



HAL
open science

Frequent Itemsets and Clustering Unravel the Evolution of Evo Morales Discursive Strategies

Sergio Peignier, Patricia Zapata

► **To cite this version:**

Sergio Peignier, Patricia Zapata. Frequent Itemsets and Clustering Unravel the Evolution of Evo Morales Discursive Strategies. 2020. hal-03022614v1

HAL Id: hal-03022614

<https://hal.science/hal-03022614v1>

Preprint submitted on 24 Nov 2020 (v1), last revised 25 Nov 2023 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Frequent Itemsets and Clustering Unravel the Evolution of Evo Morales Discursive Strategies

Sergio Peignier¹ and Patricia Zapata²

¹*Univ Lyon 1, INSA Lyon, INRA , BF2I, UMR0203, F-69621, Villeurbanne, France , sergio.peignier@insa-lyon.fr*

²*Universidad Mayor de San Andrés, Carrera de Lingüística e Idiomas, La Paz, Bolivia , pzapata@umsa.bo*

1 Introduction

Evo Morales, is a Bolivian politician, leader of the Movement for Socialism party, and former president of Bolivia from 2006 to 2019. His administration has been part of the latin-american Socialism of the 21st century. The leaders belonging to this movement are characterized by their ability to use their rhetorical skills to seduce people. Indeed, Morales has been described as a “charismatic populist” [1], who aimed at identifying himself with his audience [2], by means of a populist discourse [3]. Indeed, his speeches are organized as a Triadic Scenario where people, guided by Morales himself, fight against a common enemy (e.g., political class, elite), which is responsible for a disastrous social situation [3, 2]. Moreover, he often refers to social welfare and economic progress as a powerful recruitment process strategy[2], in order to stir positive feelings in the audience, as a rhetoric of effects. And thus, studying Morales’ rhetoric skills is a crucial step towards understanding his political success. Interestingly, discursive strategies change according to external events (e.g., during crisis [4, 5]) and evolve along time, as shown for instance in [6]. Here, we studied the evolution of Morales’s discursive strategies combining frequent itemsets mining and clustering, with the linguistic speech analysis methodology of P. Charaudeau.

In this work, we studied 477 tweets from Evo Morales, posted between the 1st of November and the 2nd of December 2019, i.e., during the particularly turbulent period, that characterized the end of his administration.

Dates	Events
10/20	General Elections (first results suggest the need of a 2nd tour).
10/21	Live vote count interruption, opposition denounces electoral fraud.
10/22	General strike proclaimed, demonstrations/riots until 11/10. Clashes between opponents vs. police and Morales supporters.
10/23	Morales' government asks the OAE to audit the elections.
10/25	Morales is re-elected without needing a 2nd tour.
10/31	OAE electoral audit begins.
11/08	Police mutiny.
11/09	Government call to convene a dialogue, but the opposition refused.
11/10	Preliminary OAE report confirms intentional electoral manipulation. Workers union, police and army call Morales to resign. Morales and senior members of the government resign. Pro-Morales riots until 11/22, clashes with police and army.
11/11	Morales leaves Bolivia.
11/12	Senator J. Añez assumes the presidential office.

Table 1: Timeline of main events since Bolivian general election

2 Materials and Methods

2.1 Dataset

Context During October and November 2019, Bolivia experienced a dramatic socio-political crisis, and underwent major political changes, as summarised in the timeline 1. This period can be separated in three major intervals:

- 1) Between the 1st and the 20th of October, Bolivia underwent a political campaign towards general elections.
- 2) Between the 21st of October to the 10th of November, Bolivia experienced a socio-political post-electoral crisis.
- 3) Between the 11th and the end of November, Bolivia went through an administration change and an important socio-political crisis.

Therefore characterizing Morales' discursive strategies during this period is an important step towards interesting the mechanisms driving the evolution of discourse under political crisis.

Dataset description This paper studies the evolution of Evo Morales' discursive strategies on Twitter, by focusing on the tweets he posted between the 1st of October 2019 and the 2 of December 2019. In practice, the tweepy python library [7] was used to download 477 Morales' tweets, containing 12871 words in total, and representing a vocabulary of 3678 distinct words.

2.2 Methodology

Pre-processing As in many Natural Language Processing applications, a first step consisted in filtering stop-words and rare-words. Stop-words filtering aims

at removing too common words conveying little semantic information, whereas rare-words filtering aims at removing “noisy words” that are not used often. Here, we selected a curated set of 34 stop-words¹, and words that were only used once in the dataset were considered rare, and were removed accordingly. This step was ensured using custom Python scripts based on the nltk library [8]. The pre-processed dataset contains 10693 words in total, and a vocabulary with 1500 distinct words.

Frequent itemsets extraction In order to detect sets of words that are frequently used together, we used a frequent itemset mining procedure.

More formally, let a tweet s be an ordered list of words from vocabulary $V = \{w_1, w_2, \dots\}$, and let a tweets dataset be $S = (s_1, s_2, \dots)$ an ordered list of tweets s_i received at time i . The dataset vocabulary V , denotes the set of possible items, the transactions dataset is simply $T = (t_1, t_2, \dots)$, with $t_i = \{w \in s_i\}$ for each tweet $s_i \in S$, and let $m \subseteq V$ denote an itemset. The support $supp(m, T)$ of an itemset m in T , is simply the frequency of transactions in T containing m , i.e., $supp(m, T) = \frac{|\{t \in T \mid m \subseteq t\}|}{|T|}$.

In practice, we used the popular FP-growth algorithm [9], to extract 2552 frequent itemsets $\mathcal{M} = \{m_1, m_2, \dots\}$ such that, $\forall m \in \mathcal{M}$, $supp(m, T) \geq 0.015$.

Frequent itemsets selection The frequency (i.e., number of occurrences) of word w in the dataset T is denoted $freq(w, T)$, and thus the total number of words in T is $N = \sum_{w \in V} freq(w, T)$. The coverage of a set of words m in T is $coverage(m, T) = \frac{\sum_{w \in m} freq(w, T)}{N}$.

In order to select an informative subset of itemsets $\mathcal{M}^* \subseteq \mathcal{M}$ with size $|\mathcal{M}^*| = M$, we select iteratively the itemsets leading to the highest gain in coverage, using the greedy strategy described in Algorithm 1.

In practice, selecting a subset \mathcal{M}^* with size $M = 200$ and vocabulary V^* , led to a overall $coverage(V^*, T) = 0.613$. The cumulative gain in coverage using Algorithm 1, is represented Figures 1.

Algorithm 1 Itemsets selection

```

1: procedure GREEDYITEMSETSSELECTION( $T, \mathcal{M}, M$ )
2:    $\mathcal{M}^* \leftarrow \emptyset$  ▷ Initialize  $\mathcal{M}^*$ 
3:   for  $i \leftarrow 1$  to  $M$  do
4:      $V^* \leftarrow \bigcup \mathcal{M}^*$  ▷ Vocabulary in  $\mathcal{M}^*$ 
5:      $m^* \leftarrow \arg \max_{m \in \mathcal{M}} (coverage(m \setminus V^*))$  ▷ Highest coverage itemset
6:      $\mathcal{M}^* \leftarrow \mathcal{M}^* \cup \{m^*\}$  ▷ Update  $\mathcal{M}^*$ 
7:   return  $\mathcal{M}^*, V^*$ 

```

¹Mainly demonstrative adjectives, indefinite and definite articles

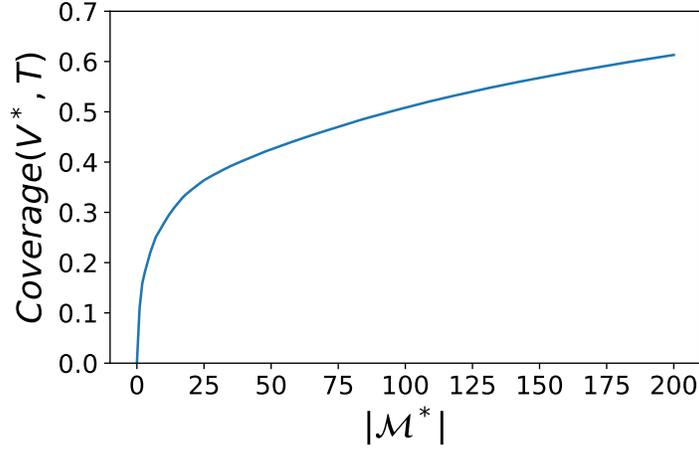


Figure 1: Model Coverage gain vs. number of itemsets

Itemsets clustering To make sense of the selected itemsets, we have assessed their presence in the different tweets. To do so, for each itemset m , we have computed a boolean vector $B^m = (\mathbb{1}_m(t), \forall t \in T)$, where $\mathbb{1}_{m \subseteq V}(t \subseteq V)$ is an indicator function such that $\mathbb{1}_m(t) = 1$ if $m \subseteq t$, and $\mathbb{1}_m(t) = 0$ otherwise.

In order to highlight vectors B^m general trends and smooth their short-term fluctuations, we applied a Central Moving Average with a window size $u = 50$, to compute $X^m = CMA_u(B^m)$ the activity vector of each itemset m .

Then we used the well-known K-means algorithm [10], to partition itemsets into k clusters, according to their respective activity vectors $\{X_m, \forall m \in \mathcal{M}^*\}$, so clusters group itemsets that tend to be active simultaneously.

In order to choose the number of clusters, we have computed the Sum of Squared Distance (SSE), between vectors X_m and their respective centroids, for different number of clusters from $K = 1$ to $K = 50$. Then, we applied the Kneedle algorithm [11], to determine the elbow point in the SSE curve (i.e., a classical indicator for choosing the number of clusters), that suggests to take $K = 9$ clusters, as shown Figure 2. Finally the k-means algorithm was executed 1000 times, the clustering leading to the best SSE was selected. Thus, itemsets \mathcal{M} were clustered in $K = 9$ groups ,

Finally, for each cluster $C_k = \{m \in \mathcal{M}\}$, we computed its activity vector $X^{C_k} = \sum_{m \in C_k} X^m$, as the element-wise sum of activity vectors X^m of each itemset m belonging to cluster C_k .

Discourse Analysis Methodology The clusters' itemsets were studied using Patrick Charaudeau's well-known discourse analysis methodology [12]. This methodology, structured according to the Aristotelian classification of Rhetoric art, identifies three families of discursive strategies, called Ethos Pathos and

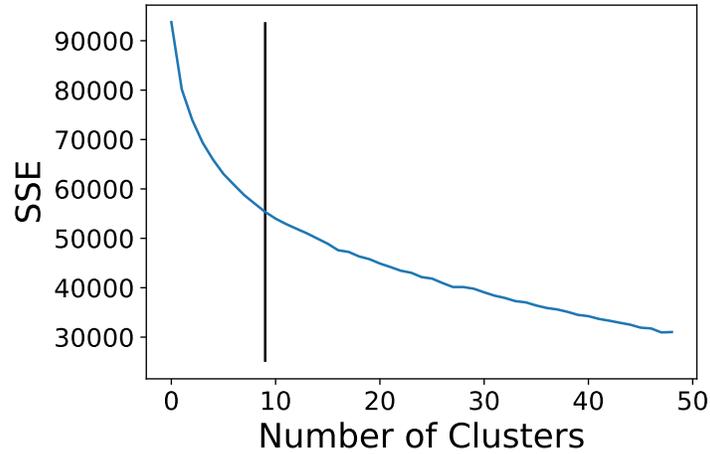


Figure 2: SSE vs. number of clusters, knee point for 9 clusters

Logos.

- Ethos strategies allow the politician to define his discursive identity, and identify himself with the audience, to increase its credibility and charisma.
- Pathos strategies aim at using rhetoric of effects to bringing out feelings and passions, and also recruitment processes to lead the audience to accept the politician’s project willingly.
- Logos strategies rely on rational arguments, mainly based on simplified causal reasoning, and discursive formulas analogous to slogans.

3 Results

Clusters of itemsets are represented in Table 2, and their temporal activities are depicted in Figure 3 as a row-wise z-scored heat-map. According to these results, during his electoral campaign, Morales mainly relies on recruitment processes based on economical progress and social welfare, which often arise feelings of hope and pride (clusters 7, 2 and 1). Furthermore, Morales also used a triadic scenario to oppose neoliberalism and people, and to delegitimize his main opponent Carlos Mesa, by linking him to the former president Gonzalo Sánchez de Lozada, how was overthrown in 2003 by a social revolt, during the so-called “Bolivian gas war”.

Interestingly, at first Morales seems to minimize the social conflicts that arose soon after the elections. Indeed, he keeps developing recruitment processes as before, and claiming his electoral victory (cluster 4). Then, he seems to develop a recruitment process based on nationalism, unity and moral values, in order to

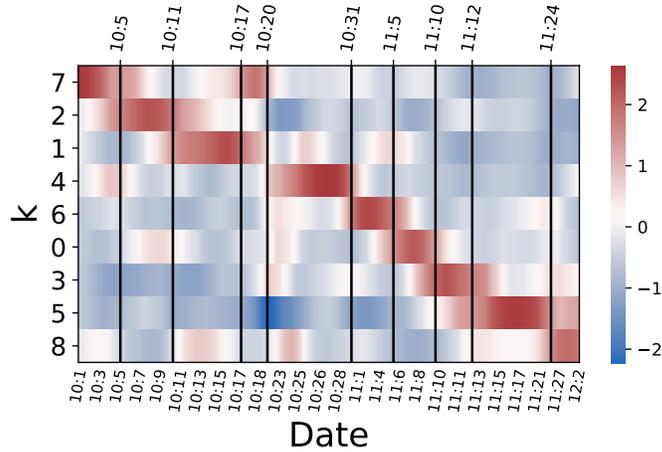


Figure 3: Standardize average use of each cluster along time

support the request he made to the OAE, to audit the electoral process. We can also perceive that Morales employs the word racism, possibly to delegitimize the opposition (cluster 6). Finally in cluster 6, Morales keeps on developing similar recruitment processes, and he also employs many terms related to peace. And thus, he aims a coping with the social crisis by showing himself as a peacemaker. Nevertheless, he simultaneously use verbs in infinitive, with an imperative value, that suggest he calls upon his supporters and people to defend his government.

Finally, after being overthrown, Morales almost completely drops the previous recruitment process strategies, in order to devote almost entirely to the description of a new triadic scenario. On the one hand, he denounces his enemies, (i.e., the army, the police, opponents leaders such as Luis Fernando Camacho and the new president Jeanine Añez), as a member of a dictatorship, and he also denounces their actions as a violent coup. On the other hand, he describes his allies, which would also correspond to the victims, i.e., indigenous and poor people. This description aims at evoking strong feelings in the audience, such as injustice, empathy, piety, fear, sadness and hatred. Moreover it aims at delegitimizing his opponents, by depicting them as a violent and racist group, and it aims at recruiting new allies based on their ethnic and ideological belonging.

4 Conclusion and Discussion

References

- [1] Linda C Farthing and Benjamin H Kohl. *Evo's Bolivia: Continuity and change*. University of Texas Press, 2014.

k	Itemsets
7	familias trabajadores jóvenes nuevas-generaciones neoliberales liberación educación proyectos salud autoridades informó entregamos garantizamos garantiza saludamos gracias·unidad prensa oruro años cada
2	pobreza octubre masacre sánchez-carlos-mesa-lozada cómplices EEUU-gas industrialización-nacionalización muy-futuroseguro estabilidad económica plan seguridad garantizar fortaleza historia mejor gran muchas trabajamos política junto juntos nuestras cariño reunimos sucre homenaje agradecido-apoyo mientras fue así día
1	más-ahora económico-crecimiento trabajo procesodecambio rrnn construcción obras ley inversiones conciencia será grande grandes estamos cierre-campaña unidos vamos concentración santacruz lapaz potosí reportó desde hasta
4	como-es-hoy-bolivia soberanía agua integración desarrollo beneficio solidaridad·hermano nueva tenemos triunfo voto cochabamba argentina países entregar nunca porque todas
6	pueblo-por-para-que familia entre-bolivianos racismo nuestra-querida-su hermanos-sus agradecemos-nos somos auditoría compromiso país político pero solo está primera ni era
0	sectores-sociales profesionales humildes campo ciudad social-paz patria boliviana economía pacíficamente proceso libre hermana hermanas hemos casa nuestro-no defender cuidar trabajar escuchar todo ante manera medio
3	pobres golpe-estado-democracia violencia-mi grupos boliviano derecho políticas pacífica elecciones democrática constitucional cpe rt-evoespueblo mundo llamado evo pido me presidente todos han ha o si nuevo
5	indígenas alto contra-vida represión muertos ffaa-policía áñez-camacho facto golpistas golpista dictadura militares agradezco mis respeto diálogo libertad denuncia comunidad-internacional después tienen
8	pueblos-identidad movimiento-indígena movimientos fuerza derecha dignidad-lucha derechos nacional gobierno-nuestros saludo defensa son sin ser tiene sobre quieren

Table 2: Clusters itemsets coloured by discursive strategy: blue: commitment, authority (Ethos), Yellow: ally (Pathos), Red: enemy (Pathos), Orange: people as victim (Pathos), Green: Recruitment process based on development, welfare, moral values (Pathos), Violet: Identification with people, charisma (Pathos)

- [2] Patricia Zapata and Sergio Peignier. *Análisis del Discurso Socialista Latinoamericano Basado en Inteligencia Artificial*. Instituto Andrés Bello, 2017.
- [3] George Philip and Francisco Panizza. *The triumph of politics*. John Wiley & Sons, 2013.
- [4] Antoon De Rycker and Zuraidah Mohd Don. *Discourse and crisis: Critical perspectives*, volume 52. John Benjamins Publishing, 2013.
- [5] Laura Candón Gautier. El discurso político y su reflejo en los medios: la credibilidad en tiempos de crisis (2008-2011)/political discourse and its reflection in the media: credibility in times of crisis (2008-2011). *Historia y Comunicación Social*, 19:129–137, 2014.

- [6] Serge De Sousa. À l'épreuve des temps... temps lexical et temps politique dans le discours de Fidel Castro (1959-2008). *A. Dister, D. Longré et G. Purnelle (Éds), JADT*, pages 337–349, 2012.
- [7] Joshua Roesslein et al. tweepy. <https://github.com/tweepy/tweepy>, 2019.
- [8] Steven Bird, Ewan Klein, and Edward Loper. *Natural language processing with Python: analyzing text with the natural language toolkit*. " O'Reilly Media, Inc.", 2009.
- [9] Jiawei Han, Jian Pei, Yiwen Yin, and Runying Mao. Mining frequent patterns without candidate generation: A frequent-pattern tree approach. *Data mining and knowledge discovery*, 8(1):53–87, 2004.
- [10] James MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA, 1967.
- [11] Ville Satopaa, Jeannie Albrecht, David Irwin, and Barath Raghavan. Finding a" kneedle" in a haystack: Detecting knee points in system behavior. In *2011 31st international conference on distributed computing systems workshops*, pages 166–171. IEEE, 2011.
- [12] Patrick Charaudeau. Réflexions pour l'analyse du discours populiste. *Mots. Les langages du politique*, (97):101–116, 2011.