CORPS: A Corpus of Tagged Political Speeches for Persuasive Communication Processing

Marco Guerini
Carlo Strapparava
Oliviero Stock

ABSTRACT. In political speech, even if the audience is sympathetic to the speaker and does not need to be persuaded, it tends to react or respond to signals of persuasive communication (including an expected theme, a name, an expression, and the tone of the voice). In this article, we describe the creation of a corpus of political speeches tagged with audience reactions, such as applause, as indicators of persuasive expressions. We hypothesize that corpora of this kind can be usefully employed in the qualitative analysis of political communication. In addition, we present a corpus-based approach for persuasive expression mining that relies on techniques from natural language processing (NLP). We show how the approach can support the analysis of political communication, providing insights well beyond those of traditional word-counting analysis techniques.

KEYWORDS. Persuasion, natural language processing, political communication, corpora collection

Natural language processing (NLP) is a subfield of artificial intelligence and computational linguistics that deals with automated generation and understanding of natural human languages. Persuasive NLP focuses, in particular, on the use of language for inducing desired beliefs and behaviors in the receiver(s). In order to automatically produce and analyze persuasive communication, specific resources and methodologies are needed. To this end, we built a resource called CORPS (CORpus of tagged *P*olitical *S*peeches) that contains political speeches tagged with audience reactions.

Marco Guerini is a Young Researcher at FBK-Irst in Trento. He graduated with a degree in Philosophy (with specialization in logic and cognitive science) and holds a Ph.D. in computer science (in information and communication technologies). His current research is in the field of computer–human interaction, and focuses on data-driven approaches to persuasive communication. He is a member of program committees of various workshops and a reviewer for conferences and journals.

Carlo Strapparava graduated with a degree in Computer Science at University of Pisa with a thesis on theoretical computers. He is a researcher at FBK-irst (former ITC-irst) since 1988 and actually is Senior Researcher in Human Language Technology Research Unit. His research activity covers artificial intelligence, natural language processing, affective computing, and computational humor. He regularly serves in the program committees of the major NLP conferences.

Oliviero Stock has been at ITC-irst since 1988 and was its Director from 1997 to 2001. His activity is mainly in artificial intelligence, natural language processing, intelligent interfaces, and cognitive technologies. He has been Chairman of the European Coordinating Committee for Artificial Intelligence, President of the Association for Computational Linguistics and of the Italian AI Association, and is an ECCAI Fellow and a AAAI Fellow.

Address correspondence to: Marco Guerini, FBK-irst, Povo I- 38050, Trento, Italy. (E-mail: marcoguerini@gmail.com).

Journal of Information Technology & Politics, Vol. 5(1) 2008
Available online at http://jitp.haworthpress.com

doi:10.1080/19331680802149616

In this article, we describe the construction of this resource and present some experiments for acquiring a lexicon (i.e., dictionary) of persuasive expression from it, using a specific measure of the persuasive impact of words. In particular, we focus on the analysis of expressions that provoke audience reactions, such as applause, representing an audience validation of a speaker's rhetoric. We further argue for the advantages of this measurement in the automatic analysis of political communication, showing that traditional approaches based on word usage (e.g., counting how many times the word *war* is used in a speech) fail to highlight important rhetorical phenomena.

Here we focus mainly on lexical (i.e., word-level) aspects of persuasive communication, and show how our approach can help to address a number of political analysis questions: For example, how do political speeches change after key historical events? What can be said about the lexical choices of well-known persuasive speakers? How does the perception of the enemy change in different historical moments? Still the corpus is potentially useful for many other NLP and political analysis tasks (that involve, for example, reasoning about syntactic and rhetorical aspects of the speeches) that do not necessarily have a strict focus on persuasion.

This article is structured as follows: We first give an overview of key concepts connected to persuasion, such as argumentation and rhetoric, and briefly describe the state of the art in related areas (NLP and political sciences). We then describe CORPS, the resource we built for statistical acquisition of persuasive expressions, and some issues related to the annotation of this specific resource. The following sections introduce some NLP techniques tailored to CORPS structure for persuasive expression mining. Finally we provide some examples that show the advantages of using CORPS for political communication analysis.

PERSUASION, AFFECT AND NLP

According to Perelman & Olbrechts-Tyteca (1969), persuasion is a skill that human beings use in communication, in order to make their partners perform certain actions or collaborate in various activities. Here we introduce some related key concepts.

Argumentation and Persuasion

In artificial intelligence, the main approaches focus on the argumentative aspects of persuasion. Still, argumentation is considered a process that involves rational elements, while persuasion includes elements like emotions. In our view, a better distinction can be drawn considering their differences of attention: While the former focuses on message correctness (its being a valid argument) the latter is concerned with its effectiveness (its impact). The recent area of natural argumentation (Reed & Grasso, 2007), tries to bridge the two by focusing, for example, on the problem of the adequacy of the message.

Emotions and Persuasion

Since persuasion includes nonrational elements, it is a superset of argumentation, but this does not rule out that there is a role for emotion within argumentation: through arousal of emotions or through appeal to expected emotions, as stated by Miceli, deRosis, and Poggi (2006). Indeed, emotional communication has become of increasing interest for persuasive NLP.

Rhetoric

Rhetoric is the study of how language can be used effectively. This area of study concerns the linguistic means of persuasion (one of the main means, but not the only one). This is the area we are focusing on in this article.

Irony

Irony refers to the practice of saying one thing while meaning another. Irony occurs when a word or phrase has a surface meaning, but another contradictory meaning beneath the surface. Irony is a widely used rhetorical artifice, especially in advertisement.

Persuasion and NLP

Past works on persuasion and NLP have focused mainly on natural language generation

(NLG). A notable exception is Araucaria by Reed & Rowe (2004). NLG is the branch of NLP that deals with the automatic production of texts in human languages, often starting from nonlinguistic input (Reiter & Dale, 2000). Usually this field is described as investigating communicative goals, the dynamic choice of what to say, the planning of the overall rhetorical structure of the text, and the actual realization of sentences on the basis of grammar and lexicon. Persuasive text generation deals with the production of texts that are meant to affect the behavior of the receiver. For example STOP, one of the best known NLG systems for the clinical smoking domain, uses domain-specific rules, based on expert knowledge acquisition (Reiter, Sripada, & Robertsonn, 2003). Promoter by Guerini, Stock and Zancanaro (2007), instead, uses strategies gathered from different persuasive theories and subsumed in a general planning framework. Other persuasive NLG systems are more argumentation-oriented, using classical argument structure theories such as Toulmin (1958), Perelman & Olbrechts-Tyteca (1969), and Walton (1996).

Since emotional reasoning is usually performed in order to modify/increase the impact of the message, effective NLP is strictly connected to persuasive NLP. An annotated bibliography on affective NLG can be found in Piwek (2002).

Opinion mining is a topic at the crossroads of information retrieval and computational linguistics concerned with the identification of opinions (either positive or negative) expressed in a document, for example, "The tax proposal was simple and well received." Recent research has tried to automatically identify whether a term has a positive or a negative connotation, as seen, for example, in Carenini, Ng, and Zwart (2005), Wilson, Wiebe, and Hwa (2004), Breck, Choi, and Cardic (2007). In Carenini et al. (2005), a method for feature extraction that draws on an existing unsupervised method is introduced. The work in Wilson et al. (2004) presents methodologies that use a wide range of features, including new syntactic features, for opinion recognition. Wilson, Wiebe, and Hoffmann (2005) propose an approach that tries to identify opinions at the phrase-level considering the word polarity in context rather than a priori.

Opinions, once extracted, must be summarized (in case of) and presented to the user. The advantages and limits of this extraction-based approach are discussed in Carenini, Ng, and Pauls (2006). Opinion mining deals with texts that are meant to persuade, but its focus is on polarity (valence) recognition for evaluative language retrieval, while persuasive expression mining deals with the extraction of pieces of text that are meant to persuade, regardless of their possible evaluative use.

Persuasion and Automatic Analysis of Political Communication

While there is a huge amount of theoretical and empirical research on politicians' rhetoric, only in recent years has there been a growing interest in bridging the gap between qualitative analysis of political communication and computational linguistics in order to automatize tasks that were usually carried out manually. A well-detailed discussion on the broader problem of integrating information technologies with social science research can be found in Cousins and Mcintosh (2005).

Furthermore, the automatic analysis of political communication is mainly focused on text categorization. Text categorization deals with the task of assigning a document to a predefined set of categories, such as determining party position in a text (e.g., Republican or Democratic); see, for example, the work by Purpura & Hillard (2006) and Purpura, Hillard, and Howard (2006). In Purpura & Hillard (2006), a topic-spotting classification algorithm was used for the task of coding legislative activities into subject areas; the algorithm used a traditional bag-of-words document representation. In Purpura et al. (2006), the authors presented a method based on support vector machines for classifying political e-mails according to the party affiliation of the person who sent them (either Republican or Democrat).

Franzosi, in many of his works, for example Franzosi (2004), focused more closely on persuasion issues of political communication, in particular on narrative and semantic aspects. He created a large scale corpus of annotated political news from opponents' journals, using the PC-ACE tool to manually annotate them. His aim was to understand the characteristics of social events during the fascist period, so the interest on persuasive aspects was quite incidental.

Finally, an automatic analysis of the lexical aspects of political communication, similar to the work presented in this article (but not considering words' persuasive impact), can be found in the work of Laver and colleagues: Benoit & Laver (2003), Laver & Benoit (2002), Laver, Benoit, and Garry (2003), Laver & Garry (2000), or in Bligh, Kohles, and Meindl (2004). We will further discuss these works in the section dedicated to the use of CORPS for political communication analysis.

CORPS

In this article, we adopt persuasive expression mining techniques as a component for persuasive NLP systems in an unrestricted domain. As for emotions, we restrict our focus to valenced expressions (i.e., those that have a positive or negative connotation). We collected a specific resource aimed at persuasion: a CORpus of tagged Political Speeches (CORPS), as examples of long and elaborated persuasive texts.

In collecting this corpus, we relied on the hypothesis that tags about public reaction, such as APPLAUSE, are indicators of hot-spots where persuasion attempts succeeded or, at least, a persuasive attempt had been recognized by the audience; on this point see Bull & Noordhuizen (2000) on mistimed applause in political speeches. We can then perform specific analyses—and extractions—of persuasive linguistic material that caused the audience reaction.

Given that the corpus is composed of transcriptions of speeches mostly given at public mass gatherings, in general the audience is favorable to the speakers and the context is one of support. Of course, by giving value to the audience reactions, we do not mean that the audience is actually effectively persuaded of some ideas or induced to do something that it did not believe in beforehand, even if the audience can be reassured, inspired, or helped in making sense of events. To the contrary, the audience tends just to react to signals, including an expected theme, a name, an expression, or the tone of the voice. Often the signals are creative, in the sense that the speaker may have produced new forms through creative rhetorical elaboration, but eventually they are recognized. Therefore the audience, so to say, resonates to a fragment of speech, which is meant to be of a persuasive genre and mostly concerned with a concept or a conceptual framework of which the audience is already persuaded. To be successful, the speaker's expression that immediately leads to the audience reaction must have been coherently composed. So we believe that there is a wealth of material that, by virtue of the validation provided by the audience reaction, can be used by a machine to automatically learn and use in different situations, where it may have the goal of effectively persuading someone, or simply to reproduce politicians' speech or be used for analyzing the pragmatic characteristics of novel political speech.

Given the textual nature of the corpus, rhetorical artifices based on prosody and other speech features cannot be addressed. These artifices are used to highlight key passages of a speech, with the help of high-impact words or concepts.

At present, there are approximately 900 speeches in the corpus and about 2.2 million words (see Figure 1 for a survey on the main speakers' number of speeches). The speeches are all in native English language, and all represent monological situations (i.e., there is only one speaker addressing an audience). We took this decision since dialogical situations, like in political debates, are not in our current focus of research and pose further problems in labeling and analysis.

These speeches have been collected from the Internet, and an automatic conversion of audience reactions tags has been performed to make them homogeneous in formalism and labeling. For example some discourses contain the tag {BIG-APPLAUSE} while others have {LOUD-APPLAUSE}. See Table 1 for a summary of audience reactions tags and their conversion.

FIGURE 1. Number of speeches per speaker.

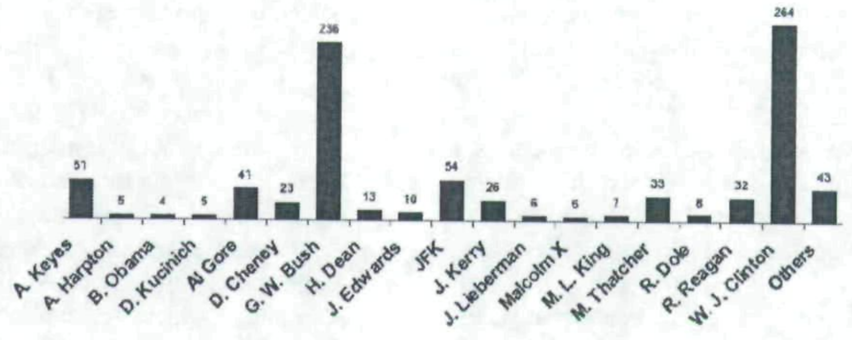


TABLE 1. List of Main Tags

Tag	Note
{APPLAUSE}	Main tag in speech transcription
{SPONTANEOUS-DEMONSTRATION}	Tags replaced: "reaction" "audience interruption"
{STANDING-OVATION}	–
{SUSTAINED-APPLAUSE}	Tags replaced: "big applause" "loud applause" etc.
{CHEERS}	Cries or shouts of approval from the audience. Tags replaced: "cries" "shouts" "whistles," etc.
{BOOING}	In this case, the act of showing displeasure by loudly yelling "Boo." Tags replaced: "hissing"
{TAG1 ; TAG2 ; ...}	In case of multiple tagging, tags are divided by semicolon. Usually there are at most two tags.
Special Tag	**Note**
{AUDIENCE-MEMBER} [text] {/AUDIENCE-MEMBER}	Tag used to signal a single audience member's intervention, such as claques speaking.
{OTHER-SPEAK} [text] {/OTHER-SPEAK}	Tag used to signal speakers other than the subject (like journalists, chairmen, etc.)
{AUDIENCE} [text] {/AUDIENCE}	Tag used to signal audience's intervention.

Metadata regarding the speech has also been added: title, event, speaker, date, and description. See Table 2 for a complete description of the structure of the speeches.

TABLE 2. Structure of a Speech Entry in CORPS

{title} [mandatory - describing the speech] {/title}
{event} [not mandatory - derivable from the title] {/event}
{speaker} [mandatory] {/speaker}
{date} [mandatory] {/date}
{source} [mandatory - internet address] {/source}
{description} [if present in the source] {/description}
{speech} [speech transcription with audience reactions tags] {/speech}

A special tag COMMENT is used for particular cases, for example:

- {COMMENT = "A moment of silence was observed"}
- {COMMENT = "An audience member claps"}
- {COMMENT = "Recording interrupted"}

With regard to the problem of interannotator and intersource agreement, it should be noted that:

(a) The automatic conversion of audience reactions tags drastically reduces the problem of the heterogeneity in tag vocabularies; in fact various sources were considered in order to collect this corpus. These discrepancies can be virtually eliminated at the analysis stage by further clustering tags into coherent groups of audience reactions (see following sections).

(b) Since tags represent audience reactions, in principle there is an evident high interannotator agreement. In some sense it is the audience itself that annotates the corpus.

As for the problem of label informativeness, especially if focusing on the problem of mistimed applause, it should be noted that there are no explicit annotations on applause duration, delay, or similar in this corpus (see for example Atkinson, 1984), so it is difficult to state if and when there has been a mismatching. Still, we believe that for our purposes this is not a problem, because persuasive dynamics are still present. An *interruptive applause* indicates that there has been an impact on the audience even if not intended by the speaker. A *delayed applause* indicates that there has been a persuasive attempt that has not been promptly recognized by the audience.

Moreover, given the four categories of mistiming proposed by Bull and Noordhuizen (2000), at least some cases can be individuated:

- An *isolated applause* is individuated by {COMMENT = "An audience member claps"} when explicitly recorded by annotators. Obviously this tag is not considered as the tag APPLAUSE.
- An *interruptive applause* is individuated by a fragment of speech where a sentence is broken up by an audience intervention (no End Of Sentence mark, dangling sentence parsing, and usually before the tag there is also a double dash to signal the interruption).
- Also, the special cases of speakers interrupting applause can be individuated when the speaker explicitly asks to continue.

Text annotation techniques vary according to the degree of manual intervention involved in the annotation process: Text annotation tasks can be accomplished entirely manually, entirely automatically, automatically after a manual training period, or semiautomatically. The semiautomatic approach we used to collect CORPS limits the amount of costly, manually annotated data. The procedure involved the following:

- The use of specific HTML parsing algorithms to extract the metadata from the Web pages (when large scale and homogeneous corpus were available).
- Conversion to make homogeneous the tag names (as mentioned before).
- A manual check for consistency of the final output, for example (a) the Web sources were not uniformly formatted and (b) annotators made typos in tagging.

EXPLOITING THE CORPUS

Among techniques at the root of traditional social science research, there is text analysis and various methodologies that can focus on the lexical, syntactic, or semantic level. In analyzing CORPS, the focus has been posed on the lexical level, both from a persuasive and affective point of view, and partially on the syntactic level. The NLG uses are briefly mentioned in the conclusions.

To reduce data sparseness, we used a lemmatizer and a part-of-speech tagger on the whole corpus, that gave for each token in the text the corresponding lemma and pos. So, at the lexical level we considered lemmata (e.g., the verb to *win*) rather than *tokens* (i.e., the form of the word, as it appears in the text: *win, wins, won*). In the following sections, if not differently stated, the term *word* indicates a lemma#pos—where pos can be *v* for verbs, *a* for adjectives, *r* for adverbs and *n* for nouns. So the word *to win* is represented as win#v. In the lexical analysis we further considered the following:

- Windows of different width *wn* (where *wn* is the number of tokens considered) preceding audience reactions tags
- The typology of persuasive communication (audience reaction)

As for what concern the last point in this list, we individuate three main groups of tags:

- *Positive-Focus*: This group indicates a persuasive attempt that sets a positive focus in the audience. Tags considered (about 16,000): {APPLAUSE}, {SPONTANEOUS-DEMONSTRATION}, {STANDING-OVATION}, {SUSTAINED-APPLAUSE}, {AUDIENCE-INTERVENTION}, {CHEERING}.
- *Negative-Focus*: It indicates a persuasive attempt that sets a negative focus in the audience. Note that the negative focus is set towards the object of the speech and not on the speaker themselves (e.g., "Do we want more taxes?") Tags considered (about 100): {BOOING}, {AUDIENCE} No! {/AUDIENCE}.
- *Ironical*: Indicate the use of ironical devices in persuasion. Tags considered (about 4,000): {LAUGHTER}.[1]

It should be noted that, rhetorically, positive-focus reactions can be obtained also by means of (sub) fragments of speech that set a temporary negative focus in the audience, or even by means of a complete focusing on negative aspects (usually political opponents' behavior). In fact, about 30% of the time, the rhetorical device used in political speeches to evoke applauses is CONTRAST [e.g. see Atkinson (1984); Heritage & Greatbatch (1986)].

Let us consider the speech that John F. Kennedy gave in Berlin on the 26th of June 1963, and in particular the following fragment that led to a {APPLAUSE ; CHEERS} reaction: "Freedom has many difficulties and democracy is not perfect. But we have never had to put a wall up to keep our people in—to prevent them from leaving us." This fragment sets a double negative focus. First, by means of a CONCESSION, Kennedy sets a negative focus on the limits of the American social model: "Freedom has many difficulties and democracy is not perfect," then by means of a CONTRAST he sets a stronger negative focus on the Soviet social model: "But we have never had to put a wall up to keep our people in—to prevent them from leaving us." Still, the overall effect of the fragment, based on an implicit CONCESSION and an explicit CONTRAST, is to set people to a positive point of view on the American social model.

CORPS AND PERSUASIVE EXPRESSION MINING

Though there have been various works focusing on the lexical level of political speeches, for example Laver et al. (2003) and subsequent works such as Martin & Vanberg (2007), those works were focused only on political positions recognition, a task similar to text categorization, and they treated all the words as potentially equivalent, leaving aside aspects such as emotional content or, more generally, persuasive impact.

For the analyses presented hereafter, we used the following resources and tools:

a) The TextPro package to perform lemmatization, POS analysis, named entity recognition, and sentence splitting; see Pianta & Zanoli (2007) and Zanoli & Pianta (2007).

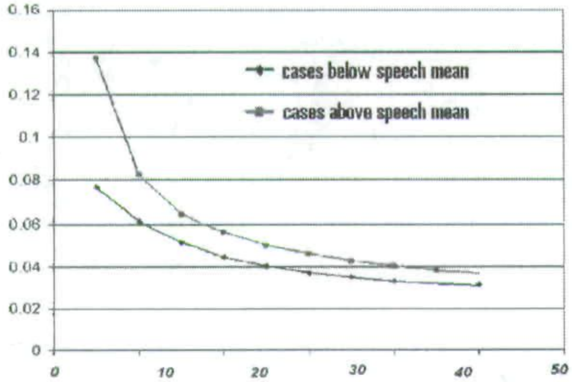
b) SentiWordNet[2] scores. Esuli & Sebastiani (2006), to compute the valence of speeches lexical entries (words). An example of SentiWordNet items is given in Table 3.

We conducted a preliminary analysis of the corpus focusing on the relation between valence and persuasion: The phase that leads to audience reaction (e.g., APPLAUSE), if it presents valence dynamics, is characterized by a

TABLE 3. Examples of SentiWordNet Entries

POS	Offset	PosScore	NegScore	SynsetTerms
a	602378	0.0	0.875	wrong#a#1 incorrect#a#1
r	60640	0.75	0.0	better#r#1
n	7017251	0.0	0.0	victory#n#1 triumph#n#1

FIGURE 2. Relation between valence and persuasion.



valence crescendo. That is to say, persuasion is not necessarily achieved via modification of valence intensity, but, when this is the case, it is by means of an increase in the valence of the fragment of speech.

To come to this result, we computed, for every window, its mean valence ($\overline{w}$), calculated by summing up all the valences of the lemmata (SentiWordNet scores) corresponding to the tokens in the fragment and dived by wn, and subtracted the mean valence of the corresponding speech ($\overline{s}$). In this way we obtained two classes of windows:

- Windows with mean-valence above the mean-valence of the speech ($\overline{w} > \overline{s}$)
- Windows with mean-valence below the mean-valence of the speech ($\overline{s} > \overline{w}$)

We then summed up all the values for the two classes and normalized the results by dividing it for the total number of cases in the class (n_c). We repeated the procedure for various window widths ($5 < wn < 40$), see Figure 2 and Equation 1. The results show that cases above the speech mean are fewer but far stronger. We are planning to have a finer grained analysis by means of cluster-based approaches and variable window width.

$$y = \frac{\sum abs \mid \overline{w} - \overline{s}}{n_c} \quad x = wn \qquad (1)$$

We then focused on the impact of the lexicon used in the speeches assuming that, for persuasive purposes (both in analysis and generation), not all the words have the same importance. We extracted persuasive words by using a coefficient of persuasive impact (pi) based on a weighted tf-idf (term frequency-inverse document frequency) (see Equation 2, $pi = tf \times idf$).

$$tf_{\mathrm{i}} = \frac{n_i \times \sum_{ni} s_i}{\sum_k n_k} \quad idf_{\mathrm{i}} = \log \frac{|D|}{|\{d : d \ni t_i\}|} \quad (2)$$

The tf-idf weight is a statistical measure used to evaluate how important a word is to a document in a corpus. To calculate the tf-idf weight, we created a virtual document by unifying all the tokens inside all the windows (of dimension $wn = 15$) preceding audience reactions tags, and considering the number of documents in the corpus as coincident to the number of speeches plus one (the virtual document). Obviously, from the speeches we subtracted those pieces of text that were used to form the virtual document. Given this premise we can now define the terms in Equation 2:

- n_i = number of times the term (word) t_i appears in the virtual document
- $\Sigma n_i\, s_i$ = sum of the scores of the word (the closer to the tag, the higher the score)
- $\Sigma_k n_k$ = the number of occurrences of all words = $wn \times |tags\ number|$
- $|D|$ = total number of speeches in the corpus (included the virtual document)
- $|\{d : d \ni t_i\}|$ = number of documents where the term t_i appears (we made a hypothesis of equidistribution)

Four lists of words were created according to the group of audience reactions tags they refer to: positive-focus words, negative-focus words, ironical words, and a persuasive words list—computed by considering all tags together. Analyzing the 100 top words of these lists, ordered according to their pi score, we found that the negative valence mean of positive-focus and negative-focus groups is the same, while for the negative-focus group the positive valence mean is about ¼ with regard to the positive-focus

TABLE 4. List of Most Persuasive Words

Positive-focus words	Negative-focus words
bless#v deserve#v victory#n justice#n fine#a relief#n November#n win#v help#n thanks#n glad#a stop#v better#r congressman#n lady#n regime#n fabulous#a uniform#n military#a wrong#a soul#n lawsuit#n welcome#v appreciate#v Bush#n behind#r grateful#a 21st#a defend#v responsible#a safe#a terror#n cause#n bridge#n prevail#v choose#v hand#n love#v frivolous#a sir#n honor#n defeat#v end#v fight#n no#r Joe#n ready#a wear#v future#a direction#n foreign#a death#n single#a democratic#a	horrible#a criticize#v waste#n opponent#n timidity#n shuttle#n erode#v torpor#n Soviets#n invasion#n scout#n violation#n Castro#n troop#n authority#n Guevara#n Kaufman#n Sachs#n Goldman#n ferociously#r solvent#n page#n front#a international#a direction#n monstrosity#n Cambodia#n unbearable#a drilling#n Soviet#a increase#v intelligence-gathering#a Carolina#n Gerald#n trusted#a drift#n operation#n WTO#n entry#n mcgovern#v coward#n household#n Neill#n

group (t-test; $\alpha < .01$). These results could be explained by a high use of the CONTRAST relation (that brings negatively valenced words when talking about opponents) in the positive-focus group, while this is not the case for the negative-focus group.

In Table 4, a comparison between the positive-focus and negative-focus top-50 most persuasive words is given (note that named entities have not been discarded). It is arguable whether these words are "universally" persuasive (i.e., they could be biased by speaker style, audience typology, context of use, and so on). To partially overcome the problem, the corpus was balanced by choosing speakers that are equally distributed within the two major parties in the US (Democratic and Republican). At present we do not address negation, hypothetical clauses, and similar, but we believe that they do not invalidate the *pi* of a word. Let us hypothesize that the word bad#a has a high *pi* in the positive-focus list and the word is mainly used in contexts like *not bad*. The word bad#a should not be discarded from the positive-focus list, rather it would be useful to have its co-occurrence score with the word no#r.

CORPS AND ANALYSIS OF POLITICAL COMMUNICATION

Analysis of public reaction can substantiate intuitions about the speaker's rhetorical style. Given the formal annotation of the corpus together with the *pi* measure we presented, this analysis can be made automatically on a large scale, allowing us to gain interesting insights. In fact there are rhetorical phenomena that do not come into light with traditional approaches—based on words' *usage* (counting of their occurrences). Considering also the words' *impact* (their persuasiveness coefficient *pi*), a much finer analysis is possible, for example:

How Do Political Speeches Change After Key Historical Events?

There are works such as Bligh et al. (2004) that investigated the lexicon of Bush's speeches before and after September 11, 2001 (9/11), with tools for automatic analysis of political discourses (DICTION 5.0) focusing on charisma traits. Using CORPS, and analyzing some of the speeches of George W. Bush before and after 9/11 (70 speeches before and 70 after, from 12 months before to 16 months after) at the lexical level, we found that while the positive valence mean remains totally unvaried, the negative increased by 15 percent (t-test; $\alpha < .001$).

Then we ran a quantitative/qualitative analysis on Bush's persuasive words before and after 9/11 to understand how his rhetoric changed, making two lists of persuasive words, one for the speeches before 9/11 and another for the speeches after 9/11. We focused on some paradigmatic words and found some interesting results. The words are presented in Table 5. In the first column there is the lemma#pos (word), in the second and third columns its position (persuasiveness)[3] in the lists before and after 9/11, and in the fourth and fifth columns the number of occurrences in the speeches. An *x* indicates that the word is not persuasive (i.e., when it appears in the corpus but never in proximity of an audience reaction, the persuasiveness ceases around position 2,500 in the lists). A hyphen indicates the word is not present in the corpus at all.

Bligh et al. (2004) followed a simple approach based on words' usage. Here, instead, we adopted also the words' impact and created a matrix, for every word, that records an increase or decrease of use compared with an increase or decrease of persuasiveness. Some interesting phenomena emerged. Let us consider the words military#n or treat#v. Both words are used almost the same number of times before and after 9/11 (respectively 23 vs. 29 times and 25 vs. 20 times). So their informativeness, based on number of occurrences, is null. But considering the persuasiveness score, we see that their impact varies quite a bit (respectively from position 197 to 36 and from 54 to 473). Let us also consider the word tax#n; if we consider only the number of occurrences, we could infer that before 9/11 this topic was much more "felt" (702 occurrences vs. 81), but if we look at persuasiveness we see that before 9/11 the word tax#n never got audiences' reactions, while after it become very popular (position 93). The same, but in an opposite direction, holds for war#n: mentioned three times more after 9/11 (80 vs. 254), but never got applause.

The results were divided in four blocks, according to thematic areas. In the first block there are words that became very popular after 9/11. They usually (indirectly) refer to war, usually from a positive point of view. These words were not considered before 9/11 (i.e., justice#n was not persuasive at all before 9/11 but jumped to the ninth position after; at

TABLE 5. Bush's Words Before and After September 11th

Lemma	Ranking before	Ranking after	Occur before	Occur after
win#v	112	7	27	52
justice#n	x	9	15	111
prevail#v	x	15	2	20
defeat#v	x	16	1	44
right#r	x	25	94	55
taliban#n	x	27	1	44
mighty#a	615	30	4	26
military#n	197	36	23	29
victory#n	826	65	9	26
evil#a	–	129	0	44
death#n	4	450	65	32
war#n	36	x	80	258
soldier#n	70	296	20	47
tax#n	x	93	702	81
refund#n	15	–	10	0
wage#n	121	–	4	0
drug-free#a	87	x	9	3
commander-in-chief#n	76	850	25	14
leadership#n	81	261	40	75
future#n	83	394	54	51
dream#n	99	321	77	30
soul#n	23	126	47	32
generation#n	122	442	27	56

Notes. In the second and third column, the number represents the rank in the list of persuasive words; an "x" indicates a *pi* = 0; an "–" indicates the word is not present in the corpus at all. In the fourth and fifth columns the total number of occurrences.

the same time its frequency increased ten times after the attack). The second block represents words that were popular before the attack but became unutterable after 9/11 (e.g., death#n fell from position 4 to 450, with the frequency cut in half). These words generally refer to the negative aspects of war or to war itself. The third block contains some words that represent the shift in the political agenda before and after 9/11: taxation, contrasting drugs use, leadership. The fourth block shows some abstract and moving words that became less used and popular after 9/11, partially in contrast to the findings of Bligh et al. (2004).

What Can Be Said of the Lexical Choices of a Specific Speaker Who Obtains a Certain Characteristic Pattern of Public Reaction?

Known as "the great communicator," Ronald Reagan's rhetoric has been the focus of many qualitative research studies, such as Collier (2006); such research has also focused on particular aspects of his style, such as his use of irony, see Weintraub (1986) and Stevenson (2004). We tried to test whether these findings were consistent with our corpus. By considering 32 of Ronald Reagan's speeches, we first found that the mean tag density of this collection is one-half of the mean tag density of the whole corpus (t-test; $\alpha < .001$). At first sight this result is somewhat strange, because his being a "great communicator" is not bound to his "firing up" rate (far below the average rate of others speakers). But interestingly, focusing only on the subgroup of ironical tags, we found that the density in Reagan's speeches is almost double as compared to the whole corpus (t-test; $\alpha < 0.001$). The results are even more striking if they are compared to the mean ironical-tags ratio mtr_i (the mean of the ratio of ironical tags to positive-focus and negative-focus tags per speech; see Equation 3) of the two groups. In Reagan's speeches the mtr_i is about 7.5 times greater than the mtr_i of the whole corpus (about 3.5 vs. about 0.5; t-test; $\alpha < .001$). That is to say, while normally there is one tag of LAUGHTER for every two other tags, such as APPLAUSE, in Reagan's speeches there is one tag, such as APPLAUSE, for every three or four tags of LAUGHTER.

$$mtr_i = \sum \frac{|\,ironical - tags\,|}{|\,positive - focus\,| + |\,negative - focus\,|} \tag{3}$$

With regard to Reagan's overall style, his criterion was, "Would you talk that way to your barber?" as reported in Collier (2006). He wanted his style to appear "simple and conversational." To verify this statement, we made a hypothesis that a simple and conversational style is more polysemic than a "cultured" style (richer in technical, and less polysemic, terms). We first calculated the mean polysemy of Reagan's speeches and compared it to the mean polysemy of the whole corpus, finding no statistical difference between the two; also, in this case, words' usage analysis was not informative. Then we focused on the persuasive lexicon: We made a list of Reagan's persuasive words and compared it to the persuasive words list of the rest of the corpus (we considered all the words whose *pi* was not 0). We found that the mean polysemy of Reagan's persuasive words is almost double as compared to the whole corpus (t-test; $\alpha < .001$).

How Does the Perception of the Enemy Change in Different Historical Moments?

A specific analysis on the valence of the lexical context surrounding named entities that elicit negative-focus audience reactions in different periods of time can provide interesting insights. Looking at Table 4, it is clear that there are various named entities in the list of negative-focus words at the top-most positions, while this is not the case for positive-focus words. Given the small amount of negative-focus tags, our approach will include a second, inductive analysis step: After individuating named entities that elicit negative-focus reactions (i.e., the "enemies"), those same entities will be searched in the corpus (in the surroundings of positive-focus tags) by assuming that they are inserted in a CONTRAST relation, that sets a temporary negative focus on enemies' behavior, as described before in this section.

Persuasive Opinion Mining

Not all the opinions expressed in speeches or texts have the same persuasive impact. "Successful" opinions, for example G. W. Bush speaking about W. J. Clinton, can be extracted considering those followed by a reaction of the audience. The role of rhetorical constructs will be taken into account in future research.

CONCLUSIONS AND FUTURE WORK

We have presented the CORPS corpus, which contains political speeches tagged with audience reactions. CORPS is freely available for research purposes (for further details see http://hlt.fbk.eu/corps), and we want to promote its scaling up. Along with the corpus, we have described techniques for statistical acquisition of persuasive expressions (such as a measure of persuasive impact of words) with a view to contributing to various persuasive NLP tasks. Effective expressions are of paramount importance in this context.

In the present work we have limited ourselves to lexical analysis, and of course if the corpus is not big enough this may lead to errors. In the long run, we will add more complex elements, such as syntax and negation and, most important, rhetorical analysis of the text and possibly speech (e.g., pitch) analysis. These more complex techniques will further help in the correct persuasive lexeme identifications, not to mention in the task of analyzing political speech more deeply and modeling of persuasive expression understanding and production.

Currently we are investigating natural language generation uses. For lexical choice, in text generation microplanning, the adoption of techniques that use corpus and domain information for choosing appropriate lemmata inside synsets has been proposed, among others, by Jing (1998). In our approach, the choice is performed considering lemma impact rather than lemma use (the lemma with the highest *pi* is extracted). If the typology of the persuasive communicative goal also is defined (positive-focus, negative-focus, ironical), the choice can be further refined by selecting the lemma according to the specific *pi* (i.e., accessing the proper list of persuasive words). These strategies are currently implemented in the Valentino prototype, Guerini, Stock, and Strapparave (2008), and will be added to the realization component of the existing Promoter prototype (Guerini et al., 2007).

With a methodology similar to the one used for persuasive words, we also extracted chunks of persuasive sentences. In this case, the window width was based on the number of sentences. We plan to use these chunks for extracting high-impact sequences of words and rhetorical patterns. In particular we want to (a) understand if the results presented by Atkinson (1984) and by Heritage & Greatbatch (1986) about rhetorical devices used to provoke audience reactions can be replicated/verified on a larger scale corpus, and (b) refine the analysis by focusing on complex patterns of rhetorical relations.

A paradigmatic example of application scenarios that can be envisaged is a summarization system that relies on audience reaction tags for extracting automatically key material from political speeches.

NOTES

1. If LAUGHTER appears in a multiple tag (e.g., together with APPLAUSE), by default this tag is associated to the ironical group. This is not the case for BOOING that occurs always alone.

2. WordNet is a large lexical database of English. Nouns, verbs, adjectives, and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept and indexed by an offset. Synsets are interlinked by means of conceptual-semantic and lexical relations. SentiWordNet is a lexical resource in which each WordNet synset is associated to three numerical scores: Obj(s), Pos(s), and Neg(s). These scores represent the objective, positive, and negative valence of the synset. Each entry takes the form lemma#pos#sense-number.

3. We use the rank in the list, instead of the *pi, Structures of social action* for readability purposes.

REFERENCES

Atkinson, J. (1984). Public speaking and audience responses: Some techniques for inviting applause. In J. Atkinson & J. Heritage (Eds.), *Structures of social action* (pp. 370–409). Cambridge, England: Cambridge University Press.

Benoit, K., & Laver, M. (2003). Estimating Irish party positions using computer wordscoring: The 2002 elections. *Irish Political Studies, 18*(1), 97–107.

Bligh, M. C., Kohles, J. C., & Meindl, J. R. (2004). Charisma under crisis: Presidential leadership, rhetoric, and media responses before and after the September 11th terrorist attacks. *The Leadership Quarterly, 15*(2), 211–239.

Breck, E., Choi, Y., & Cardie, C. (2007, January). Identifying expressions of opinion in context. In *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI-2007)*, Hyderabad, India.

Bull, P., & Noordhuizen, M. (2000). The mistiming of applause in political speeches. *Journal of Language and Social Psychology, 19*, 275–294.

Carenini, G., Ng, R., & Pauls, A. (2006). Multi-document summarization of evaluative text. In *Proceedings of EACL*. Trento, Italy.

Carenini, G., Ng, R., & Zwart, E. (2005). Extracting knowledge from evaluative text. In *Proceedings of the 3rd International Conference on Knowledge Capture* (pp. 11–18).

Collier, K. (2006). Writing for the great communicators: Speechwriting for Roosevelt and Reagan. In *Prooceedings of the Southwest Political Science Association Meetings*. San Antonio, Texas.

Cousins, K., & Mcintosh, W. (2005). More than typewriters, more than adding machines: Integrating information technology into political research. *Quality and Quantity, 39*, 581–614.

Esuli, A., & Sebastiani, F. (2006). SentiWordNet: A publicly available lexical resource for opinion mining. In *Proceedings of the 5th Conference on Language Resources and Evaluation* (pp. 417–422). Genova, Italy.

Franzosi, R. (2004). *From words to numbers: Narrative, data, and social science*. Cambridge: Cambridge University Press.

Guerini, M., Stock, O., & Strapparava, C. (2008). Valentino: A tool for valence shifting of natural language texts. In *Proceedings of LREC 2008*. Marrakech, Morocco.

Guerini, M., Stock, O., & Zancanaro, M. (2007). A taxonomy of strategies for multimodal persuasive message generation. *Applied Artificial Intelligence Journal, 21*(2), 99–136.

Heritage, J., & Greatbatch, D. (1986). Generating applause: A study of rhetoric and response at party political conferences. *American Journal of Sociology, 92*, 110–157.

Jing, H. (1998). Usage of WordNet in natural language generation. In S. Harabagiu (Ed.), *Proceedings of the Conference on the Use of WordNet in Natural Language Processing Systems*" (pp. 128–134). Somerset, NJ: Association for Computational Linguistics.

Laver, M., & Benoit, K. (2002). Locating ads in policy spaces: Wordscoring Dail speeches. *Irish Political Studies, 17*(1).

Laver, M., Benoit, K., & Garry, J. (2003). Extracting policy positions from political texts using words as data. *American Political Science Review, 97*(2), 311–331.

Laver, M., & Garry, J. (2000). Estimating policy positions from political texts. *American Journal of Political Science. 44*(3), 619–634.

Martin, L. W., & Vanberg, G. (2007). A robust transformation procedure for interpreting political text. *Political Analysis, 16*(1), 93–100.

Miceli, M., deRosis, F., & Poggi, I. (2006). Emotional and non-emotional persuasion. *Applied Artificial Intelligence, 20*, 849–879.

Perelman, C., & Olbrechts-Tyteca, L. (1969). *The new rhetoric: A treatise on argumentation*. Notre Dame: University of Notre Dame Press.

Pianta, E., & Zanoli, R. (2007, June). Tagpro: A system for Italian pos tagging based on svm. *Intelligenza Artificiale, Numero Speciale Strumenti di Elaborazione del Linguaggio Naturale per l'Italiano, 4*(2), 8–9.

Piwek, P. (2002). *An annotated bibliography of affective natural language generation* ([Technical Report] No. ITRI-02-02). University of Brighton.

Purpura, S., & Hillard, D. (2006). Automated classification of congressional legislation. In *Proceedings of the Seventh International Conference on Digital Government Research*. San Diego, CA.

Purpura, S., Hillard, D., & Howard, P. (2006). *A comparative study of human coding and context analysis against support vector machines (svm) to differentiate campaign emails by party and issues.*

Reed, C., & Grasso, F. (2007). Recent advances in computational models of argument. *International Journal of Intelligent Systems, 22*(1), 1–15.

Reed, C., & Rowe, G. (2004). Araucaria: Software for argument analysis, diagramming and representation. *International Journal of AI Tools, 14*(3–4), 961–980.

Reiter, E., & Dale, R. (2000). *Building natural language generation systems*. Cambridge, England: Cambridge University Press.

Reiter, E., Sripada, S., & Robertson, R. (2003). Acquiring correct knowledge for natural language generation. *Journal of Artificial Intelligence Research, 18*, 491–516.

Stevenson, D. J. (2004). *Laughter and leadership*. Paper presented at The International Center for Studies in Creativity, Buffalo State College.

Toulmin, S. (1958). *The use of arguments*. Cambridge, England: Cambridge University Press.

Walton, D. N. (1996). *Argumentation schemes for presumptive reasoning*. Mahwah, NJ: Lawrence Erlbaum Associates.

Weintraub, W. (1986). Personality profiles of American presidents as revealed in their public statements: The presidential news conferences of Jimmy Carter and Ronald Reagan. *Political Psychology*, *7*(2), 285-295.

Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*. Vancouver, Canada.

Wilson, T., Wiebe, J., & Hwa, R. (2004). Just how mad are you? Finding strong and weak opinion clauses. In *Proceedings of AAAI* (pp. 761-769).

Zanoli, R., & Pianta, E. (2007, June). Entitypro: Exploiting svm for Italian named entity recognition. *Intelligenza Artificiale, Numero Speciale Strumenti di Elaborazione del Linguaggio Naturale per l'Italiano*, *4*(2), 69-70.