

# Assessing the Impact of Language Style on Emergent Leadership Perception from Ubiquitous Audio

Dairazalia Sanchez-Cortes<sup>1,2</sup>, Petr Motlicek<sup>1</sup> and Daniel Gatica-Perez<sup>1,2</sup>

<sup>1</sup> Idiap Research Institute, Martigny, Switzerland

<sup>2</sup> Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland  
(dscortes,motlicek,gatica)@idiap.ch

## ABSTRACT

Leaders stand out for what they say and how they say it. This work describes the impact of the language style of emergent leaders in small group discussions based on 7 hours of audio from English spoken discussions recorded with a ubiquitous platform. For the language style analysis, word categories are extracted from manual transcriptions of the discussions as well as from automatically detected keywords. The most relevant word categories are then used to predict the emergent leader in each group. Our findings reveal that non-privacy sensitive word categories like amount of words, conjunctions and assent are good predictors of emergent leadership. The emergent leader can be correctly inferred in a fully automatic approach with up to 82% accuracy using categories derived from keywords, and up to 86% using categories derived from full manual transcriptions.

## Categories and Subject Descriptors

H.3.1 [Information Storage and Retrieval]: Content Analysis and Indexing

## General Terms

Human Factors

## Keywords

Emergent leadership, Language style, Dominance, Spoken Keywords

## 1. INTRODUCTION

In organizations today, group leaders are the interface to the problems that the organization faces, give them priorities, and propose solutions to be implemented by the group members. With globalization, the fast development of technology and the personal shifting among different geographical locations, there is an increasing need for effective self-organization and cooperation in newly formed groups. To cover this need in organizations, the emergent leader concept is becoming relevant. Emergence of leadership is an spontaneous phenomenon that occurs everywhere, whenever there

is a group of unacquainted people in which decisions need to be made. An emergent leader is defined as a person that integrates the group towards a common goal having her/his leading force initiative based on the sympathy of the group [27].

Emergent leadership is a universal phenomenon. A leader plays a critical role in social movements, as well as governmental, educational, medical, industrial and military settings. The role of the leader has implications on cooperation, cohesion, communication and coordination towards accomplishment of goals. Providing feedback to the leader, on how she/he engaged and persuade her/his work-team, or if she or he created a cohesive group, could enhance future communication patterns that could be reflected in the group's behavior. Such feedback information is starting to be embedded in awareness tools [16, 3].

The study of organizational phenomena like leadership is becoming relevant in ubiquitous and social computing as recording and analyzing sensor data is becoming more common in daily-natural scenarios, moving from lab rooms to in-field gathering of data [20, 6]. Today, interacting groups can be recorded with high-quality portable commercial microphone arrays designed for small groups [1, 9]. The high quality audio allows extraction of reliable prosodic features as well as automatic speech recognition.

Leaders stand out in such a way that the group believes he or she, rather than someone else, can make the most of individual differences towards a common objective [27]. The first attempt to identify emergent leaders through their verbal and nonverbal communication was presented in [27], where observers were able to identify the emergent leader by listening to filtered audio or looking at the full transcription of the interaction. Recent work in computing has examined the problem of detecting emergent leaders in small groups, using audio-video recordings in which groups from unacquainted people are recorded while solving a task. On one hand, in [25] emergent leaders were automatically inferred by considering only nonverbal information extracted from audio and video, with an accuracy of up to 85%. On the other hand, emergent leaders can be identified in asynchronous and remote small group scenarios by considering the interactions in the group as a social network or a virtual team, in which the main source of information are posted messages in a discussion board or web-based collaborative tools [28, 4].

Many aspects of our identity and relationships are embedded in the words we say or write [22, 7, 14, 10]. Existing findings in psychology reveal a strong connection between

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MUM'12, December 04 - 06 2012, Ulm, Germany

Copyright 2012 ACM 978-1-4503-1815-0/12/12 ...\$15.00.

personality traits and the language embedded in written or spoken forms [17, 22]. Language cues also provide information in the prediction of successful relationships [12]. Language has also been used to analyze the presidentiality of candidates by using manual transcriptions from publicly available interviews, speeches and debates [26].

The use of high quality audio recordings in scenarios in which privacy is not an issue, allows verbal content analysis. We present a solution to the problem of identifying emergent leaders from automatically transcribed spoken words in face-to-face group interactions, using high quality audio from ubiquitous devices. We are not aware of any work that has attempted to use automatic transcription of spoken words in interactions for predicting emergent leaders. We study two novel research questions in the context of predicting emergent leadership in small groups using ubiquitous sensing and automatic analysis. First, is there a correlation between how the emergent leader is perceived and his verbal language style (as opposed to nonverbal cues)? And secondly, can emergent leadership be inferred from only partial verbal samples of the full conversation? The language style is extracted using a psychological validated state-of-the-art content analysis module (LIWC), and investigated both under ideal condition (clean manual speech transcriptions) and realistic automated condition where a highly accurate keyword spotter is used in the audio channel. Our findings first reveal a significant correlation between language styles and the perceived emergent leader in the group. Second, a simple word counting approach can also provide an accurate inference of perceived dominance, a variable related (although not identical) to leadership. Third, by using fully automatic extraction of verbal content we can correctly identify the emergent leaders with an accuracy of up to 82%.

In Section 2, we describe our approach, followed by the description of the corpus used in Section 3. In Section 4, we explain the verbal feature extraction, as well as the leadership inference methods. In Section 5, we present the analysis of our results. Our final discussion and conclusions are presented in Section 6.

## 2. OVERVIEW OF OUR APPROACH

The analysis for the verbal content is based on a subset of the ELEA corpus [25], containing approximately 7 hours of audio. The corpus contains English spoken meetings, which are formed with unacquainted people, and recorded in several places using a commercial portable microphone [1]. The corpus contains groups of three and four participants that are asked to solve the winter survival task as a group while being recorded. It also contains questionnaire outputs on how they perceived the other participants during the interaction.

From the audio recordings, we generate a manual word transcription from the whole interaction. The scripted conversations are performed by a professional company that required only the audio files. In addition, we run an automatic keyword spotting detection system. Potential (most probable) keywords detected by the system and assigned with their confidence are automatically obtained.

From both the full manual transcriptions and detected keywords, we proceed to extract word categories using the Linguistic Inquiry and Word Count software (LIWC) [2]. LIWC is a text content software that allows language analysis (from text, transcribed speech, blogs etc.) based on high

level categories defined by psychologists.

For the emergent leader analysis, we generate individual files per participant accumulating their respective dialog segments and keywords, and then extract the language categories per participant using LIWC. As we work with transcriptions and ASR, we do not compute the categories that involve punctuations like periods, dots, exclamation, etc.

With the resulting 64 word categories, we perform a feature selection applying support vector machine-recursive feature elimination (SVM-RFE). Finally, using the top relevant features from the categories, we use a supervised method to automatically infer the emergent leader using the top relevant features. Figure 1 shows our approach. In our work, we address two goals:

1. Identify the top relevant word categories that have a connection with emergent leadership and dominance.
2. Inference of the emergent leader and the perceived dominant person in a small group in a fully automated way using a portable microphone array.

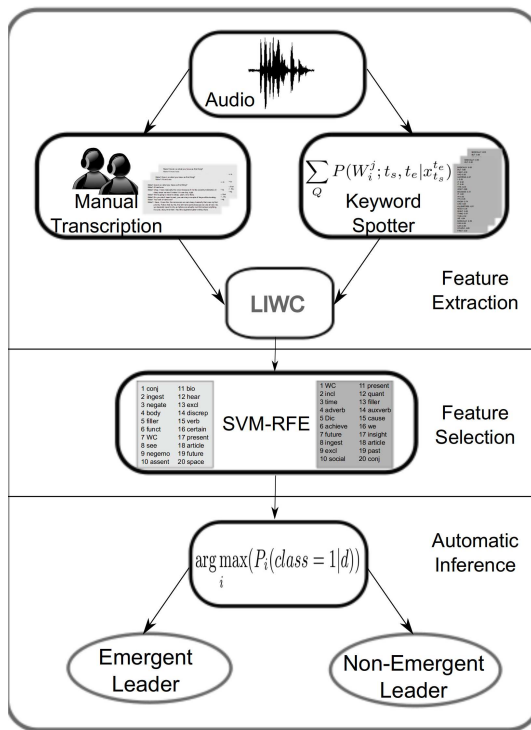


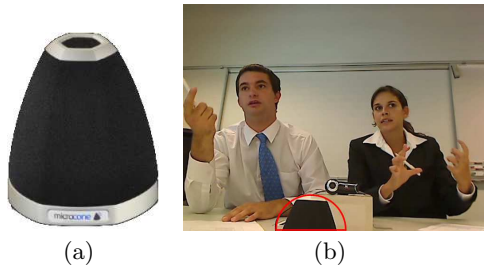
Figure 1: Our approach.

## 3. SENSORS AND DATA

To record the conversation in the group we use the Microcone, a commercial microphone array designed to record small group conversations of up to six participants [1]. The microcone is fully portable and works directly with a laptop and thus constitutes a pervasive platform in the workplace. Figure 2(a) shows the microcone device and an example of a recorded group from the corpus (2(b)). The microcone records the conversation at 16kHz and provides an automatically estimated speaking segmentation by the end of the recording. The segmentation per speaker is performed

using a filter-sum beamformer followed by a post-filtering stage [1].

The corpus consists of approximately 7 hours of recorded meetings. There are 20 four-person meetings and 9 three-person meetings composed of previously unacquainted people. The participants were asked to participate in the Winter Survival Task with no roles predefined. We chose the Winter Survival Test given that it is the most cited test in studies related to small group performance and leadership. It is focused on ranking a list of items in order to survive an airplane crash in winter [15]. The ranking is performed first individually and then as a team, to favor interaction between the participants and allow the emergence of the leader. Video was also available for ground-truth generation.



**Figure 2: Setup of the corpus: (a) The microcone audio recording device and (b) Recorded group from the corpus, the microcone is placed in the center of the table (highlighted in red).**

The participants in our data were given questionnaires that measure the perceived interaction (in a scale from 1-5) through the variables: perceived leadership (PLeadership: a person that is involved, directs the group) and perceived dominance (PDominance: a person that is in a position of power, dominates). The questionnaire contains 16 items, from which 4 variables are measured capturing leadership, dominance, competence and liking. Additionally, questionnaires from self-reported personality are available. In this study, we concentrate only on the perception of leadership and dominance. The scores from the perceived interaction are averaged considering the number of participants in the group to have a group perception score, such that the participant with the highest score in PLeadership is considered as the emergent leader in the group. Similarly, the participant with the highest score in PDominance is considered the dominant person in the group.

## 4. AUTOMATIC ANALYSIS OF EMERGENT LEADERSHIP

This section details the procedure followed in order to obtain the word categories from the conversations and estimation techniques that we use. Among the goals that we address, automatically extracted word categories are needed in order to infer the emergent leader and the perceived dominant person in the group using a supervised model.

### 4.1 Full Transcription

For each audio recording, we generate the full transcription of the conversations. The transcriptions were generated manually, performed by a professional company from only

the audio recordings. The level of the transcriptions includes time stamps, gender, natural utterances, description of events like crosstalk or laughter. Additionally, whenever there were long pause segments (silence), they are interpreted as sequences of dots (e.g., a 5 second pause of speech is captured as “.....”) and unintelligible speech or inaudible words are described as “xxxx”. Few manual transcriptions failed in tracking speakers due to similarities in tone of voice. For the incorrectly transcribed recordings, the corresponding video files were used to correct the transcriptions. Each group transcription is segmented to have individual speech transcription files per participant.

### 4.2 Keyword Spotter

As a keyword spotting system (KWS), we use a relatively complex system based on Large Vocabulary Continuous Speech Recognition (LVCSR). Our LVCSR is built on HMM/GMMs and is trained on 16 kHz multi-distant microphone recordings from several standard meeting corpora (ICSI, NIST, AMI) [11]. The LVCSR system decodes the input speech in several passes. The acoustic models are trained discriminatively and in speaker adaptive manner. We also use state-of-the-art discriminatively trained posterior-based speech features trained using Neural Networks.

During the decoding, a 50k dictionary is used together with a 3-gram Language Model. Such system reaches a Word Error Rate (WER) of about 3% on the well-known Wall Street Journal task (reading Wall Street Journal Sentences, in this case independent-head microphone recordings provided by same datasets are used for training of acoustic models)<sup>1</sup>.

To perform the search of selected keywords in meeting recordings, the recordings are first pre-processed using the LVCSR system that produces word recognition lattices (representing the set of most probable hypotheses generated by the LVCSR). The word lattices are then converted into a candidate term index accompanied with times and detection scores. The detection scores are represented by the word posterior probabilities, estimated from the lattices using the forward-backward reestimation algorithm [8], and defined as:

$$P(W_i; t_s, t_e) = \sum_Q P(W_i^j; t_s, t_e | x_{t_s}^{t_e}), \quad (1)$$

where  $W_i$  is the hypothesized word identity spanning the time interval  $t \in (t_s, t_e)$ ;  $t_s$  and  $t_e$  denote the start and end time interval, respectively;  $j$  denotes the occurrence of word  $W_i$  in the lattice;  $x_{t_s}^{t_e}$  denotes the corresponding partition of the input speech (the observation feature sequence) and  $Q$  represents a set of all word hypotheses sequences in the lattice that contain the hypothesized word  $W_i$  in  $t \in (t_s, t_e)$ .

Keyword spotting system is performed on our full corpus to extract verbal cues from all the recordings. Obtained manual transcriptions are then used to evaluate the keyword spotter on target data.

Our recordings are acoustically very challenging due to several reasons:

- the corpus is recorded using multi-distant microphones,

<sup>1</sup>[www.amiproject.org/ami-scientific-portal/documentation/annual-reports/pdf/D4.2.pdf/view](http://www.amiproject.org/ami-scientific-portal/documentation/annual-reports/pdf/D4.2.pdf/view)

- there are many cross-talk segments created by multiple speakers,
- the interaction represent an open vocabulary scenario.

Since the obtained manual transcriptions are not segmented (according to time into speech/silence), we evaluate the keyword spotting system only in terms of ASR using WER. The achieved WER is about 60%. Although this is a quite high number, it also includes errors due to many cross-talks and mainly due to lack of speech/non-speech segmentation (many insertions in non-speech parts of recordings appear and are used for scoring). By applying automatic segmentation, keyword spotter could perform reasonably well as it has been shown on real lecture recording scenarios [19].

### 4.3 Automatic Analysis of Verbal Content

In a spontaneous conversation, people do not control often how or when to use pronouns, articles or auxiliary verbs, but their unconscious use has an impact in the listener and reflects the individual linguistic style. For instance, the preferences in the use of function words have been proved to be related to individuals with skills to socialize [7].

In order to analyze the linguistic style by using high level word categories, we use LIWC to process the full manual transcriptions per participant. Similarly, we extract the word categories from the verbal content captured with the keyword spotter, using only the words with high score posterior probability. In order to keep most accurate words, we use a high threshold score from -1 to 0, where 0 is the highest score for a given word (100% accuracy).

The LIWC dictionary is composed of almost 4,500 words [21]. Each word belongs to one or more word categories. For example, the word “agree” is part of three word categories: *affect*, *posemo* and *assent*. So, whenever the word “agree” is found, the scores in the categories *posemo* and *assent* will be incremented. Additionally, the positive emotion category (*posemo*) belongs to a higher category that considers all the emotion words (*affect*), so this general category will be also incremented. More details on the categories and the dictionary can be found in [2].

For the extraction of categories, we consider 64 categories. Since the scenario is a conversation, punctuations are not considered as relevant as the verbal content per se, thus discarded from the analysis. The results from the categories are interpreted as percentages of full content (speech transcription or keywords per participant), except for the *WC* category (word count) which only accumulates the total number of words per participant.

Considering the resulting 64 categories from the full transcription, 12 categories were discarded from the analysis due to low to null occurrences (e.g. money, religion, etc.). This gives 52 categories for analysis.

As a first exploratory step, we performed a correlation analysis in order to see the potential of applying a classification method. A Pearson correlation was computed per group, followed by a Fisher transformation, then a T-test was applied to validate the significance, and the mean correlation values were computed using the Fisher inverse function.

### 4.4 Feature Selection

To analyze the relevance of the features coded from the word categories described in Section 4.3, we use SVM-RFE

based on a 5-fold cross-validation approach.

Considering the 52 categories obtained from LIWC, we performed SVM-RFE in order to obtain the top most informative features. For the process, we used two SVM-RFE respectively for groups with three and four participants, considering as Emergent-Leader (class 1) and Non-Emergent-Leader (class 0), and similarly for Perceived-Dominant (class 1) and Non-Dominant (class 0). The labels are assigned considering the outputs from the questionnaires described in Section 3, such that the person with the highest score in perceived leadership is assigned with the label Emergent-Leader per group. Similarly, the perceived dominant label is assigned. We use a 5-fold cross validation approach to obtain the top ranked features.

### 4.5 Inferring the Emergent Leader and the Perceived Dominant Person

To infer the emergent leader in the group and the perceived dominant person, we use a supervised method in which the input is a feature vector composed of the top 20 categories selected from the SVM-RFE described in the previous section.

We employ a linear kernel SVM ( $k(x, x') = \langle x, x' \rangle$ ) by using a leave-one-meeting-out approach. As SVM outputs, we obtain posterior probabilities instead of class labels [23], resulting from fitting a sigmoid function

$$P(\text{class} = 1|d) = \frac{1}{1 + e^{Ad+B}}, \quad (2)$$

where  $d$  are the decision values of the SVM, and  $A$  and  $B$  are estimated by minimizing the negative log-likelihood function. The outputs are then interpreted such that we assign only one Emergent-Leader label (class 1) per group, computed by:

$$\text{Emergent} - \text{Leader}^f = \arg \max_i (P_i(\text{class} = 1|d)), \quad (3)$$

where  $f$  represents the input feature vector (composed of the top word categories),  $nPart$  is the number of participants in the group, and  $i = 1, 2, \dots, nPart$ . Finally, the emergent leader in the group is the one with the highest posterior probability belonging to class 1. We follow similar procedure for the perceived dominant person in the group.

Considering the amount of participants and meetings in the corpus, the random accuracy in this case is 27.6%, given that we have 20 meetings with four participants and 9 meetings with three participants.

## 5. RESULTS

In this section we present the results for the prediction of the emergent leader in the group and the perceived dominant person using automatically extracted features. First, we present the correlation analysis results using the manual transcriptions. Then we present the resulting top categories using SVM-RFE and their accuracy in the inference of the emergent leader.

### 5.1 Leadership and Dominance Language Styles

In this section we present correlations of the top word categories and the perceived leadership and dominance. The Pearson correlations are computed per group, followed by

Category	PLeadership	PDominance
WC	0.819*	0.680*
assent	-0.747*	-0.549*
funct	0.645*	0.568*
WPS	0.503*	0.301
article	0.415	0.564*
time	0.400 <sup>+</sup>	0.466
conj	0.391 <sup>+</sup>	0.562*
work	0.349 <sup>+</sup>	0.379 <sup>+</sup>

**Table 1: Correlation values between word categories from the manual transcription and perceived variables PLeadership and PDominance. Significance values  $+$ :  $p < 0.05$ ,  $*$ :  $p < 0.01$ .**

a Fisher transformation. Then, mean correlation values are computed, followed by a T-test that validates the significance. As a final step, we compute the inverse Fisher transformation of the mean values.

Table 1 shows only the top high correlated values for the two variables. As we can observe, for the variable PLeadership the highest correlation corresponds to the category *WC* (word count), followed by the *assent* category. The negative correlation suggest that the perceived leaders use less words from this category (for instance “agree”, “mm\*”, “ok”, “yeah”, “yes”). Further, the *WPS* category (words per sentence) shows a high correlation with PLeadership, suggesting that the emergent leader produces longer sentences. Similarly for the PDominance case, the highest correlation is again with the category *WC*, followed by the *funct* category (total function words, this category includes pronouns and articles).

## 5.2 Feature Selection

The features described in this section correspond to the resulting outputs from the SVM-RFE based on a 5-fold cross-validation approach. The features correspond to the top word categories from the manual speech transcription and keywords. For the analysis, we used separated SVM-RFE training and test sets for groups with three and four participants.

### 5.2.1 Full transcription

In the full transcription case, for groups with four participants the top 20 ranked features that discriminate between Emergent-Leaders and Non-Emergent-Leaders are shown in Table 2, left. Similarly, the top 20 ranked features according to the SVM-RFE that can discriminate between the classes Perceived-Dominant and Non-Dominant are listed in Table 2. As we can observe, the ranking of relevance of the categories differs between the two sets (i.e., for groups with three and four participants).

Although the orders in the ranking of the features for the variable PLeadership between the groups differ, there are 9 categories that appear to be relevant in the two sets. The similar categories in PLeadership capture linguistic processes (*WC*, *article*, *verb* and *conj*), cognitive processes (*certain*), perceptual processes (*see* and *hear*), biological processes (*bio*) and a spoken category (*assent*). Similarly for PDominance, from the 20 top categories, there are 9 common relevant categories for the two groups although in different ranking. The categories include linguistic processes (*WC*, *article* and *negate*), affective processes

Top	PLeadership		PDominance	
	Gp 3	Gp 4	Gp 3	Gp 4
1	conj	WC	conj	WC
2	ingest	see	incl	motion
3	negate	hear	article	tentat
4	body	you	WC	negate
5	filler	achieve	hear	excl
6	funct	article	negate	filler
7	WC	adverb	assent	we
8	see	quant	verb	achieve
9	negemo	certain	social	certain
10	assent	bio	discrep	incl
11	bio	assent	ipron	space
12	hear	we	ingest	article
13	excl	pronoun	funct	Sixltr
14	discrep	social	i	you
15	verb	conj	cause	ingest
16	certain	death	negemo	hear
17	present	verb	work	feel
18	article	affect	motion	negemo
19	future	cause	nonfl	quant
20	space	WPS	tentat	future

**Table 2: Top 20 word categories from the SVM-RFE, resulting from categories extracted from the manual transcription.**

(*negemo*), cognitive processes (*incl* and *tentat*), perceptual processes (*hear*), biological processes (*ingest*) and relativity (*motion*).

Note also that PLeadership and PDominance have categories in common. Although this could suggest that both concepts share a language style, there are also clear differences in the top categories that captures PDominance (among groups with three and four participants). For example, the categories *motion*, *negate* (*negations*), *tentat* (*tentative*), *ipron* (*impersonal pronouns*) and *negemo* (*negative emotion*), that captures PDominance are not relevant for PLeadership.

### 5.2.2 Keywords

The top ranked categories derived from the most accurate keywords are shown in Table 3. As we can observe, the category word count (*WC*) is the most relevant category for the Emergent-Leader class in the two sets of groups. In contrast for the variable PDominance, only for groups with four participants the category *WC* is relevant, and for groups with three participants *WC* is not in the top 20 ranked features.

Considering the top categories for the variable PLeadership, only 4 categories are similar among the two set of groups (see Table 3, left). For the case of PDominance, 8 categories are commonly relevant among the groups in the top 20 (see Table 3, right).

Although the automatic keyword spotter can not recover the full conversation (due to overlaps, words that are not in the dictionary, and other reasons) it captures the interaction by recovering some top categories from the full transcription.

For groups with three participants, the categories extracted from the keywords for the variable PLeadership are consistent with the categories considering the full transcription, such that there is an overlap of 8 categories among the top 20. Additionally, some other categories are captured,

Top	PLeadership		PDominance	
	Gp 3	Gp 4	Gp 3	Gp 4
1	WC	WC	affect	assent
2	incl	relativ	you	WC
3	time	conj	adverb	cogmech
4	adverb	posemo	posemo	time
5	Dic	space	quant	funct
6	achieve	negemo	filler	bio
7	future	hear	work	past
8	ingest	we	we	negate
9	excl	pronoun	future	inhib
10	social	work	time	leisure
11	present	achieve	verb	work
12	quant	i	tentat	filler
13	filler	cogmech	relativ	adverb
14	auxverb	funct	Sixltr	motion
15	cause	feel	leisure	ipron
16	we	discrep	certain	relativ
17	insight	tentat	social	social
18	article	bio	negate	percept
19	past	negate	Dic	insight
20	conj	social	assent	verb

**Table 3: Top 20 word categories from the SVM-RFE, resulting from categories extracted from the keyword spotter.**

if the hierarchical category is considered. For example the *verb* category (top 15 from the full transcription in Table 2), includes the subcategories *adverb* and *auxverb*, that are captured in the top 20 categories from the keywords (see Table 3, top 4 and 14). For the case of four participants, 7 out of the top 20 categories are recovered, and it also captures subcategories that appear in high categories in the full transcription, e.g., the *posemo* and *negemo* categories (top 4 and 6 from Table 3) are subcategories of the *affect* category (top 18 from Table 2).

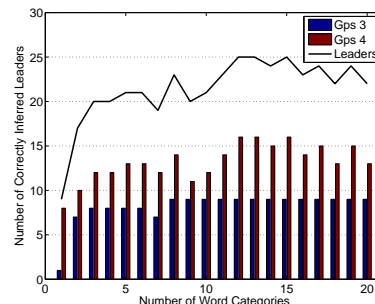
Similarly for the variable PDominance, 6 categories overlap considering the top ranked categories for groups with three participants. For groups with four participants, although only 4 categories overlap one-to-one, higher categories captured from the keywords appear as relevant categories in the full transcription, e.g., *cogmech* category (top 3 from Table 3) includes the subcategories *tentat*, *excl*, *certain* and *incl* (tops 3, 5, 9 and 10 respectively from Table 2).

As we can observe from Tables 2 and 3, the top ranked categories capture the discussion, such that the interaction involves decisions about motion, time and the needs in order to survive an accident. Although some top categories are linked to the context, other categories more related to the language style (free of the context) are relevant in order to discriminate between Emergent-Leaders and Non-Emergent-Leaders (e.g., *WC*, *conj*, *excl* and *assent*). While the *WC* does not exactly reflect language style, it provides relevant information in the perception of leadership and dominance. Despite the fact that the keyword spotter can not recover exactly the top 20 relevant categories for PLeadership (from the manual transcription), it captures a significant number of top categories in a one-to-one-base (7 categories out of 20).

### 5.3 Inference with Manual Transcriptions

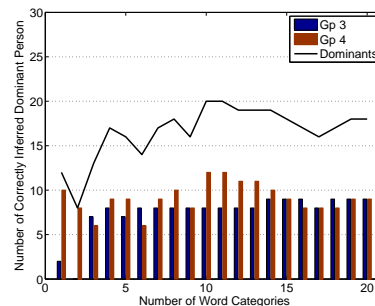
The results in this section show performance on the top 20 features obtained from the SVM-RFE and the manual transcriptions described in the previous section.

Figure 3 shows the performance of the perceived emergent leader inference method, using categories derived from the full transcription. As we can observe, the best overall performance is 86.2% reached by using the top 12 and 13 ranked features, having 25 correctly inferred emergent leaders, out of the total of 29 leaders in the corpus. For groups with three participants, the best performance is already reached starting with the top 8 features. For the case of four participants, more features are needed to achieve the highest accuracy (12 features).



**Figure 3: Performance of Perceived Emergent Leaders in the corpus, considering the full transcription.**

For the perceived dominant person in the group (Fig. 4), the top 10 and 11 ranked features provide the best achieved performance, namely 69.0%, i.e., 20 dominant participants are correctly inferred. As we can observe for groups with four participants, one single feature provides 50% accuracy (the category *WC*), and the best performance is reached with the top 10 and 11 ranked features. For the case of groups with three participants, more features are needed, starting from top 14 on, to get the best performance, inferring accurately all the 9 perceived dominant participants in the three person groups.



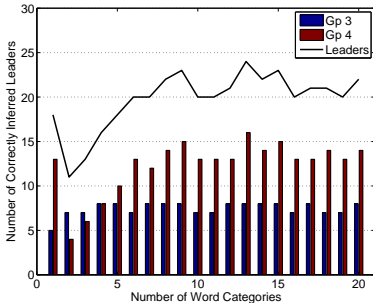
**Figure 4: Performance of Perceived Dominant person in the corpus, considering the full transcription.**

### 5.4 Inference with Keywords

The results in this section show performance for the resulting top 20 features from the SVM-RFE and the keywords described in Section 5.2.2.

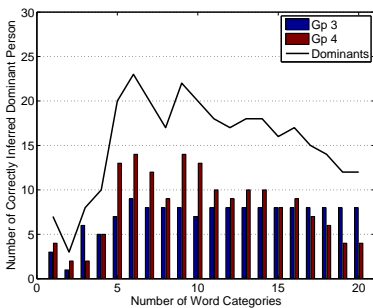
Figure 5 shows the performance of the perceived emergent leader inference method, using the most accurate detected keywords. As we can observe, the best performance is 82.8%, reached by using the top 13 ranked features, having 24 correctly inferred emergent leaders, out of 29 in the

corpus. This performance is only marginally lower than the one obtained with manual transcriptions. For groups with three participants, the top 4-5 features provide the same performance as using the top 12 to 15. For groups with four participants although the performance is relatively high with only one feature (*WC*, the 13 out of the 20 perceived leaders), 13 top ranked features are needed to reach the highest performance (16 out of 20).



**Figure 5: Performance of Perceived Emergent Leaders in the corpus, considering only keywords**

Finally, for the perceived dominant person in the group, the best performance is 79.3% by considering the top 6 categories, which results in 23 correctly inferred most dominant participants out of the 29. Surprisingly the performance using keywords is higher than the one obtained with manual transcriptions. This performance could be due to the fact that the keyword spotter could overestimate words from categories that are highly correlated with dominance, e.g. *negemo* (negative emotions). As an aside research line, we plan to explore the hypothesis that the keyword spotter could overestimate words from specific word categories.



**Figure 6: Performance of Perceived Dominant person in the corpus, considering only keywords.**

## 5.5 Comparison with Previous Literature

The emergent leadership phenomenon as well as the dominance trait have been explored previously in social psychology and social computing areas. Psychologists have shown strong correlations between leadership and participation (amount of time that a person spent talking) derived from manual transcriptions or filtered audio [27]. Also, high amount of speaking time within small group interactions is correlated with dominance [18].

Based on previous findings from psychologists and in order to have a relative comparison in the context of predicting the emergent leader using as feature the amount of speaking time. First, we computed the amount of speaking time

per participant, derived from the automatic speech/non-speech segmentation provided by the microcone [1]. We then used an unsupervised method that infers the leader in the group. The inference selects the participant with the maximum amount of speaking time in the group as the emergent leader. Such method gives us an accuracy of 72.5%, which performs relatively similar as reported in the literature in the social computing area. For example, 60% accuracy was reported in [25] considering only amount of speaking time in order to predict leadership and up to 72% accuracy with more prosodic features and using complex machine learning techniques.

For the perceived dominant person, our accuracy using a rule based inference using as feature the amount of speaking time reaches about 55.2%. Previous works reported accuracies ranging from 40 to 85%, considering only speaking time feature [25, 13]. The performance increases considerably, ranging from 75-90% accuracy when more prosodic features are used with more complex supervised methods [24, 5, 13].

## 6. DISCUSSION AND CONCLUSIONS

In this paper, we presented the language style that characterize people being perceived as leaders and dominants. For the language analysis, word categories were extracted from manual transcriptions and an automatic keyword spotter using 7 hours of recorded interactions. By using the top relevant word categories and a supervised method, we inferred the emergent leader and the perceived dominant person in the group.

Our analysis on the relevance of categories shows that perceived leadership uses words related to *motion* (*to walk or to stay*), *space* (*map, far, close, etc*) and *basic needs* (*eat, sleep, etc*). This is clearly scenario-dependent and suggests that the emergent leader gets involved in the scenario, in this case the decision making process of how to survive an airplane crash in winter. Although some categories are linked to the context, other relevant categories like *WC*, *conj*, *excl* and *assent* (i.e., free of the context and non-privacy sensitive) are informative in order to discriminate the emergent leader.

While the *WC* feature (word count) does not exactly reflect language style, it provides relevant information in the perception of leadership and dominance. The *WC* category can be considered as a feature correlated with the amount of speaking time. In that sense, our findings are consistent with previous work, such that the amount of speaking time is highly correlated with dominance [18, 13] and the perception of a leader [27, 25].

Despite the fact that the keyword spotter can not recover exactly the top 20 relevant categories for leadership extracted from the manual transcription, it captures a significant number of the top categories on a one-by-one-base (7 categories out of 20). The performance of the keyword spotter regarding inference of the emergent leader is promisingly high (82.8%), as compared with the performance of the manual transcription (86.2%). For the case of perceived dominance, the use of the manual transcription provides only 69% accuracy, while for the case of keyword spotter output, a better performance is reached (up to 79.3%). The results using an automatic keyword spotter suggest that accurate inferences can be done in a relatively short amount of time. It is worth to mention that the performance of the keyword spotter is high, considering that the audio track is challeng-

ing due to an open vocabulary scenario and many cross-talk segments. Finally, a keyword spotting system can be implemented in a realistic scenario (in an automated and fast process).

We can detect several future research lines. First, the verbal content has been analyzed considering only English spoken meetings, such that we can not generalize our findings in the perception of emergent leadership to other spoken languages. More data would be needed to confirm our findings. Second, prosodic features could be used as additional source of information, more specifically to augment information derived from categories that capture positive or negative emotions.

**Acknowledgments:** This research was supported by CONACYT (Mexico) through a doctoral scholarship, the SNSF project SONVB and the SNSF National Center of Competence in Research (NCCR) on Interactive Multimodal Information Management (IM)2.

## 7. REFERENCES

- [1] Dev-audio pty ltd. <http://www.dev-audio.com/>.
- [2] Liwc inc. <http://www.liwc.net/index.php>.
- [3] K. Bachour, F. Kaplan, and P. Dillenbourg. An Interactive Table for Supporting Participation Balance in Face-to-Face Collaborative Learning. *IEEE Transactions on Learning Technologies*, 3(3):203–213, 2010.
- [4] T. Carte, L. Chidambaram, and A. Becker. Emergent leadership in self-managed virtual teams. *Group decision and negotiation*, 15(4):323–343, 2006.
- [5] M. Charfuelan and M. Schroder. The vocal effort of dominance in scenario meetings. In *Interspeech*, 2011.
- [6] G. Chittaranjan, J. Blom, and D. Gatica-Perez. Mining large-scale smartphone data for personality studies. *Personal and Ubiquitous Computing*, pages 1–18, 2012.
- [7] C. Chung and J. Pennebaker. The psychological functions of function words. In *Social communication*, pages 343–359. Psychology Press, 2007.
- [8] G. Evermann and P. Woodland. Large vocabulary decoding and confidence estimation using word phoneme accuracy posterior probabilities. In *International Conference on Acoustics, Speech and Signal Processing*, volume 3 of *ICASSP*, pages 2366–2369, Jun 2000.
- [9] D. Gatica-Perez. Analyzing group interactions in conversations: a review. In *International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pages 41–46, sep 2006.
- [10] S. A. Golder and M. W. Macy. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science*, 333(6051):1878–1881, 2011.
- [11] T. Hain, L. Burget, J. Dines, G. Garau, V. Wan, M. Karafiat, J. Vepa, and M. Lincoln. The ami system for the transcription of speech in meetings. In *International Conference on Acoustics, Speech and Signal Processing*, volume 4 of *ICASSP*, pages 357–360, Apr 2007.
- [12] M. E. Ireland, R. B. Slatcher, P. W. Eastwick, L. E. Scissors, E. J. Finkel, and J. W. Pennebaker. Language style matching predicts relationship initiation and stability. *Psychological Science*, 22(1):39–44, Jan 2011.
- [13] D. Jayagopi, H. Hung, C. Yeo, and D. Gatica-Perez. Modeling dominance in group conversations using nonverbal activity cues. *IEEE Transactions on Audio, Speech and Language Processing*, 17(3), 2009.
- [14] M. Jensen, T. Meservy, J. Burgoon, and J. Nunamaker. Automatic, multimodal evaluation of human interaction. *Group Decision and Negotiation*, 19:367–389, 2010.
- [15] D. W. Johnson and F. P. Johnson. *Joining together: Group theory and group skills*. Boston: Allyn and Bacon, 1994.
- [16] T. Kim, A. Chang, L. Holland, and A. Pentland. Meeting mediator: enhancing group collaboration with sociometric feedback. In *Conference on Computer Supported Cooperative Work, CSCW*, 2008.
- [17] F. Mairesse, M. A. Walker, M. R. Mehl, and R. K. Moore. Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research*, 30:457–501, 2007.
- [18] M. S. Mast. Dominance as expressed and inferred through speaking time: A meta-analysis. *Human Communication research*, 28(3):420–450, 2002.
- [19] P. Motlicek, F. Valente, and P. N. Garner. English spoken term detection in multilingual recordings. In *Interspeech*, Sep 2010.
- [20] D. Olguin, B. Waber, T. Kim, A. Mohan, K. Ara, and A. Pentland. Sensible organizations: Technology and methodology for automatically measuring organizational behavior. *Systems, Man, and Cybernetics, IEEE Transactions on*, 39(1):43–55, 2009.
- [21] J. Pennebaker, M. Francis, and R. Booth. *Linguistic Inquiry and Word Count: LIWC2001*. Mahwah, NJ: Erlbaum Publishers, 2001.
- [22] J. Pennebaker and L. King. Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, 77(6):1296–1312, 1999.
- [23] J. C. Platt. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In *Advances in Large Margin Classifiers*, pages 61–74. MIT Press, 1999.
- [24] R. Rienks, D. Zhang, D. Gatica-Perez, and W. Post. Detection and application of influence rankings in small group meetings. In *International Conference on Multimodal Interfaces, ICMI*, 2006.
- [25] D. Sanchez-Cortes, O. Aran, M. S. Mast, and D. Gatica-Perez. A nonverbal behavior approach to identify emergent leaders in small groups. *IEEE Transactions on Multimedia*, 14(3):816–832, 2012.
- [26] R. B. Slatcher, C. K. Chung, J. W. Pennebaker, and L. D. Stone. Winning words: Individual differences in linguistic style among u.s. presidential and vice presidential candidates. *Journal of Research in Personality*, 41(1):63 – 75, 2007.
- [27] R. T. Stein. Identifying emergent leaders from verbal and nonverbal communications. *Personality and Social Psychology*, 32(1):125–135, 1975.
- [28] P. Temdee, B. Thipakorn, B. Sirinaovakul, and H. Schelhowe. Of collaborative learning team: An approach for emergent leadership roles identification by using social network analysis. In *Technologies for E-Learning and Digital Entertainment*, volume 3942 of *Lecture Notes in Computer Science*, pages 745–754. 2006.