards to what.

As a starting point, we use the notions of animacy and transitivity to formulate a methodology to probe the interface joining the formally defined but practically indivisible linguistic functions of syntax, semantics, and pragmatics. To do so reliably we need powerful tools for automatic linguistic analysis – we have recourse to quite competent morphological and syntactic analysis, but make do with simple algorithms for reference resolution and other higher level dependencies. Our aim is not to build a knowledge base, a lexical resource, or an ontology for the purposes of natural language processing. Our purpose is to understand how much information is couched in the form of linguistic expression and to explore and possibly expand the limits of processing that can be done using algorithmic rather than knowledge intensive methods.

2. Starting Points – Prototypical Attitudinal Expressions

A prototypical attitudinal expression has three constituents: an attitude, a target for the attitude, and an expressor. Someone has an attitude as regards something. This someone is always an animate agent. The object of the attitude is typically topical for the text. The attitudinal expression conforms in most cases to expectations for such expressions: the attitude is expressed through attitudinally loaded terms, using syntax which accommodates some of the three constituents and relates them to the topical frame of the narration or discussion at hand. Our claim in this paper is that attitude is not conveyed by lexical choice alone. We claim that attitude is largely expressed through the form of the utterance – with no requirement for any one lexical item in the utterance to be prototypically attitudinal. Lexical items are initially attitudinally loaded by virtue of their distributional history (see, e.g., Sahlgren (2002) or Sahlgren and Karlgren (2003) for a discussion on distributional semantics) but can be coerced to be more or less attitudinally loaded through their syntactic context. Similarly, experiments by Riloff and Wiebe (2003) make use of syntactic patterns to find subjective expressions, because lexical resources by necessity are incomplete and static in the face of the variety of emotional expression available to authors.

The experiments and examples given below focus on clauses with predicative complements. We believe that they are more often than other constructions loci for attitudinal expression. Our aim is to be able to find attitudinally loaded patterns with constituents identified and categorized from prototypical macro-patterns, specified on the syntactic and informational level.

It is W-ly X to Y.
Z believes that Y is very X.
The X really Y'd Z off.
An X seems quite often Y.

3. Text Topicality: Players

We model text topicality by *players* or *discourse referents*. Discourse referents – a theoretical concept since Coling 1969 (Karttunen, 1969), but hitherto not directly applied to information access technology – introduce a representation of text on a higher level of abstraction than terms are able to, and are text-internally and syntactically detectable, independent of text-external domain-specific knowledge bases. Identifying potential players in text (as opposed to entities that are mentioned non-topically, in passing) will need syntactic analysis, at least some initial steps towards anaphora resolution, a theory of topicality in text, and some statistical finesse. For our purposes in these experiments, we are interested in animate players with emotive potential, on the one hand and in topical players, on the other, as targets of attitudinal expression. The former have a central role in the attitudinal expression itself but need not be textually as important as the latter can be presumed to be.

We do not aim to push the envelope as regards identification of discourse referents themselves – the literature on how to identify and formalise discourse referents is plentiful albeit unproven in large scale processing experiments such as the ones we envision (e.g., Grosz et al., 1995; Sidner, 1979, 1986; Rich and LuperFoy, 1988; Fraurud, 1988). In the experiments presented in this paper we use a commercially available dependency based syntactic analysis tool to identify player candidates and filter the candidate set using text global term frequency calculations. Simpler syntactic analysis – adjacency based patterns, e.g. – would not, even for a fixed word order

language such as English, give us enough information to identify, e.g., the subject of a matrix clause with any reliability.

4. Text Topicality: Moves

Players, in our model of text topicality, engage in *moves*. Moves are primarily encoded in *transitive clauses* – briefly put, in clauses that describe action, with animate agents and well anchored in the discourse space (Halliday, 1967, 1968; Hopper and Thompson, 1980). Expressing attitude is one form of action, and a move in our model.

5. Identifying Players

Candidate players are found in a text using a combination of lexical and syntactic criteria. Our claim is that there are players of different kinds in a text, and that their different informational roles will have them occupying different syntactic functions. To exemplify, we run an experiment on a longish review or column of Michael Moore's recent movie "Fahrenheit 9/11", published in Slate in June 2004. In Figure 1, results from different combinations of criteria are displayed, ordered by descending frequency and truncated at the lowest frequencies.

The subject position, unrestricted, displays a large variety of nouns and names. Note that [Michael] *Moore* and [Richard] *Clarke*, George W. Bush's former chief of counterterrorism, are subjects much more frequently than is *Bush*. The object position shows the reverse. "*Film*", an ordinarily inanimate noun, follows the pattern of *Moore* in most of the sets.

Moving to predicate clauses, which we claim to be a prototypical locus of attitudinal expression, we find, in the third column, subjects to verbs of belief to be a much more clearly restricted set of players. The example article is a very opinionated text, with author as the main source of attitude; Michael Moore does not appear in this position. The fourth column, subjects of predicative clauses with adjectival complements, is a position we claim is indicative of the target of the attitudes. Here, we find *Moore*, *Osama*, *Laden* and *Iraq* – but no Bush, who is a side topic of the text! This shows how players can be usefully distinguished with respect to their syntactic function – or rather, their *moves* – in ways that are relevant to their attitudinal role.

6. The Case for Animacy: Adjectival Attributes and Genitive Attributes

The examples discussed in the previous section show how syntax mirrors the attitudinal role of the player. This is no coincidence. Certain types of player will occur more typically in certain syntactic positions. This is by virtue of their informational position and by their ontological status: strongly animate agents, typically human, active, and that stand out, are more likely to be in focus, both as regards attitudes expressed by the players in question or about them by other players or the author. In the following example, with data shown in Figure 2, we have selected a lexical item "*Clinton*", which can be expected to engender attitudinal expression and extracted adjectival attributes attached to it. By simple examination it is evident that the adjectives in the second column, derived in this fashion, have more pronounced attitudinal loading than the baseline adjectives shown in the first column.

Similarly, constructions that are marked for animate agents can be expected to hold attitudinal expressions more often than others. The third column shows adjective-noun constructions with a genitival modifier – the position marked X is reserved for animate agents. The fourth column

shows the same extraction pattern, but with an agent selected for high topical focus – as above, "*Clinton*". The adjectives in the fourth column are mostly and typically highly attitudinal. This demonstrates how the status of the referent is a useful selector for attitudinal items.

Subject nouns	Object nouns	Subjects of verbs of belief with predicative complements	Subjects to predicatives with adjectival complements	Noun heads of predicative complements	Noun heads of predicative complements with adjective attributes
		***We** believe it is difficult*	***Iraq** would still be the personal property of…*	*… is a **pacifist***	*… is a **brave** man*
Moore film Clarke war shot Saudi Saddam regime people meeting Laden Iraq half Fahrenheit civilian Bush Baghdad attack Afghanistan administration	film Bush army way vote troop removal president Orwell Moore line life courage coalition chance	we here there I	it this there that Moore Laden point Osama meeting Iraq he describe company capital	fact Moore film Afghanistan war time point Iraq way sort pacifist man word United States right interview family exercise day course Baghdad audience airport	man edifice cowardice book army

Figure 1. Examples of player candidates and filtered sets of players.

7. The Case for Syntactic Structure: Situational Reference

Texts abound with self reference, clause reference, situational reference and other types of meta-level references. Examples of such references are the pronoun *It* in: "*I kissed the ticket collector on the train yesterday. **It** was nice*." and the pronoun *That* in "*Sometimes there is no correlate. **That** is an annoying problem*." Most practically oriented studies on referential expressions gather such cases under the heading "situation reference". Resolving what the pronoun in the example above is referring to is at present problematic or near-impossible, but for the present purpose, collecting the attitude expressed towards them is not. In the examples above, we know that the author regards something as *nice* and something as *annoying*, even if we are unable to identify that entity.

Figure 2 gives in the first column the most frequent of all adjectives found in three months of Los Angeles Times newsprint; in the second column, the most frequent of adjectives that function as predicative complements; in the third, the most frequent of adjectives that are predicative complements to the three frequently situational pronouns *it*, *this*, and *that*.

Adjectives, all	Adjective attributes to "*Clinton*"	Adjective-noun constructions with any genitive attribute	Adjective-noun constructions with genitive attribute "*Clinton*"
New	early	X's executive director	Clinton's white house
good	encouraging	X's general fund	Clinton's strong commitment
high	former	X's good friend	Clinton's proposed alliance
big	standard	X's general manager	Clinton's tough talk
federal	actual	X's central bank	Clinton's proposed reform
american	agitated	X's young brother	Clinton's prominent role
orange	entire	X's general plan	Clinton's political quagmire
public	frequent	X's technical program	Clinton's federal budget
great	gregarious	X's national championship	Clinton's vehement response
own	high-ranking	X's valuable player	Clinton's strong defense
long	leaving	X's close friend	
national	longtime	X's advisory council	
former	now-famous	X's winless streak	
large	opportunistic	X's Vietnamese community	
local	outraged	X's super bowl	
small	proposed	X's short story	
same	real	X's Greek row	
old	regular	X's good player	
free	staunch	X's athletic director	
southern	underfunded		
major			
young			
white			
political			
late			
real			

Figure 2. Animate and focused heads accommodate more typically attitudinal adjectives. Data from one month of Los Angeles Times newsprint, first column sorted in descending frequency and truncated to fit.

Even from a cursory glance it is evident that the third column has more prototypical attitudinal adjectives than the first; the difference between the second and third is more open to discussion. Whatever the added value of the more stringent filtering criterion, it is clear that the position in predicative complement seems to be well established for attitudinal lexical items. Our first claim is that a predicative complement, especially in a situational reference setting, is a locus for attitude in text.

Further inspection of the second and third columns shows that, even in the top ten list of adjectives, there are items that are non-typical attitudinal items: *ready, likely, available, possible* are all examples of non-attitudinal lexical items. Our second claim is that the clauses and contexts they participate in may well be attitudinal in any case ("*Now comes the big one, and we're*

ready!"; "*Others simply are not ready to…*"; "*She's always fully available to help*."), in support of our contention that attitude is not simply a lexical question: no simple list of attitudinal terms will select attitudes from texts; no simple algorithm will allow us to draw up such a list, even provisionally.

All adjectives	Adjectives as predicative complement	Adjectives as predicative complement under *it*, *this*, and *that*
new	good	Hard
good	able	good
high	available	easy
big	hard	difficult
orange	sure	important
long	important	true
public	easy	bad
federal	ready	possible
own	likely	great
great	bad	nice
N = 35 000	N = 7 500	N = 2 500

Figure 3. Examples of how certain syntactic constructions are repositories of attitudinal expression.

8. Using Syntactic Patterns more Systematically

The earlier experiments showed that animacy and clausal structures, evaluated informally, seem to carry weight. To investigate this more formally, we formulated a number of progressively more restrictive criteria for identifying attitudinal expressions which we searched for in two sets of texts taken from the Wall Street Journal. One set (N=3398) is composed of editorials, opinion pieces, and letters to the editor; the other (N=3500) of reporting news articles, our assumption being that there are more overt expressions of attitude in the former set. The number of times the respective criteria matched in each text were calculated, normalized by number of clauses and are shown in detail in Figure 4. All reported results are statistically significant.

The first criterion we looked for was number of adjectives per clause, under the relatively weak assumption that attitudinal expression often finds its realization in adjectival form. For the second criterion we looked for occurrences of "*good*" or "*bad*" per clause – assuming these two most prototypical attitudinal adjectives would be noticeable. The third criterion adds syntactic constraints, looking for adjectives in a predicative complement – constructions such as "*It is good*" or "*... is a wonderful feeling*". The fourth criterion further looks to see if the complement is an object to another verb in constructions such as "*… believe this is a serious question*".

While the two evaluation sets admittedly are rather crudely fashioned, the results are unequivocal: there are clear stylistic differences between the two categories of text and this difference is better modeled using a syntactic distinguishing criterion. The most stringent object criterion does not seem to carry as strong a distinguishing power between the two sets as does the simpler predicative complement one, but is still clearly statistically significant.

Criterion	Opinion texts	Reporting texts	Rank sum
Adjectives per clause	1.75	1.64	12452777.5
"good" or "bad" per clause	0.0333	0.0131	12866718.5
Adjectives in predicative complement	0.354	0.183	14453290
Adjectives in predicative complement which is in object relation to other verb	0.0432	0.0368	12105678

Figure 4. Average number of occurrences of a attitudinal item or syntactic construction, for a collection of opinion pieces and reporting newspring, respectively. All differences significant ($p > 0.95$) by Mann Whitney U rank sum test, criterion level 11857416.2673481.

9. Generalizing from Syntactic Patterns to the Lexicon

A further examination of differences between the two sets defined above show how the sets of adjectives differ between sets and between the different contexts examined. The top ten adjectives in each condition are shown in Figure 5. The two document sets have slightly differing ranking of adjectives used – but six out of the top ten are shared between the document sets. In the most stringent criterion the overlap is slightly lower – only five out of the top ten are shared. The top ten adjective lists give the same impression as the lists in Figure 2: the more restrictive contexts seem to be loci for attitudinal expression, and the expressions found give purchase to our claim that adjectives gain their attitudinal loading from being used in certain expressions. This would also seem to point at the potential for mining attitudinally loaded items from expressions, if some statistical finesse is observed.

10. Conclusions

Our conclusion at this juncture is fairly abstract, but well supported by the data in our preliminary experiments reported here. The expression of attitude is done through a combination of syntactic and lexical means – most lexical items and constructions that participate in attitudinal expressions are typically also found in non-attitudinal expressions. Our further claim is consistent with the data but as of yet unproven – that lexical items gain attitudinal loading from the contexts they participate in. On a more theoretical note, we conclude that to explore the expression of attitude in text, it is necessary to explore the interaction of syntax and pragmatics – lexical resources by themselves will not be sufficient, be they static or dynamic.

	Opinion texts	Reporting texts
Adjectives	new good political american public economic federal national own high	new high federal big large good american national low major
Adjectives in predicative complement	good able true political important new likely high bad wrong	high good likely able big new low large strong bad
Adjectives in predicative complement which is in object relation to another verb	good able bad wrong important necessary willing only new easy	good likely high able strong big bad new low willing

Figure 5. Examples of adjectives in predicative complements, with and without object verb criterion.

11. Bibliography

Fraurud, K. (1988) Pronoun Resolution in Unrestricted Text. *Nordic Journal of Linguistics*. 11, 47-68.

Grosz, B., Joshi, A., and Weinstein., S. (1995) Centering: a Framework for Modelling the Local Coherence of Discourse. *Computational Linguistics*, 21 (2), 44-50.

Halliday, M. A. K. (1967) Notes on Transitivity and Theme in English I. *Journal of Linguistics*, 3(1), 37-81.

Halliday, M. A. K. (1967) Notes on Transitivity and Theme in English II. *Journal of Linguistics*, 3(2), 199-244.

Halliday, M. A. K. (1968) Notes on Transitivity and Theme in English III. *Journal of Linguistics*, 4 (2), 179-215.

Hopper, P., and Thompson., S. (1980) Transitivity in Grammar and Discourse. *Language* 56, 251-99.

Karlgren, J., and Sahlgren, M. (2001) From Words to Understanding. In Uesaka, Y., Kanerva, P. & Asoh, H. (Eds.): *Foundations of Real-World Intelligence*, 294-308, Stanford: CSLI Publications.

Karttunen, L. (1969) Discourse Referents. *Proceedings of the Third International Conference on Computational Linguistics* (COLING). International Committee on Computational Linguistics and KVAL research group on quantitative linguistics.

Rich, E., and LuperFoy, S. (1988) An architecture for anaphora resolution. *Proceedings of the Second Conference on Applied Natural Language Processing (ANLP-2),* 18-24.

Riloff, E., and Wiebe, J. (2003) Learning Extraction Patterns for Subjective Expressions. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-03)*, 105-112.

Sahlgren, M. (2002) Towards a Flexible Model of Word Meaning. In Karlgren, J. (Ed.) Papers from *Acquiring (and Using) Linguistic (and World) Knowledge for Information Access*, AAAI Spring Symposium 2002, Technical Report SS-02-09. AAAI Press.

Sidner, C. L. (1979) Towards a Computational Theory of Definite Anaphora Comprehension in English Discourse. Technical Report No. 537. M.I.T., Artificial Intelligence Laboratory.

Sidner, C. L. (1986) Focusing in the Comprehension of Definite Anaphora. In Grosz, B., Sparck Jones, K., and Webber, B. (Eds.) *Readings in Natural Language Processing*., Morgan Kaufmann Publishers.