Political Positions and Political Cleavages in Texts

Jan Kleinnijenhuis and Wouter van Atteveldt

If humans are political animals, and language is their most versatile communication tool, then the old question what should be extracted from political texts to understand politics deserves renewed attention. Recent advances in the information and communication sciences have resulted in new means to process political texts, especially advances in the domains of knowledge representation, natural language parsing, automated content analysis, and semantic web technologies. Applying these innovations to uncover what matters in politics is far from trivial, however. The aim of this chapter is to give an introduction to an analysis of political texts aimed at inferring their political meaning.

First, a few concepts from the history of political thought will be reviewed, so as to arrive at a feeling for what should be revealed by means of an analysis of political texts. Next, three sets of Natural Language Processing tools will be distinguished to analyze political texts: tools to assess whether a concept (or 'object') occurs, tools to assess whether an (a-symmetric) relationship between concepts occurs, and tools to assess the nature of the relationship between concepts. These three sets of tools relate respectively to advances in ontology construction and entity recognition, advances in statistical associations and network theory, and advances in part-of-speech tagging and grammar parsing. The aim is to show how automation of the three sets of tools could be employed in the near future could give reliable and valid answers to frequently asked questions in political communication.

Political language and content analysis

Webster's Unabridged dictionary defines *politics* in various ways. Politics is "the total complex or interacting and usually conflicting relations between men living in society", it is concerned "with governing or with influencing or winning and holding control" and with "actions, practices or policies" to achieve goals with respect to issues. In short, politics is usually about achieving goals by means of policies, about conflict or cooperation, and about winning or losing.

Machiavelli (1469-1527) observed that princes can acquire new principalities by means of their own virtue, or by fortune. Their *virtue* prescribes them to take great pains to satisfy the people and make them content with laws that serve their issue interests, but also to raise prestige by the art of conflict and cooperation, thus by revealing themselves without any preservation in favor of one side against another, with the risk of vigorous wars. *Fortune* is what befalls on leaders, such as unexpected natural disasters, or popular support. Virtue can be practiced, whereas fortune can only be anticipated. Machiavelli's primary concern is political language rather than politics itself, that is, how others would *speak* about the acts of actors and about what befalls upon them. Even winning a war was perceived by

Machiavelli as the art of inspiring the army with confidence in itself and in its general, thus as the art of inducing others to speak about themselves in relationship to their leader. Interpreted from the contemporary perspective of this chapter, already Machiavelli elucidated that political language centers around four types of statements: [1]

Virtue: what an actor can do:

1. take an issue position (pro or con a cause, e.g. poverty, crime, unemployment);
2. cooperation or conflict with, support or criticism from other actors (e.g. building a coalition government, waging war)

Fortuna: what befalls upon an actor:

3. real world developments with regard to issues (e.g. famine, unemployment)
4. success or failure, gains or losses of actors (e.g. gains and losses in wars, opinion polls, or in political debates).

Machiavelli is considered as the forerunner of the age of absolutism, but his basic ideas are equally important for a democracy. In a democracy, each of these four types of statements is important to attract voters. Voters choose a party when they agree with the issue positions of that party (Tomz & Van Houweling, 2008; Westholm, 1997), and even when the issues on which that party holds a strong reputation dominate the campaign (Budge & Farlie, 1983; Hayes, 2005; Petrocik, 1996). Voters choose a party with a strong profile in terms of attacks and criticisms from political adversaries and support from within and from societal actors (Shah, Watts, Domke, & Fan, 2002). They prefer incumbent parties in case of favorable real world developments (e.g. economic growth) and a challenger in case of deteriorating real world developments (e.g. unemployment)(Schumpeter, 1944), but in the absence of objective knowledge they are susceptible to the positive or negative portrayal of these developments in the media (Hetherington, 1996; Sanders & Gavin, 2004; Soroka, 2006). Of these four driving factors, the attribution of successes and failures to parties in the media is the most important predictor of shifts in party preferences (Kleinnijenhuis, Van Hoof, Oegema, & De Ridder, 2007).

Rather unsuccessful attempts have been made to exclaim one of these four types of statements as the most fundamental one. In their seminal study of the historical origins of party systems and voter alignments Lipset and Rokkan observed that both parties and the loyalty of voters to these parties rest on old conflicts and cleavages (Lipset & Rokkan, 1967), for example between workers and owners (between the poor and the rich, the have-nots and the haves), or between the permissive, secular, urban groups and orthodox religious, rural groups (e.g. in the Dutch context the "rekkelijken" versus "preciezen" in the sixteenth and seventeenth century). From the perspective of Lipset and Rokkan disagreement about issue positions mirrors the historical conflicts and cleavages between actors. A central claim in Marxist theories is that that political issue positions simply mirror class cleavages, i.e. one particular type of cleavages. An opposed view is apparent from the literature about issue voting and election campaigns. Disagreement about issues tends to be seen as the heart of politics. Attack politics and news about conflicts between parties are either regarded as a mirror of issue (dis)agreement, or as degenerated forms of political communication that merely enhance political cynicism (Ansolabehere &

---

[11] This interpretation of Machiavelli is based especially on Chapters 6-7,17-18,21 and 25 of *The Prince* and chapters 33-35 of *The Discourses*.

Iyengar, 1995; Cappella & Jamieson, 1997; Patterson, 1993). The intellectual origins of this opposed view date back to ideas about political apathy of Alexis de Tocqueville (De Tocqueville, 1840 (1951, 8e édition)), that were revitalized in the theory on the *spiral of silence* of Elisabeth Noelle-Neumann (Noelle-Neumann, 1980) and also through the work of Jon Elster (Elster, 1993). The comparative research literature on party manifestoes (Budge, Robertson, & Hearl, 1987, cf. in this volume the chapters of Collette and Petry, of Kaal, Van Elfrinkhof and Maks, and of Krouwel) also contributed to the view that issue positions drive cooperation and conflict between parties. As Kaal, Van Elfrinkhof and Maks point out, a striking feature of the genre of party manifestoes is that they deal solely with the issue positions of a single party, rather than, for example, with the conflicts and the teamwork in that party that produced a particular manifesto, or with the pattern of conflicts and cooperation that will emerge in election campaigns, in government coalition negotiations, or within the next government coalition. Remarkably little longitudinal empirical research has been done, however, to verify whether one "fundamental" type of statements is systematically "mirrored" in other types of statements indeed.

The simple truth may be, that not a single causal order is impossible or even unlikely. Disagreement about a topic is sometimes the result of conflicts between actors, but conflicts between actors may also follow from disagreement about a topic. Sometimes the two may be unrelated, for example when a political party attacks an ideological similar party, or induces third parties to attack or to neglect an ideological similar party, because precisely ideological similar parties are serious electoral competitors. Winning or losing may be the outcome of conflicts, but vesting a reputation as a loser may also cause conflicts, as Machiavelli stated over and over again. Longitudinal research is needed to answer the important research question which causal order is likely under which conditions, but to arrive at relevant data for this type of research new tools are required to analyze the emphasis, turns, shifts and moves in political language, both comprehensively and in-depth.

Meta-language about political language

More than two thousand years ago the old Greeks invented political theatre, political dialogues and democracy. We still inherit important concepts to talk about political language from them, such as sign and signifier, symbols, and last but not least the subject-object-predicate triple. Although the Greek concepts of subject, object and predicate are ambiguous and outdated from the point of view of today's theories about grammar, logic and semantics, they are still worthwhile to discuss where language is all about.

The Australian linguist Robert M.W. Dixon observes that in all languages sentences deal with a *subject* who or what directs its action or energy towards a target or *object* (Dixon, 1992, 2005). The nature of this action or energy is a two-place *predicate*. The subject and the object are either animated entities, which we will label *actors*, or circumstances and other non-animated entities, which we will label as *issues* here, although in non-political context labels such as *variables, circumstances* or *states of affairs* would presumably be more intuitive. Subject-predicate-object-triples resemble the a-symmetric *xRy*-triples in relational logic, which was pointed out succinctly by Ludwig Wittgenstein (1989-1951) in his famous statement: "Namen gleichen Punkten, Sätze Pfeilen, sie haben Sinn" ("names resemble points;

propositions resemble arrows, they have sense" (Wittgenstein, 1922: 3.144)). The boundaries of one's propositions would also be the boundaries of one's world, according to the early Wittgenstein, but in his later work he was precisely interested in the exchange and the misunderstandings between different world views (Wittgenstein, 1953), thereby recognizing the source of propositions as an integral part of the proposition itself (effectively *s :xRy*, in stead of merely *xRy*, in which *s* represents the source, *x* the subject, *R* the predicate, and *y* the object).

Fritz Heider (1896-1988), another Austrian who moved into the Anglosaxon world, developed balance theory in a remarkably short paper. Balance theory deals with triangles of three statements, like *xRy*, x*Sz*, and x*Tz*. Balance theory predicts that usually the third relationship can be predicted correctly from the first two (*T* from *R* and *S*), on the basis of the principle that friends of friends, but also enemies or enemies, tend to be friends, whereas enemies of friends, as well as friends of enemies, will usually be enemies. Balance theory, and later theories of cognitive consistency, such as congruence theory and the theory of cognitive dissonance, hold that people will try to avoid cognitive representations that violate balance by a number of Freudian escape routes, such as the negation of information, blaming the messenger, or the rationalization of previous choices with ingenious new arguments (Severin & Tankard, 2005).

Charles Osgood was the first to develop a coding instruction to extract *xRy* - statements from full sentences and complete texts (Osgood, Saporta, & Nunally, 1956). Many different elaborations of this method have been proposed (Deetjen, 1977; Kleinnijenhuis, 2008; van Cuilenburg, Kleinnijenhuis, & de Ridder, 1986). Since *xRy*-statements build up a network, the enterprise to extract them from texts has been labeled as *semantic network analysis* (Krippendorff, 2004; van Atteveldt, 2008), which is the topic of the next paragraph.

Elementary *xRy*-statements may look all too familiar from the point of view of contemporary logic as applied in semantic web approaches (Antoniou & van Harmelen, 2004), but it should be pointed out that in logic the focus is usually on deriving theorems from sets of axioms in which the predicate **R** is invariant between axioms, or at least predicates **R**, **S**, **T**, … differ categorically from each other. The predicates in political language, and in theories of balance and cognitive consistency can be mapped onto a positive-negative continuum, however. In contemporary logic, axioms that give rise to contradictions are deemed untenable and therefore uninteresting, whereas in cognitive consistency theories the primary diagnostic of belief systems is the degree to which they are unbalanced and therefore ambiguous when it comes to drawing inferences. In most logics statements are either true or untrue, whereas in political language statements have both a magnitude (a frequency, a saliency), and an angle (a direction, a continuous positive-negative scale): they mount up to a vector model (Kleinnijenhuis & Pennings, 2001). Moreover, the direction of a predicate has not only a mean interpretation as its first statistical moment, but also higher order moments, such as a variance, a skewness and a kurtosis. In short, contemporary logic aborts where political language, political dialogues and cognitive consistency theories start, namely after a contradiction and after a variety of interpretations.

Three tools for analyzing political texts

Although the aim of this article is to address whether political meanings can be inferred from texts by the automation of the extraction of subject-predicate-object-triples from texts, we will start with two more basic questions. Is it feasible to automate a textual analysis to extract the occurrence and the co-occurrence of concepts in texts? From a theoretical perspective the latter questions are equivalent to the questions whether it is possible to automate first-order and second-order agenda setting research.

Does a concept occur? First-order agenda Setting and entity recognition

Agenda setting (McCombs, 2004; Rogers, Dearing, & Bregman, 1993), at least first-order agenda-setting, is concerned with the *transfer of issue saliency* on the agenda of one actor to the agenda of another actor. Agenda setting theory predicts, for example, that a huge attention for a particular issue in the media will be followed by huge public attention for that issue. A transfer of issue saliency is also at the heart of the issue ownership theory, which attempts to explain party competition. The issues that are prominent in party manifestoes will tend to become the issues that are central in election campaigns. Pre-established reputations of parties with respect to the issues that dominate the campaign, that may well go back to the old cleavages between parties, determine which party will win at the elections, and ultimately also the policy emphasis and policy expenditures of parties in government. In short, party competition is primarily directed at increasing the *emphasis* on owned issues, rather than at pro-or con arguments (Budge & Farlie, 1983; Budge, Klingemann, Volkens, Bara, & Tanenbaum, 2001; Hayes, 2005; Petrocik, 1996). In mediatized democracies, politics is basically the *politics of attention* (Jones & Baumgartner, 2005).

If the emphasis on issues and attributes could be measured with single word lists or with elementary boolean search strings, then it would be easy to automate agenda setting research, but this is often not the case. However, ontology matching may help to automate agenda setting research nevertheless.

The ontological problem: (named) entity recognition

Let's start with an example article to elucidate the ontology matching problem that will be encountered when one tries to count the occurrence of concepts, rather than the occurrence of words. The example article is compiled by the authors from a number of available articles about the Israeli-Palestinian conflict.

1. **Time running out for Mideast two-state solution**
2. Only a few months ago, president Obama welcomed Israeli Prime Minister Benjamin Netanyahu and Palestinian Authority President Mahmoud Abbas to the White House.
3. President Obama said both leaders came to Washington in an effort to restart the peace process and reach the goal of a two-state solution that ensures the rights and security of both Israelis and Palestinians.
4. Obama told reporters that the Israeli government and the Palestinian Authority had taken important steps to build mutual confidence since May.
5. Since then, president Abbas stated repeatedly that only a complete Israeli settlement freeze would create the conditions for a return to the negotiating table.
6. Today not only Israel, but also the United States dropped the Palestinian demand for a settlement freeze that would have opened up negotiations shortly.

7. If the past two years have shown nothing else, it is that the weak Ramallah government will not realize enough success to help lead the path back to negotiations that bring about a two-state solution.
8. No two-state solution is possible without Hamas, but Israel and the United States do not want to negotiate with Hamas.
9. Hamas leader Khaled Mashal called for continuing the jihad against Israel and categorically denied any possibility of talks with Israel.

Readers who understand this article, must have had an *ontology* in mind existing concepts and their relations . Ontology is the study of the things that are, and *an* ontology is a name used in Knowledge Engineering to denote a (shared) formalization of a view on the world. Table 1 presents a simple strictly hierarchical ontology (or taxonomy) to match the content of the example article.

Table 1: An ontology to match the content of the example article

Actors
    Palestinians
        government: Palestinan Authority
        leader: President Mahmoud Abbas
        capital: Rahmallah
    United States
        leader: President Barack Obama
    Hamas
        leader: Khaled Mashal
    Israel
        leader: Prime Minister Benjamin Netanyahu
Issues
    peace process, Israeli-Palestinian negotiations, peace talks
    two-state solution
    Israeli settlements
        settlement freeze [-]

An easy observation is that pre-established ontologies are indispensable, since otherwise the article would become incoherent (what has Ramallah to do with Abbas?). Explicit ontologies enable automation of agenda setting at the level of concepts, and not just only at the level of words or boolean search strings of words. From an automation perspective the advantage of an ontology is its additivity. Frequency counts of lower-order concepts (e.g. Abbas, Mashal) are sufficient to arrive of counts of higher order concepts such as the Palestinians or Hamas. Antonyms and concepts with an opposed meaning give rise to a complication. Lower order antonyms of higher order concepts are marked with a [-]-mark. The [-]-mark in the example ontology means that attention for a settlement freeze can be counted as attention for Israeli settlements, but also that protagonists of a settlement freeze should be counted as *antagonists* of Israeli settlements, and vice versa.

Research on election campaigns or government policy ontologies typically deals with 500 up to 2500 lower order concepts. A lower number of concepts is often desirable when it comes to interpretation and dispatches. For the latter ends the concepts may be mapped in accordance with the rules implied in the ontology to 5 up to 25 concepts of interest.

Co-occurrence op concepts? Conditional probabilities and associative framing

Second order agenda-setting maintains that the attributes that are associated with an object in the media to which one is exposed will become also the attributes that will be associated with the concept by the audience (McCombs, 2004). Second order agenda setting rests on the transfer of saliency, as did (first-order) agenda setting.

The question to be answered next is what makes that a concept B (dotted) is attributed to, or associated with a concept A (solid) ? In his seminal article *Features of Similarity* Amos Tversky points out that the reason why the smaller set lets you often think of the larger set is simply that the number of elements in the intersection of two sets is a larger percentage of the number of elements in the smaller set than of the number of elements in the larger set (Tversky, 1977). See Figure 1: Conditional probabilities and a-symmetric associative framing.
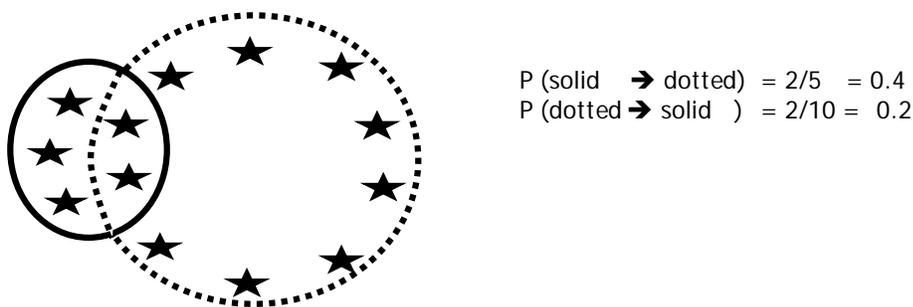
P (solid ➔ dotted) = 2/5   = 0.4
P (dotted ➔ solid) = 2/10 = 0.2

Figure 1: Conditional probabilities and a-symmetric associative framing

If you think of kerosine, you may come to think about an airplane, but it's quite unlikely that thinking about an airplane will immediately generate thoughts about kerosine. Planes are more strongly associated with visions of holiday or congress destinations. In the study of language, the size of sets is equivalent to the occurrence of concepts, whereas their intersection is equivalent to their co-occurrence.

After 9/11 politically correct journalists attempted to show that most Muslims (a large set) were actually not terrorists (a small set) by giving examples of hard working integrated Muslim immigrants, but this did not prevent that thoughts about terrorism would still generate pictures of 9/11 and negative ideas about Muslims. Conditional probabilities give the cue to associative framing of topics in the media, for example of terrorism and the Islam (Ruigrok & van Atteveldt, 2007). In Spain, however, even the Madrid train bombings by Muslim terrorists shortly before the national elections in 2004 were exploited by the governing Partido Popular as new evidence of ETA violence. In spite of the Madrid bombings, in spite of the protests against the PP-propaganda, and in spite of the victory of Spain after almost a millennium of servitude to Muslim invaders, the Spanish press from 2000-2008 never came to associate terrorism strongly with the Islam. The primary association of terrorism remained the ETA. Primary associations in the press of immigrants were not the Islam, or terrorism, but rather the economy and the "calling effect" of the regularization of immigration (Mena-Montes, 2010). Associative framing can be automated fully, since techniques to count occurrences and co-occurrences of words is straightforward,

whereas progress in ontology matching enables also counting of the occurrence and co-occurrence of higher-order concepts.

Many scholars have been puzzled by conditional probabilities, for example John Maynard Keynes, who wrote in 1929 after his renewed encounter with Ludwig Wittgenstein in Cambridge: "Well, God has arrived. I met him on the 5:15 train", thereby referring to proposition 5.15 (Wittgenstein, 1922), one of Wittgenstein's propositions about conditional probability in terms of sets. Conditional probabilities are also at the heart of Bayesian statistics. Many theorems from Bayesian statistics have an analogue in associative framing.

If the frequency of concept occurrence per textual unit is known, it is simple to compute the asymmetrical associations or co-occurrence between concepts. These associations can be conceptualized as the conditional probability of encountering one concept given that another object is encountered: given that a sentence contains a reference to Hamas, how likely are we to see a reference to Israel? This conditional probability is the association between Hamas and Israel. Taking the sentence as the contextual unit, in the example text this probability is 100% as both sentences mentioning Hamas also mention Israel. The reverse, however, is not true, as only 2 of 6 sentences mentioning Israel mentions Hamas, making the association between Israel and Hamas 33%. Figure 2 shows the network of all associations greater than 50% as extracted from the example text. The lower right shows a central cluster of strongly interconnected actors: Israel, USA, and Palestinians. Interestingly, Palestinians and Hamas are not associated with each other at all. Moreover, while both are associated with the Peace Process and the Two-state Solution, these issues are not associated with Hamas and the Palestinians, but rather with Israel and the USA. If this article would be representative of Middle East reporting, one would expect that people think of the USA and Israel when they think of the peace process, and not of Hamas or the Palestinians.
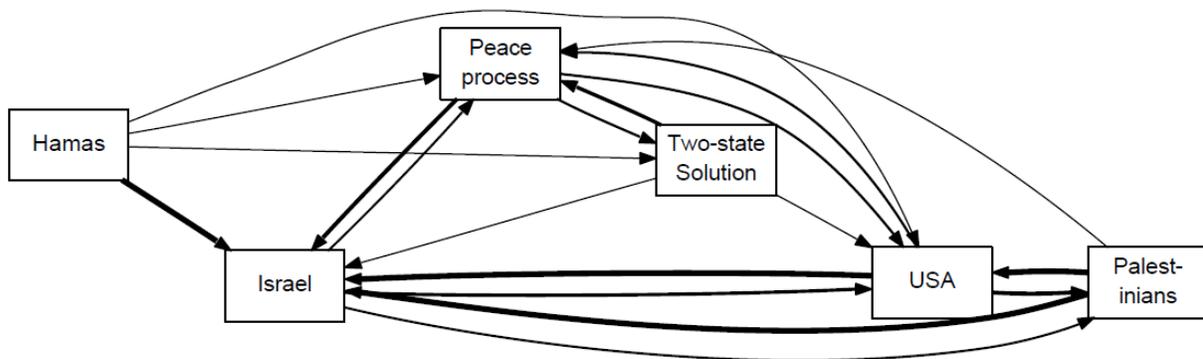

Figure 2: Association network extracted from the example text

Semantic network analysis
In the examples above we have shown how the occurrence and co-occurrence of words can provide meaningful information on the agenda and associations of relevant concepts. These techniques do not

tell us, however, how these concepts are related. Fully understanding the relations expressed in language requires fully understanding both the intricacies of natural language and the context in which the language is to be understood. This is beyond the capabilities of the computer. However, it is possible to employ grammatical analysis to analyze some of these relations. First, we should understand how humans would analyze the example text on the two-state solution.

Manual coding using the NET-method

Presumably the most straightforward way to understand the extraction of political statements from political texts is to present the extracted statements in a network. Figure 3 represents the statements in the example text about the two-state solution that are attributed to president Abbas and to president Obama. whereas Figure 4 represents the network according to the compiler of the article. In Figure 3 and 4 lower-order concepts (e.g. Obama, Abbas) are mapped In accordance with the tree-structure of the ontology to higher order concepts (e.g. USA, Palestinians). Solid arrows represent positive associations, dashed arrows represent negative associations. The arrow labels include the quoted actor, an abbreviation for the type of statement, the sentence number in the example article on which the arrow is based, and a few crucial words from the predicate that make clear why these relationships are positive or negative. Table 2 gives an overview of the abbreviations used.

Table 2: Statements types and their abbreviations

	Abbreviations of statement type	Subject/agent	object/target	Example
2-place predicates	IP: issue Position	actor	Issue	Obama freezes bonuses
	CC: conflict/cooperation, Support / Criticism	actor	Actor	Palin unleashes attack against Obama
	CSQ: consequences	issue	Actor	Bonuses are simply good for bankers
	CAU: Causation	issue	Issue	Bonuses help the economy
1-place predicates	REA: Real World developments	reality	Issue	Bonuses rose further in 2011
	SF: Success / failure	reality	Actor	Obama has lost heavily
	AEV: Actor Evaluations	actor	Ideal	Obama is doing a great job
	IEV: Issue Evaluations	issue	Ideal	Bonuses are obscene

Two additional statement types that were not discussed before pop up in the example article: CSQ (=consequences of issues for actors) and CAU (=causal relationships between issues).
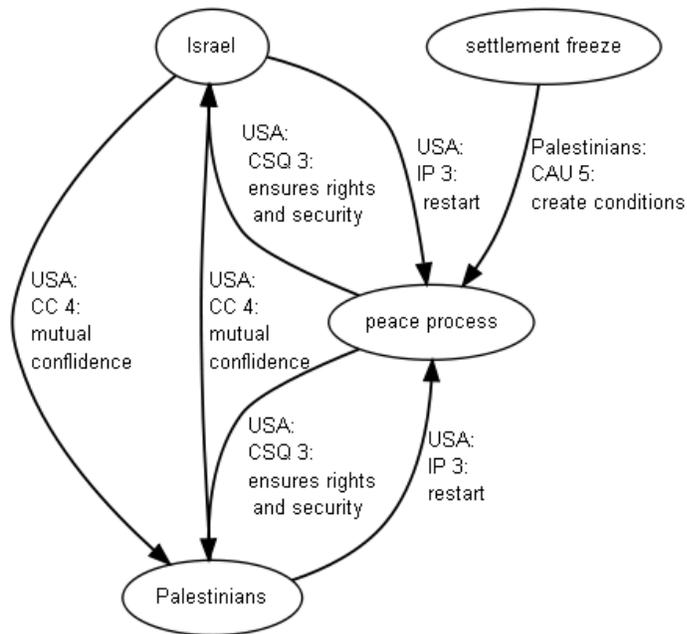
Figure 3: Quotations from Obama and Abbas, attributed respectively to USA and Palestinians

Figure 3 shows that according to Obama Israel, the Palestinians and the peace process are positively associated with each other in each direction. They mutually trust each other, both can benefit from the rights and security delivered by the peace process, and both want to restart the peace process. According to Abbas, however, a settlement freeze is a precondition for the peace process.
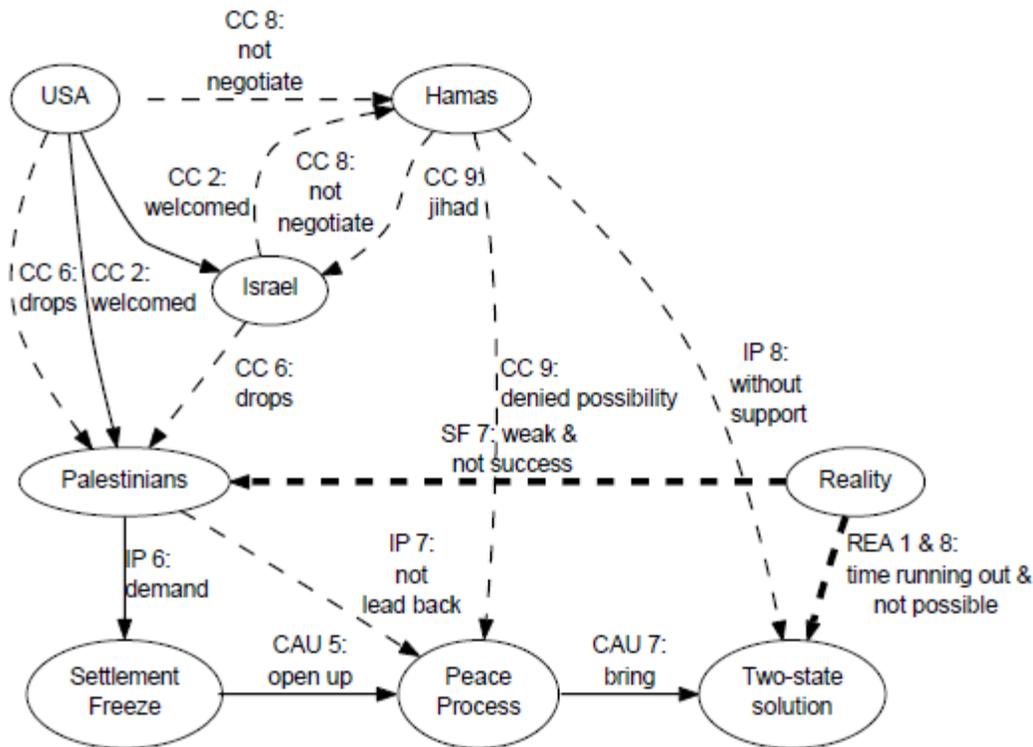
Figure 4: Statements in the example news article on behalf of its author

In contrast with Figure 3, also negative relationships show up in Figure 4. Actually, only a few relationships are positive. The USA welcomed the Palestinians, who demanded a settlement freeze, which could open up the peace process, which could bring about a two-state solution. In line with the transitivity principle, one may conclude from this chain of reasoning implies that the USA furthers a two-state solution. Other chains of reasoning undergird this conclusion. The USA do not want to negotiate with Hamas, since Hamas lends no support to a two-state solution, and denies the possibility of a peace process that could bring about a two-state solution. The USA drop the demands of the Palestinians, and welcome Israel, which also drops the demands of the Palestinians, since the Palestinians will not lead back themselves to the peace process, which could bring about the two-state solution.

The example article gives nevertheless an inconsistent, unbalanced view (Heider, 1946) of the USA position. The unbalance hinges on two inconsistencies. The Palestinians demand a settlement freeze that could open up the peace process, but they will not lead the path back to negotiations, according to the author of the example article. The USA welcome the Palestinians, but also drop their demands. Given these inconsistencies, one may also argue that the USA now rejects a two-state solution. For example, the USA welcomed Palestinians who will not lead back themselves to the peace process. More importantly, they dropped Palestinian demands although these demands for a settlement freeze could have opened the peace process, which could have brought about a two-state solution.

In summary, a semantic network analysis aims at inferring political meanings from texts by highlighting the chains of reasoning in texts, as well as their consistency or inconsistency. The semantic network analysis of the example article clearly reveals that the example article throws the ball into the

Palestinian court, since the Palestinian demands for a settlement freeze, rather than the Israeli settlements are portrayed as the major obstacle for the peace process.

Automation using semantic rules on top of an ontology, POS-tags, syntax dependency trees and a sentiment analysis of predicates

As stated above, we will show how grammatical analysis can be used to automatically extract part of the network as extracted by human coders. In particular, we will extract citations (sources) and semantic subject/predicate/object triples. Grammatical analysis yields `syntax trees', graphs containing the grammatical relations between the words of a sentence. For this example, we used the freely available Stanford parser to parse the sentences listed above (Klein & Manning, 2003). In other cases we have used the (also freely available) Alpino parser for Dutch with similar techniques (Van Noord 2006, Van Atteveldt *et al* 2008).

The key intuition behind using syntax trees is that these trees are closer to the (semantic) relation we wish to measure than the 'raw' words of the sentence. As an example, consider the sentences "John hits Mary" and "Mary, who has been the victim of domestic violence before, was hit by John". Both sentences express a hitting relation between John (the hitter) and Mary. However, the 'surface' structure is very different, with many (for this relation) irrelevant words in between John and Mary in the second example and the order of John and Mary in the sentence reversed. As will be shown below the grammatical structure of these sentences will make it clear that the relative clause (", who has been ..") is not central to the expressed relation and that the second sentence is in the passive voice.

We use the grammatical structure by defining rules that match specific patterns in the syntax trees. The concepts occurring in the relevant parts of the patterns are then translated to semantic roles between these concepts. To illustrate this, Figure 5 shows the annotated parse tree of the fifth sentences from the example above. The words in italics are the words from the text, with the labels on the edges indicating the grammatical relations between them. *President Abbas*, for example, is the subject of *stated*, while the whole subtree under *would create* is the complement of that verb.[2] As can be seen from this graph, 'irrelevant' modifiers such as *repeatedly*, *only*, and *since then* are no longer in between the predicates and their subjects, bringing the grammatical structure indeed closer to our intended semantic structure. Moreover, grammatical relations such as 'subject' are related (but not identical) to the semantic relations, giving hope that moving from grammar to semantics might be doable.

---

[2] Note that noun phrase (NP) chunks were collapsed to simplify the graph. Normally, each word would be a single node, with for example *Abbas* being the subject of *stated* and *President* a modifying node under *Abbas*.
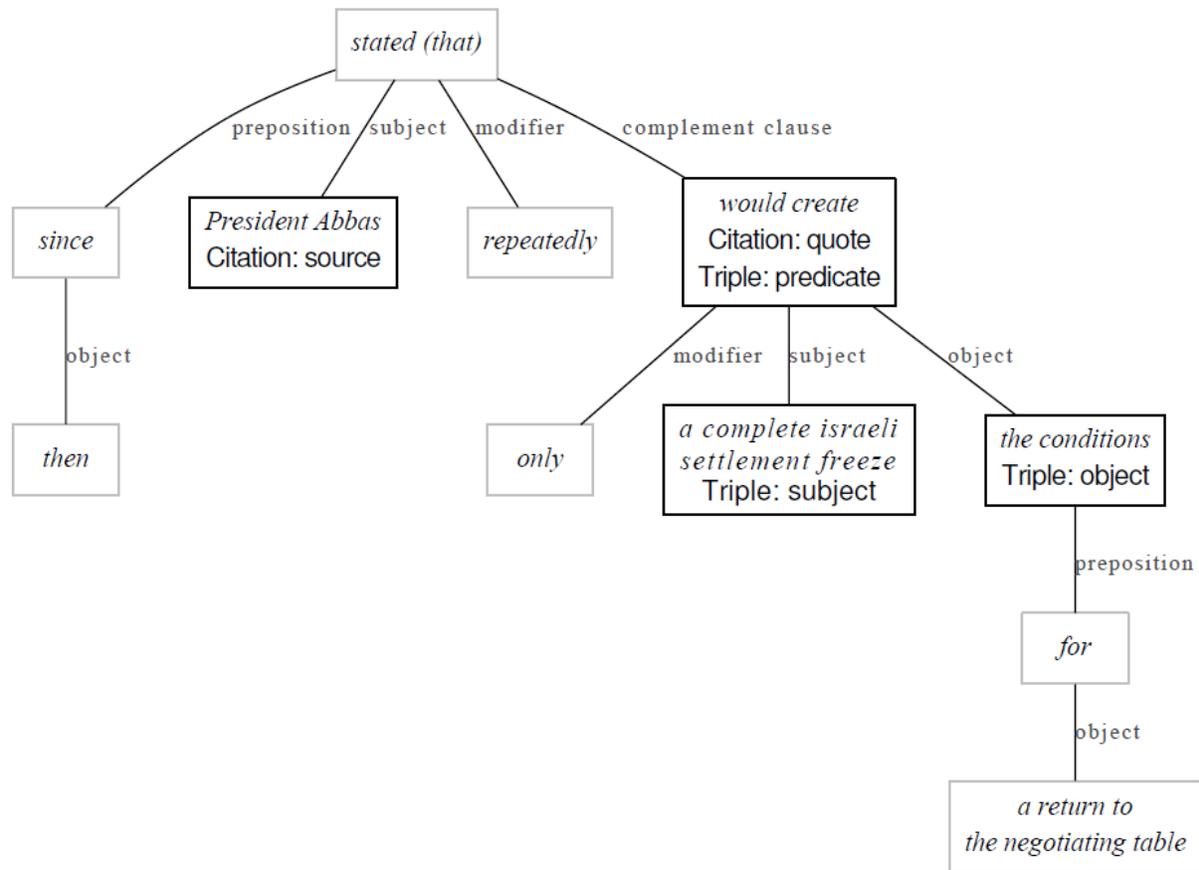
Figure 5: Semantic tree of sentence 5:
" Since then, president Abbas stated repeatedly that only a complete Israeli settlement freeze would create the conditions for a return to the negotiating table."

In order to move from grammar to semantics, we have defined a (relatively) small number of rules that match patterns on the syntactic tree. Figure 6 contains a list of the four rules that are used in the examples here. In the current example, rule 1 matches the *stated* verb as it is a speech act verb and it has a subject and complement. Thus, a **Citation** is created with *President Abbas* as source and *would create* (and all nodes below it) as quote. Rule 2 matches the verb *would create*, as it is not a speech act and has a subject and object. Thus, a **Triple** is created with the *settlement freeze* as subject, the *conditions* (again including underlying nodes) as object, and the verb *could create* as predicate.

If we use the same ontology to identify the concepts of interest, and if we can detect that create is a positive relation, we can reduce this citation containing a triple to a s: xRy relation [Palestines: SettlementFreeze + PeaceProcess].

Rule 1: Citation

Pattern: a speech act *verb* (state, say, …) with a *subject* child node and a *complement* child node

Result: **Citation(source=**subject*, **quote**=complement)*

Rule 2: Active verbs

Pattern: a non-speech act verb with a *subject* child and an *object* child

Result: **Triple(subject**=*subject*, **predicate**=*verb*, **object**=*object***)**

Rule 3: Citation from gerund

Pattern: a speech act gerund (stating, saying, …) with a *complement* parent node that has a *subject* child node and a *complement clause* child node,

Result: **Citation(source=**subject* of *complement***, quote=***complement clause***)**

Rule 4: Action Nouns

Pattern: an action *noun* (attack, policy) with a possessive child and a preposition – object grandchild

Result: **Triple(subject**=*possessive*, **predicate**=*noun*, **object**=*object* of *preposition***)**

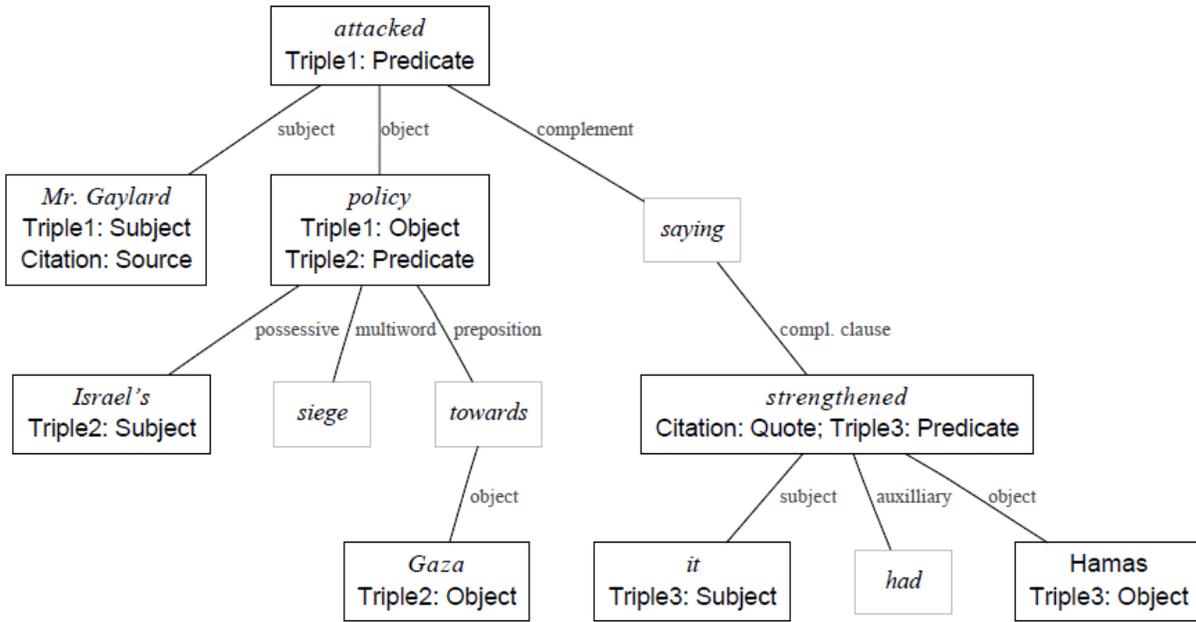**Figure 6: Selection of pattern rules for detecting semantic roles**

Figure 7: Semantic tree of " Mr Gaylard attacked Israel's siege policy towards Gaza, saying it had strengthened Hamas"

Figure 7 shows a more complicated example from an actual newspaper article: Mr Gaylard attacked Israel's siege policy towards Gaza, saying it had strengthened Hamas. This tree sets off rules: Rule 2 matches both active verbs *strengthened* and *attacked*. Rule 3 matches the gerund *saying*, finding as its source the subject of its parent node attacked and as its quote the clause below *strengthened.* Rule 4 matches the *policy,* using the prepositional object Gaza as object and the possessive *Israel's* as subject. This yields the annotations **Citation** and **Triple1** to **Triple3** displayed in the syntax tree. Similar to above, in order to create a semantic network from these annotations we would need a suitable ontology to link Mr. Gaylard to the UN and Gaza to Palestinians. Moreover, we need sentiment analysis to determine that *siege policy* and *attack* are negative while *strengthening* is positive. Finally, we need to use anaphora resolution to determine that the *it* in this sentence refers back to Israel's *policy* (Lappin & Laess 1994, Van Atteveldt et al 2008 polana). This yields three semantic roles: [UN – policy], [Israel – Palestinians], and [UN: policy + Hamas].

Note that this example showcases another complexity in extracting a semantic graph from language: graphs are by definition first order, meaning that relations cannot themselves be used as nodes in another relation. However, in natural language relations are frequently nested, as in our example of Mr Gaylard attacking Israel's policy against Gaza. To reduce this complex network to a normal graph we need to resolve these containments using transitivity rules based on cognitive consistency theories discussed above. In this case, we woud conclude that Mr Gaylard is against Israel (since he disagrees with their policy) and in favour of Gaza (since he disagrees with a policy detrimental to them).
These steps entail a substantive interpretation of the implications of statements, and as move from the manifest to the latent content of the text. Since we can formally describe the rules for these interpretations this is much nicer, however, than asking human coders to draw such inferences, since it is

sometimes difficult to keep political knowledge (and bias) away from the interpretation.

Summary

Starting from old ideas about politics and political language this chapter explored whether the occurrence of concepts, their co-occurrence, and the relationships between them can be extracted automatically so as to infer the political meanings underlying a text. From a theoretical perspective, these three objectives correspond with the automation of first-order agenda setting, second-order agenda setting and semantic network analysis. The latter is not only concerned with the extraction of issue positions of actors from texts, but simultaneously with the extraction of other political relationships, such as conflict or cooperation between actors, success or failure of actors, consequences of issues for actors, causal relationships between issues, and so on.

By showing what information the different methods (association analysis, manual coding, syntactic parsing) extract from a single example text on the Middle East conflict, the chapter illustrates how these different methods show the semantic network expressed in this text in different levels of detail. By using pattern matching on the automatically parsed syntax trees, it showed that automation of semantic network analysis can proceed beyond word counts and co-occurrence. It also illustrated the complex patterns originating from single sentences and the additional techniques required to move from extracted syntactic roles to a full semantic network.

Literature

Ansolabehere, S., & Iyengar, S. (1995). *Going negative: How attack ads shrink and polarize the electorate*. New York: Free Press.

Antoniou, G., & van Harmelen, F. (2004). *A semantic web primer*. Cambridge: MIT Press.

Budge, I., & Farlie, D. J. (1983). *Explaining and Predicting Elections: Issues Effects and Party Strategies in Twenty-Three Democracies*. London: George, Allen and Urwin.

Budge, I., Klingemann, H. D., Volkens, A., Bara, J., & Tanenbaum, E. (2001). *Mapping policy preferences: estimates for parties, electors, and governments, 1945-1998*. Oxford: Oxford University Press.

Budge, I., Robertson, D., & Hearl, D. (1987). *Ideology, Strategy and Party Change: A Spatial Analysis of Post-war Election Programmes in 19 Democracies*. Cambridge: Cambridge University Press.

Cappella, J. I., & Jamieson, K. H. (1997). *Spiral of Cynicism: The Press and The Public Good*. New York and Oxford: Oxford University Press.

De Tocqueville, A. C. H. C. (1840 (1951, 8e édition)). *De la démocratie en Amérique*. Paris: Gallimard.

Deetjen, G. (1977). *Industriellenprofile in Massenmedien: ein neuer Ansatz zur Aussagenanalyse*. Hamburg: Verlag Hans-Bredow-Institut.

Dixon, R. M. W. (1992). *A new approach to English grammar, based on semantic principles*. Oxford: Clarendon.

Dixon, R. M. W. (2005). *A semantic approach to English grammar*. Oxford: Clarendon.

Elster, J. (1993). *Political psychology*. Cambridge: Cambridge University Press.

Hayes, D. (2005). Candidate Qualities through a Partisan Lens: A Theory of Trait Ownership. *American Journal of Political Science, 49*(4), 908-923.

Heider, F. (1946). Attitudes and cognitive organization. *The Journal of Psychology, 21*(1), 107-112.

Hetherington, M. J. (1996). The Media's Role in Forming Voters' National Economic Evaluations in 1992. *American Journal of Political Science, 40*(2), 372-395.

Jones, B., & Baumgartner, F. (2005). *The politics of attention: How government prioritizes problems*. Chicago: University of Chicago Press.

Klein, D & Mannig, C. D. (2003). *Accurate Unlexicalized Parsing*. Proceedings of the 41st Meeting of the Association for Computational Linguistics, pp. 423-430

Kleinnijenhuis, J. (2008). Reasoning in economic discourse: a network approach to the Dutch Press. In K. Krippendorff & M. A. Bock (Eds.), *The Content Analysis Reader* (pp. 430-442). Thousand Oaks: Sage.

Kleinnijenhuis, J., & Pennings, P. (2001). Measurement of party positions on the basis of party programmes, media coverage and voter perceptions, in: . In M. Laver (Ed.), *Estimating the Policy Positions of Political Actors* (pp. 162-182). London and New York: Routledge, .

Kleinnijenhuis, J., Van Hoof, A. M. J., Oegema, D., & De Ridder, J. A. (2007). A test of rivaling hypotheses to explain news effects: news on issue positions of parties, real world developments, support and criticism, and success and failure. *Journal of Communication, 57*(2), 366-384.

Krippendorff, K. (2004). *Content Analysis*. Thousand Oaks: Sage.

Lappin, S., & Leass, H. (1994). *An algorithm for pronominal anaphora resolution*. Computational Linguistics, 20(4), 535-561.

Lipset, S., & Rokkan, S. (1967). *Cleavage structures, party systems, and voter alignments: An introduction*. New York: The Free Press.

McCombs, M. E. (2004). *Setting the agenda: the mass media and public opinion*. Cambridge: Polity Press.

Machiavelli, N. (1961 [1513,1532]). *The Prince [Il Principe].* Harmondsworth: Penguin Books.

Machiavelli, N. (1970 [1513]). *The discourses [Discorsi sopra la prima deca di Tito Livio].* Harmondsworth: Penguin Books.

Mena-Montes, N. (2010). *Un estudio sobre la inmigración (2000-2008): La construcción de la agenda (Agenda Building) y la evolución de los encuadres (Frame Building) en el discurso político-parlamentario, mediático y sus referencias a la opinión pública (with an extended English summary).* Madrid: Universidad Rey Juan Carlos (PhD dissertation).

Moilanen, K., & Pulman, S. (2007). *Sentiment composition*.

Noelle-Neumann, E. (1980). *Die Schweigespirale: Oeffentliche Meinung, unsere sociale Haut*. Frankfurt am Main: Ullstein.

Osgood, C. E., Saporta, S., & Nunally, J. C. (1956). Evaluative assertion analysis. *Litera, 3*(.), 47-102.

Patterson, T. E. (1993). *Out Of Order*. New York: Knopf.

Petrocik, J. R. (1996). Issue ownership in Presidential Elections, with a 1980 case study. *American Journal of Political Science, 40*, 825-850.

Riloff, E., & Wiebe, J. (2003). *Learning extraction patterns for subjective expressions*.

Rogers, E. M., Dearing, J. W., & Bregman, D. (1993). The Anatomy of Agenda-Setting Research. *Journal of Communication, 43*(2), 68-84.

Ruigrok, P. C., & van Atteveldt, W. H. (2007). Global Angling with a Local Angle: How U.S., British, and Dutch newspapers Frame Global and Local Terrorist Attacks. *The Harvard International Journal of Press and Politics, 12*, 68-90.

Sanders, D., & Gavin, N. (2004). Television News, Economic Perceptions and Political Preferences in Britain, 1997-2001. *The Journal of Politics, 66*(4), 1245-1266.

Schumpeter, J. (1944). *Capitalism, Socialism and Democracy*.

Severin, W., & Tankard, J. W. (2005). *Communication theories*. New York: Addison-Wesley / Longman.

Shah, D. V., Watts, M. D., Domke, D., & Fan, D. P. (2002). News Framing and Cueing of Issue Regimes: Explaining Clinton's Public Approval in Spite of Scandal *Public Opinion Quarterly, 66*(3), 339-370.

Soroka, S. (2006). Good News and Bad News: Asymmetric Responses to Economic Information *Journal of Politics, 68*(2), 372-385.

Tomz, M., & Van Houweling, R. P. (2008). Candidate Positioning and Voter Choice. *American Political Science Review, 102*(3), 303-318.

Tversky, A. (1977). Features of similarity. *Psychological Review, 48*(4), 327-352.

van Atteveldt, W. (2008). *Semantic Network Analysis: techniques for extracting, representing and querying media content*. Charleston, SC: BookSurge Publishers.

van Atteveldt, W., Kleinnijenhuis, J., & Ruigrok, N. (2008). Parsing, Semantic Networks, and Political Authority: Using Syntactic Analysis to Extract Semantic Relations from Dutch Newspaper Articles. *Political Analysis, 16*(4), 428-446.

van Atteveldt, W., Kleinnijenhuis, J., Ruigrok, N., & Schlobach, S. (2008). Good news or bad news? Conducting sentiment analysis on Dutch text to distinghuish between positive and negative relations. *Journal of Information Technology & Politics, 5*(1), 1-22.

van Cuilenburg, J. J., Kleinnijenhuis, J., & de Ridder, J. A. (1986). Towards a graph theory of journalistic texts. *European Journal of Communication, 1*, 65-96

van Noord, G. At Last Parsing Is Now Operational. In: Piet Mertens, Cedrick Fairon, Anne Dister, Patrick Watrin (editors): TALN06. Verbum Ex Machina. Actes de la 13e conference sur le traitement automatique des langues naturelles. Page 20--42..

Westholm, A. (1997). Distance versus Direction: The Illusory Defeat of the Proximity Theory of Electoral Choice. *The American Political Science Review, 91*(4), 865-883.

Wilson, T., Wiebe, J., & Hoffmann, P. (2005). *Recognizing contextual polarity in phrase-level sentiment analysis*.

Wittgenstein, L. (1922). *Tractatus Logico-Philosophicus: Logisch-philosophische Abhandlung*. London: Kegan Paul.

Wittgenstein, L. (1953). *Philosophical investigations*. Oxford: Blackwell.