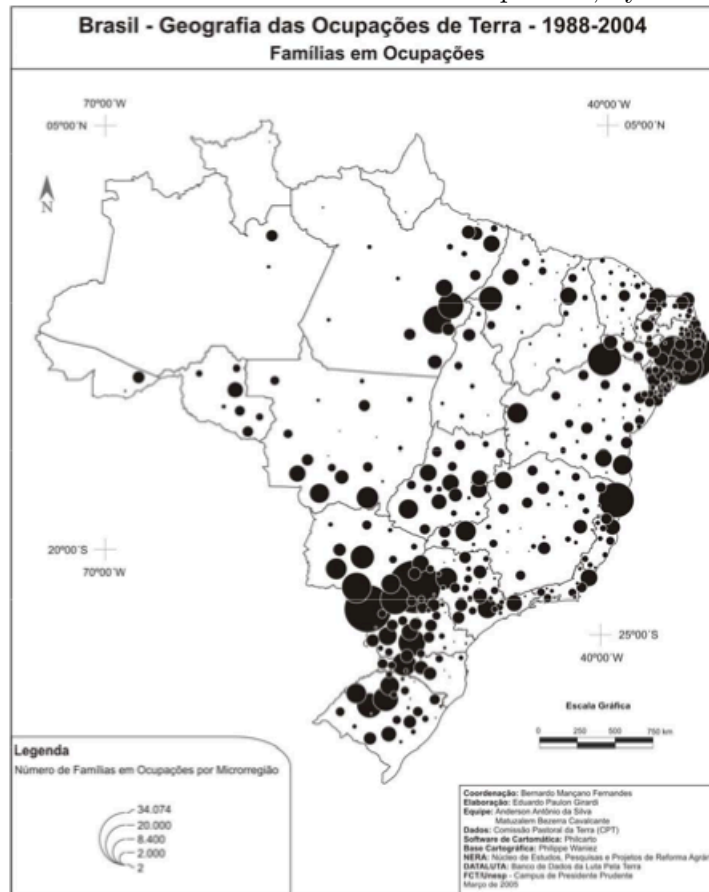


# FINAL PROJECT: CONTENT ANALYSIS

CHAD EVANS

FIGURE 1. Number of MST Families in Occupations, by Micro-region



## 1. INTRODUCTION

The Landless Workers' Movement (MST) is a social movement in Brazil principally advocating for the reformation of the land-property regime. They are also heavily involved

the day-to-day lives of their members, organizing protests, educational opportunities, careers and cooperative economical ventures. Common claims of membership are around 1.5 million people, ranging from children to adults.

Their disapproval of the current land regime is based on a historical perspective emphasizing Brazil's colonial past. The legal apparatus securing property, they argue, has a long and unjust tie to the "latifundio" powers who dominated the Brazilian agricultural economy for centuries. Economic opportunity, then, determined who could acquire title. Those without such opportunity paid the price; a steep one consisting of poverty, hardship and senseless suffering.

The methods the MST uses to challenge the land regime are audacious. The one that has drawn headlines is land "occupations." In this event, MST leadership organizes large groups of families to move onto land not "fulfilling its social purpose." The organization then provides the logistical expertise to grow produce and unite the families in a "struggle for the land" which, in most cases, is resolved in the courts. If a judge finds that the land was not "fulfilling its social purpose," he or she may rule that the families qualify for usufruct rights.

The movement, its operations and the daily lives of its members have been of interest since the eighties when the movement emerged. In the last 15 years, however, the movement has experienced an influx of greater academic attention. Scholars seem to be attracted to the epistemological challenges engendered by the movement's activities as well as the practical matters when such a large organization has an impact on so many people. This academic interest is not unique to scholars in Brazil. Researchers all over the world have sought to investigate research questions in the context of MST activities.

Given such large and growing academic interest in the MST, it would be interesting to see what MST-related ideas scholars in different parts of the world are researching. It is the intent of this paper to shed light on this question. My focus will be limited to two groups: scholars writing on the movement in Portuguese and scholars writing on the movement in English. More specifically, I will be interested to know what ideas researchers from these two groups are pursuing and whether the language-specific content has any distinguishing characteristics.

To provide structure for my project, I offer two hypotheses. First, I believe that education will be a central theme, both in Portuguese and in English. It is a pressing concern in Brazil, particularly in the interior. It seems reasonable to suspect that this theme will capture much academic interest from all over the world. Second, I hypothesize that English publications will have a tendency to focus on macro-level aspects of the movement and that Portuguese publications will be more qualitative and contextualized. By macro-level, I mean that the publications will be focused on broader issues framed in a more abstract, impersonal manner.

This project is limited in many regards. It relies on the efforts of an inexperienced scholar using new methods, some of which are very complex. Many of the computing operations, for example, sort through hundreds of thousands of words. My inferences and conclusions, then, should be considered with this in mind. Secondly, this project was limited by space. I explored countless computing functions and searches to produce the empirical base of this

project, only a part of which could be reasonably included in such a short paper. When most important, I provided figures and tables. When data was less important, I described the results briefly in the text.

## 2. DATA/CORPORA

Academic journals were utilized for information retrieval, the principle space for scholarly production. MST journals needed to be sampled, given the limitations of this project. My sampling method was atypical and not scientifically robust per se; however, I believe it was a reasonable and effective approach. I entered the query "MST Brazil" into google scholar and then collected all articles that had MST in the title, abstract or first page. I stopped the search after 40 browser pages. This query was done both in English and in Portuguese to obtain two language-specific samples. Given my sampling method, then, this research is based on electronic journals and is subject to the search algorithms employed by google scholar. Any serious academic pursuit would need to compensate for biases that may have been introduced from this sampling method.

For this project, I also sub-sampled the English corpus to obtain ten articles that would be more easily managed by the POS tagging operation. The ten articles were selected at random using an online random integer function.

Collection of electronic journal articles from the internet introduced a series of problems, some of which fundamentally altered the direction of this project. While I was able to convert the articles to text files, I was unable to retain Portuguese symbols when compiling the texts into one readable document. Cleaning was then necessary to get rid of the myriad of strange characters imputed. I cleaned with both alpha and lower-case standardization and I removed stopwords for all non-POS tagging operations. In this process, content of interest was undoubtedly removed, but I am reasonably satisfied with the amount and quality of information that was retained.

Summary statistics on the corpora can be found in Table 1. There are also summary tables and figures to describe the texts by publications per annum, by journal classification and, for English articles, by author's academic department.

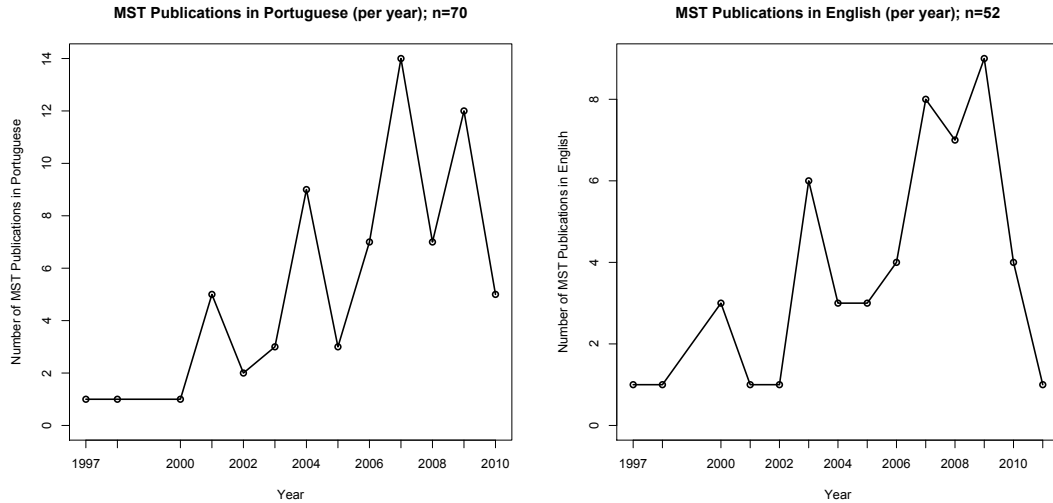
TABLE 1. Summary Statistics of the Corpora

Corpus	Raw Words	Cleaned Words	Unique Words	Cleaned Uniques	Texts
Portuguese	1,079,548	418,722	52,774	32,611	70
English	575,207	259,116	41,358	23,702	52
English Sub-sample	136,307	102,500	16,423	10,456	10

## 3. METHODS

The methods I employ in this project can be divided into four parts: exploratory searches, defined searches, tagging and chunking, and dimension reduction. For the exploratory part of this project, I used Python software to search for the words with highest

FIGURE 2. Sampled Academic Publications on MST, by Language



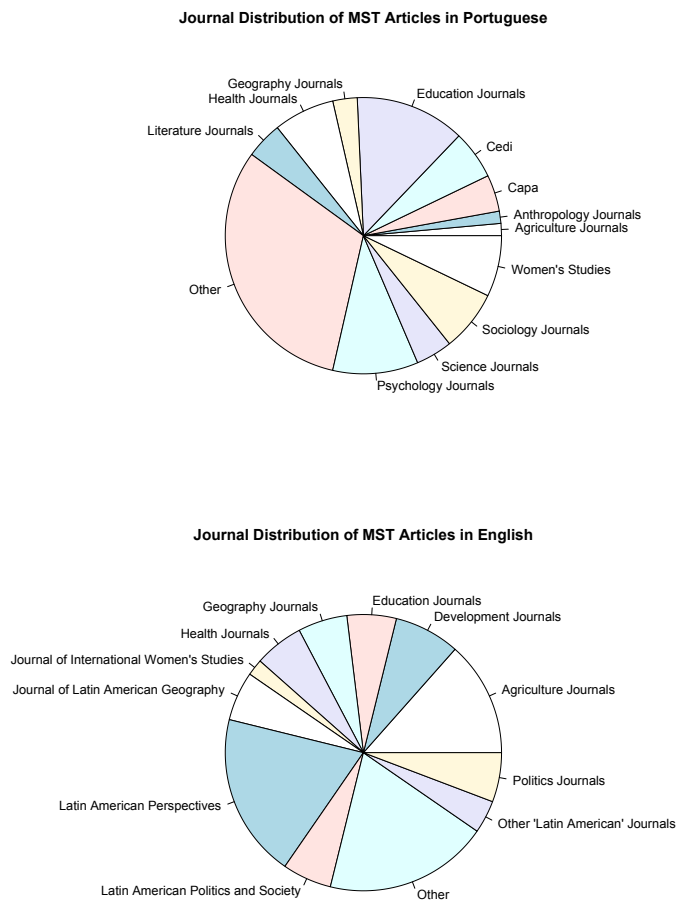
relative frequencies. I then searched for words that were both long and frequent in the texts. I also looked for words that were "unusual" in the English sample relative to the entire nltk corpus. Finally, I used collocations to find distinguishing words and phrases. These methods were all attempts to identify pertinent statements that could be subjected to closer examination in later methods.

For the second part I used Python to conduct defined searches. I started with counts of words I was interested in, based on my previous knowledge of the movement and simple curiosity. I then conducted counts on words identified in the exploratory section. Next, concordance searches were used to investigate salient items to get a better understanding of contextual roles specific words played in the corpora. I also used similarity searches to check for synonyms originally unaccounted for. For example, the Portuguese texts wrote commonly about "the fight," however, "fight" had an inferior role in the English texts. On account of similarity searches, "fight," right or wrong, is more commonly described as a "struggle" in the English texts. Next, common context searches were employed to find what themes were operating around multiple pre-defined ideas of interest. I then used bi-grams to search tokens occurring around words of interest. These themes were then tabulated to give conditional frequency distributions. Although this part focused on pre-defined ideas, it also took on an exploratory role as new ideas came to the surface during the process.

Part three consisted of information extraction from a sub-sample of ten English articles on the MST. Sub-sampling was necessary, given the speed of my computer and the time it takes to POS tag. The sub-sample was selected at random from the English Corpus, and at about  $n/5$ , the sub-sample is representative of my English corpus as a whole.

The goal of information extraction was to get at the grammatical substructure operating in and around key words and phrases. This began with POS tagging the sub-sample and

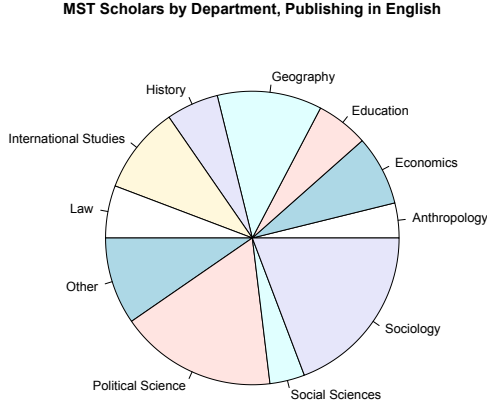
FIGURE 3. Journal Distribution of MST Publications



searching for how specified POS were operating in the texts. Next, I selected on POS types to find what specific words were occurring at high frequency per type. I was most interested in the work done by third-person present tense verbs, adjectives, comparative adjectives, and superlative adjectives.

Tags were then used to chunk and dependency parse. Chunk parsing allowed specific searches of noun phrases. More grammatical rules would have resulted in a more useful search, however, the grammar coding complicated this endeavor. I then proceeded to dependency parse the sub-sample in order to find how grammar items were operating around my key words of interest.

FIGURE 4. MST Scholars, by Department



The final methodological part of this project consisted of dimension reduction. For this end I used Python to vectorize my corpora and R to plot and investigate dendrograms and clusters. I specifically investigated how the texts would distribute using cosine distance and, based on those groupings, what words or phrases united them.

#### 4. RESULTS

**4.1. Summaries.** Based on the statistical and descriptive characteristics of the metadata, we see publications tend to be increasing by year in both the Portuguese and English corpora (Figure 2). There is some variability in per annum publication, but the only reason to reject this trend is the suspicion that earlier publications were biased by the sampling criteria of electronic form. Indeed this is of concern, but given that this criticism is only tangential to the argument as a whole, it was ignored. The drop in recent publications, that is, in 2009 and 2010, is likely due to publication delays.

Descriptive characteristics also reveal that Portuguese articles have a tendency to be printed in education and psychology journals (Figure 3). Similarly, women's studies and sociology journals are publishing relatively more MST articles in Portuguese than in English. Comparing this distribution with the English distribution, we see that there is already some indication that my first hypothesis was poorly formulated. Portuguese education journals are more actively engaged with the MST than their counterparts in English, suggesting that education is more important in Portuguese print than in English print. Finally, Portuguese journals seem to be more diversified than English journals. This point is not entirely clear based on the pie summaries, but when compiling the "other" category, there were more articles merged into "other" than in the English case.

The English corpus has a greater concentration in agricultural, political and Latin American perspectives journals. In fact, Latin American perspectives of some form occupy roughly a third of all journals in the English corpus. We also find that women's studies and sociology journals play a small role and no role, respectively, in English journals, while they were more active in the Portuguese case.

Finally, I had an opportunity to classify scholars publishing in English by department employment (Figure 4). This revealed that political scientists have taken a strong interest in the movement and also, ironically, sociologists. This last department is a bit surprising, given that there was not a single MST article published in an actual sociology journal. This phenomenon merits further consideration, perhaps a project I will undertake in the future. Lastly, one notes that the MST is being investigated by scholars from a myriad of diversified departments, evidence that the complexities of the movement are capable of galvanizing cross-disciplinary interest.

**4.2. Exploratory Searches.** Relative word frequencies are found in Table 2. Unsurprisingly, MST is the top hit in both corpora. However, one also sees space for differentiation between language prints. For example, the Portuguese corpus records high frequencies of education, formation, and fight. Distinguishing words from the English corpus include: political, reform and state.

Table 3 and Table 4 record long and frequent words in Portuguese and English, respectively. Again, the Portuguese texts are connoting community, participate, society; identity, transforms; and pedagogy and teachers (education was likely left off this list because its stem is short). The English texts in Table four are crystallizing around some similar themes, and yet, around a different set of categories as well. Of notable interest are agricultural, political, production, government and organization. These themes suggest the macro analysis I had hypothesized in the outset. They connote political and economical characteristics.

For comparative purposes, collocations and unusual words are of less utility in this project. The problems of processing Portuguese texts rendered collocations unfruitful, as basic words like agrarian were divided into "agr" and "ria" (Figure 5). This is analogous to dividing the word "agrarian" into "agr" and "arian" in English. Such a division is not the intent of collocations. Similarly, I had no grand Portuguese corpus against which to compare my Portuguese MST corpus for unusual words. These searches were fruitful, on the other hand, for the English corpus (Figure 5). Common collocations were: agrarian reform, Latin American, and land reform. Unusual words were notably fruitful in the English case with returns including mango and banana. However, sadly, unusual words proved of little substantial value in this analysis.

FIGURE 5. Collocations

<pre>In [60]: NLTKText.collocations() Building collocations list reforma agr; agr ria; pol tica; fam lias; movimentos sociais; rio janeiro; sem terra; per odo; hist ria; trabalhadores rurais; pol ticas; ind genas; poss vel; dispon vel; experi ncias; pol tico; bras lia; porto alegre; import ncia; vel http</pre>	<pre>In [40]: NLTKText.collocations() Building collocations list agrarian reform; latin american; social movements; sem terra; downloaded univ; latin america; chicago march; land reform; dos trabalhadores; rural workers; agr ria; reforma agr; american perspectives; rio grande; university press; grande sul; per cent; new york; movimento dos; trabalhadores rurais</pre>
---	---

TABLE 2. Relative Frequencies of Words

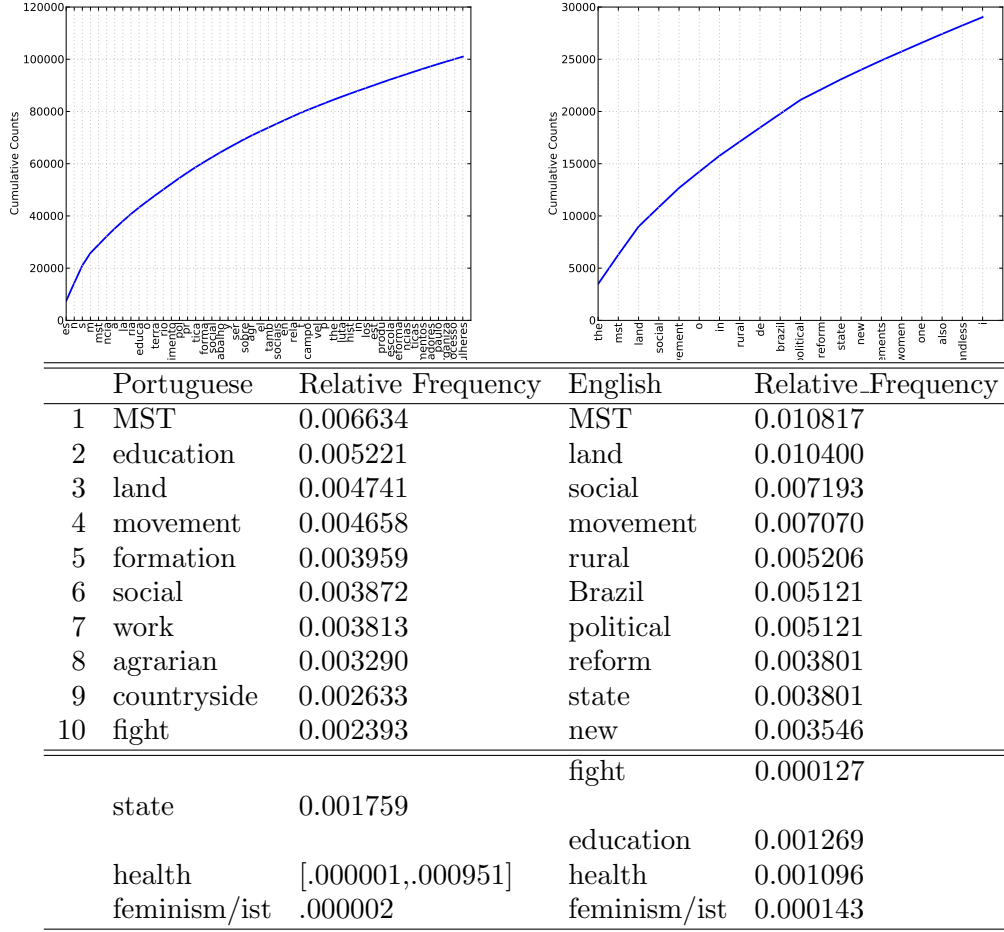


TABLE 3. Long and Frequent Words in Portuguese MST Articles

encampment	settled	settlement	activities	Brazilian
communities	development	different	identity	important
movement	participate	pedagogy	teachers	reality
represents	society	workers	transforms	university

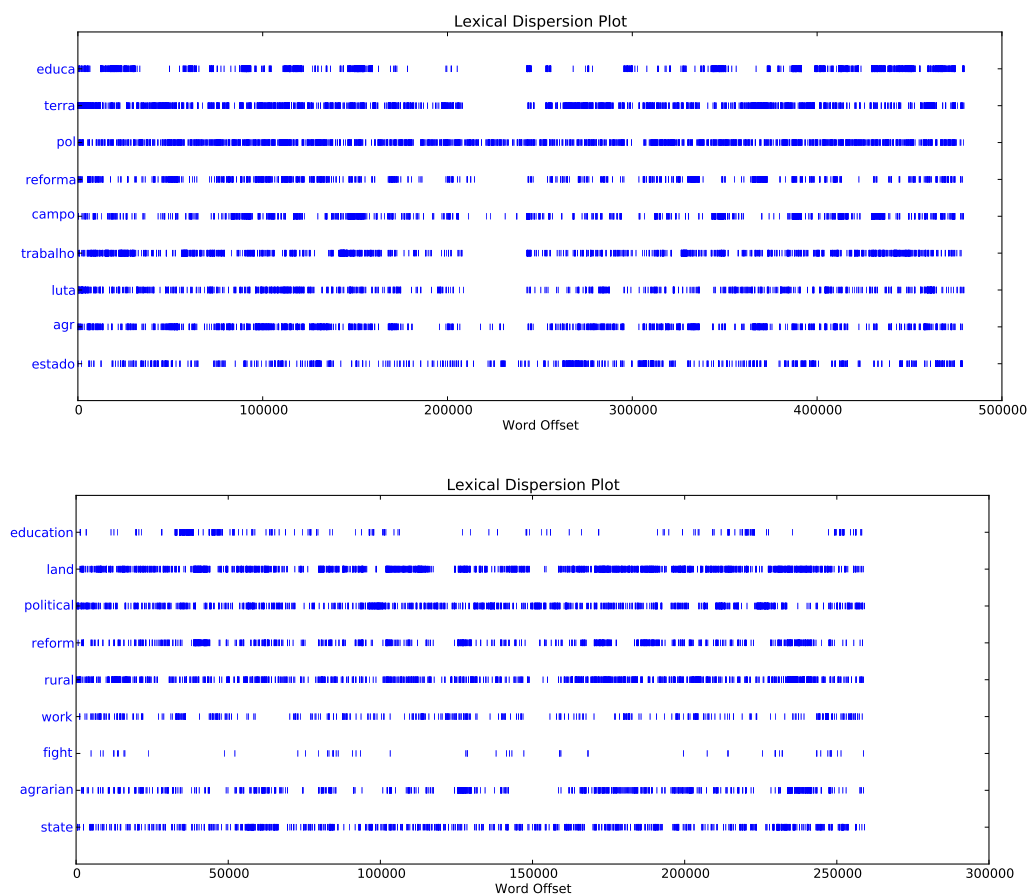
Based on the exploratory portion of this project, there are now a few themes of peaked interest. Grounds have been established for serious interest in ideas associated with education, fight, state, access, and health. However, these are only a few of the recurrent,



TABLE 4. Long and Frequent Words in English MST Articles

agricultural	brazilian	collective	community	development
different	education	government	important	international
movements	occupations	organization	participation	peasants
political	production	settlement	settlements	university

FIGURE 6. Lexical Dispersion Plots; Portuguese(above), English(below)



meaningful themes. Rather than compile a long list of words, these concepts of serious interest are identifiable as the specific words I include in the upcoming defined searches.

**4.3. Defined Searches.** I did specific word counts on a number of words, displayed as (word, counts in Portuguese texts, counts in English texts): (access, 299,319), (rights, 206,

380) and (capitalism, 101, 71). These counts are not scaled and so cannot be compared between corpora in this raw form. However, of greater interest are the following defined searches.

Concordance searches were performed on four central ideas: education, fight/struggle, social, state (Figures 7, 8, 9 and 10). These choices were made due to the fact that these concepts seem to play distinctive roles in the two corpora. In the end, concordance on fight/struggle and social did not seem to offer any notable differences, however, education and state did. In the Portuguese corpus, education is taking on a much more social role. The sentences focus on the education of infants and children, pedagogical norms and concerns, socializing people, and improving the "human condition." The English corpus, on the other hand, focuses more on "continuing education", leader education, political education and educational "data." In other words, the English corpus seems to view education in a more abstract way, focusing more on the policy of adults and not the experiences of children and young people. This finding in the English corpus calls attention to my hypotheses: education is framed conceptually in a macro-perspective. These journals concentrate on education's relationship within economics and policy. My second hypothesis relating to the Portuguese corpus continues to be retainable. These scholars are more "qualitative and contextualized," seeking to understand the pressing concerns of the Brazilian educational experience, rather than macro-perspectives focusing on government policy. This is not to say that policy is not a theme in Brazilian texts, but that the relative role of policy is less than it is in the English corpus.

FIGURE 7. Concordance of 'education' in MST Articles

os salete caldart produzindo cultura educa infantil campo movimento massivo alf  
envolvem quest es relacionadas produ educa sa cultura direitos ampliam medida a  
alhadores trabalhadoras terra tamb m educa assumir bandeiras luta cada vez o pr  
exclus social trabalho lutas sociais educa am rica um desafios coloca exatament  
pr ticas educativas matrizes g gicas educa assuma parte dilemas sociais deste f  
sos transforma social sobre pr ticas educa os mentos sociais t m sido espa orga  
ica permite enxerg desde preocupa es educa pedagogia olhar gostaria desenvolver  
tru coletividade sem terra processo educa tamb m modo produ forma humana tanto  
s acampamentos assentamentos marchas educa crian jovens adultos jeito ser colet  
tamos tempos desumanidade crescente educa somente sentido pr tica radical huma  
gem pr pria reflex pedag gica reflex educa forma humana humanizamos formamos hu  
na humanizamos formamos humanos como educa pessoa desenvolva condi humana quais  
well educated committed continuing education thus securing learning teaching sh  
well highly over half mst leaders education either technical school university  
tion mst invests large part budget education developed ties number universities  
ciousness also reflected political education programming mst leadership trainin  
cial base movement achieved higher education norm directly involved practical s  
members beyond general data income education level family a vignettes photograp  
erned promoting programs legal aid education health care burial funds extensive  
cooperative consequence low level education among it consequence structural fo  
period months undergoing political education ration effort once occupation deci  
dated often the movement committed education children adults created schools ma  
schools many camps its elementary education programs recognized unicef caldart  
ds important plishments production education even mst sfrequent appearances med

The other concordance item of interest is "state." In the Portuguese text, state is more commonly referring to "the state of." For example, authors write about the "state of illiteracy", the financial state, or simply use it as the past participle of "to be." In the English texts, "state" is more commonly referring to "state power" or some activity that the government is doing. This suggests that the state is conceptualized as an actor in the English texts and that English-print authors are more interested in the MST's relationship with the government than the MST's relationship with society or community.

FIGURE 8. Concordance of 'Fight/Struggle' in MST Articles

junto fam lia sem terra s conquistas luta coletiva muitas pessoas tamb m perde  
 ileira quest reforma agr ria fazendo luta terra afirmando iniciativas possibil  
 er sticas a radicalidade jeito fazer luta sujeitos o mst reafirmou ocupa latif  
 mou ocupa latif ndio principal forma luta terra mobiliza massa jeito faz isto  
 sto quer dizer explicita pr prias es luta contesta enquanto pr tica enquanto v  
 ginava ter exclu talvez radicalidade luta jeito faz provoca sociedade tomada p  
 tes a multiplicidade dimens es o mst luta terra eixo central caracter stico pr  
 olhas fez mente sobre jeito conduzir luta espec fica luta feita fam lias intei  
 sobre jeito conduzir luta espec fica luta feita fam lias inteiras acabaram lev  
 combina objetivos diversas dimens es luta mst acabou construindo tipo organiza  
 nte pilares identidade todo bandeira luta nasce grupo social espec fico intere  
 esses sociais o lema reforma agr ria luta todos trabalhado mst especialmente p  
 ar sociedade reconhe reforma agr ria luta n apenas trabalhadores trabalhadoras  
 sw movements dominate contemporary struggle social change latin america precis  
 wer chile via unionization labour struggle organized working class capital st  
 volution via guerrilla form armed struggle state each modality political stru  
 ggle state each modality political struggle helped incorporate elements workin  
 rer a periodization analysis waves struggle bring change implement emancipator  
 ok centre stage dominated popular struggle social antisystemic change current  
 rating nature social base dynamics struggle involved brass bernstein petras th  
 d landless families fifteen years struggle height struggle land land reform m  
 lies fifteen years struggle height struggle land land reform mst notwithstanding  
 izations term forms peaceful civil struggle including use electoral mechanism  
 rt explain divorce many ments mass struggle electoral in contrast leaders mst  
 representations movement historic struggle forces reaction banners insignias

FIGURE 9. Concordance of 'social' in MST Articles

cas distinguem trajet ria movimento social trabalhadores trabalhadoras uma tra  
 sujeitos pretende trazer volta cena social pa a ep grafe escolhi texto diz si  
 a forma o dos sem terra o movimento social como princ pio educativo acombina o  
 iza mistura versatilidade movimento social entra todo mundo tempo todo xadrez  
 nizacionais pr prias quase institui social pretende flex vel da l gica verdade  
 ex vel da l gica verdadeira empresa social convivendo irrever ncia movimento i  
 riando denomina si pr prio organiza social massas combina caracter sticas cont  
 dade todo bandeira luta nasce grupo social espec fico interesses sociais o lem  
 min rio proposta debate rela exclus social trabalho lutas sociais educa m ric  
 nam algo sobre processos transforma social sobre pr ticas educa os mentos soci  
 tica pedag gica pensarmos movimento social tamb m matrizes pedag gicas fundame  
 ticle met i accent sur ie caractkre social ainsi que sur la dynamique sociale  
 teurs considrons comme le mouvement social le plus dynamique de nos jours en a  
 ip du mouvement this paper focusses social character dynamics political leader  
 dership widely regarded as the class social character leadership movement other  
 ship movement others regard dynamic social movement latin america rural landle  
 ess workers movement mst discussion social dynamics made reference ten hypothe  
 addressed number questions related social base movements issues around collec  
 ssing studies however analysis role social dynamics political leadership ous o  
 ven saliency factor earlier studies social actor approach taken many social mo  
 es social actor approach taken many social movements analysts today calderh es  
 poststructuralist we discuss class social character leadership movement dynam  
 aracter leadership movement dynamic social movement latin america today this s

FIGURE 10. Concordance of 'state' in MST Articles

sem terra realizou janeiro cascavel estado paran hoje mst est organizado estad  
 imposs vel mudar quanto mada atual estado coisas humana pessoa exatamente con  
 s n restou outra sa n ser solicitar estado financiamento produ atrav s cria li  
 if cil porque depende correla mbito estado analfabetismo st dile com efeito ms  
 lema n s persiste sido agravado ado estado pol tica econ mica orienta a ado de  
 is industrial financeiro associados estado n obstante prisioneiro terra mst re  
 t atravancada crise agravada d vida estado medidas tomadas s rios reflexos sob  
 produ haja vista greves canavieiros estado s paulo anos eram greves oper rias  
 av s luta dando in cio ocupa terras estado rio grande sul a funda mst mbito na  
 ias acampadas aguardam provid ncias estado serem assentadas a luta legaliza re  
 direito educa fazendo cumprir dever estado oferecer educa todos cidad mst atua  
 pments settlements rio crande de sbo state survey conducted thia movement leade  
 national leaders representing every state mst operates conducted seminar may s  
 vador allende bringing working class state power chile via unionization labour  
 ggle organized working class capital state higher wages improved working tions  
 on via guerrilla form armed struggle state each modality political struggle hel  
 ted cadre leaders representing every state mst these hypotheses derived general  
 consensus ers common enemies nature state power impact world bank internationa  
 el consensus regarding saries nature state power perceives cardoso regime comple  
 ed books slater new social movements state latin amsterdam cedla stedile joao i  
 s vilaa earthquakes volcanoes market state revolution new york monthly review p  
 decreases exacerbated changes nature state support brazilian government began d  
 en sugarcane distilleries pernambuco state responsible majority total sugarcane  
 hut verge bankruptcy lins production state fallen twenty million tons million d

The results of the bi-grams searches for conditional frequency distributions are found in Table 5. These offer greater support to the notion that education is playing a more

important role in the Portuguese corpus. Regardless of the key word selected, education is by far the most likely concept to appear. In fact, education or some educational theme occurs on seven different occasions. Alternatively, education occurs only a single time when calling on the same key words in English.

TABLE 5. Conditional Frequency Distribution of Words in Portuguese MST Corpus

Land	work - 45	Pernambuco - 28	reform - 27	encampments - 22	free -
Rural	education - 22	city - 18	Brazil - 10	social -9	agenda
Work	education - 44	cooperative - 43	rural - 38	pedagogy - 30	collective -
Fight	land - 227	reform - 74	social -48	against - 42	MST -
State	Brazilian - 34	nutritional-22	Pernambuco -19	movements - 8	education
Education	countryside - 291	MST - 93	rural - 85	popular - 81	young people -
Feminism	very - 2	this - 2	women - 2	years - 1	march
Health	public - 3	collective - 1	education - 1		
Access	land - 42	service - 11	education - 10	goods - 8	school

TABLE 6. Conditional Frequency Distribution of Words in English MST Corpus

Land	reform - 387	occupations - 190	rights - 100	redistribution - 41	ownership - 27
Political	parties - 49	economy - 32	social - 29	system - 26	ecology - 25
Rural	workers - 227	women - 94	social - 71	labor - 47	areas - 42
Work	land - 11	MST - 9	together - 6	carried - 4	done - 4
Fight	land - 13	agrarian - 2	capitalism - 2	legal - 2	access - 1
State	university - 27	power - 22	level - 14	national - 12	repression - 12
Education	program - 47	social - 12	health - 10	children - 6	citizenship - 5
Feminism	left - 2	women - 2	class - 1	consciousness - 2	movement - 1
Health	care - 48	services - 25	movement - 10	system - 10	education - 8
Access	land - 78	health - 14	credit - 11	care - 10	public - 7

Instead, we see the macro-social perspective operating once again in the English conditional frequency table. These items include: reform, rights, system, national, power, and capitalism. While some social themes are connoted, the same detached, policy-level, macro-perspective seems to dominate the table.

**4.4. Information Extraction on Sub-sample of English Corpus.** My queries on tag distribution served mainly to get a feel for the work grammatical items were doing in the English sub-sample. I was most interested in what verbs were operating in third person and adjectives functioning as pure adjectives, comparative adjectives and superlatives adjectives. I used a "findtags" function to return the twenty-five most commonly occurring

items in each class. The return on verbs, comparatives and superlatives was of little value, however, the pure adjectives gave a more clear understanding of how nouns were being consistently modified. Some of the most frequent adjective returns were political, economic, public, and national. Given that these queries referred exclusively to the English texts, I interpret this as more evidence that scholars writing in English are focusing their efforts on macro-phenomena. In further support of the smaller role education is playing in the English corpus, there was not a single education-related return.

POS tagging, however, was more of a method to get to chunk and dependency parsing the text. The chunk parsing commands I issued returned noun phrases, of which there were many. In order to conduct a more useful analysis, more commands would be needed. Nonetheless, this parse seemed to corroborate the macro-perspective approach developed in the course of this paper. For the dependency parse, I focused on the adjectives modifying three words of interest: education, schools, and children. These words were chosen because higher frequencies of impersonal returns would substantiate an impersonal, macro-perspective approach operating in English print. I reasoned that the range of description would provide solid clues as to whether or not this impression was legitimate. In the end, this impression was weakly supported by the data. Education was "effective," "national," and "political," not "elementary," or "important." Children are so infrequent in the English journals that there were only 6 adjectives to describe them. Finally, schools are "Brazilian" and "democratic," with no description to alter my suspicion that English scholars, indeed, are dealing with these themes on a macro-scale, with less contextualized accounts of individuals' experience with the institution.

**4.5. Dimension Reduction.** Cluster analysis revealed interesting results in regards to how the articles clustered in the Portuguese and English MST texts. Table 8 shows a dendrogram of English MST texts (legend in Table 7) when clustering in accordance with hclust's "complete" method. We find that three categories clustered well based on cosine distance: education, health, and Latin American perspectives. I then "ten cut" to see how the clusters would distribute. Table 9 shows the results. Clusters 1, 2 and 4 captured the greatest number of articles.

Upon closer investigation, cluster 4 contains the Latin American perspectives. Other words of interest include politics, party and Incra (a Brazilian government land bureau). A closer look reveals how this cluster got so big. The dendrogram shows that cluster 4 slurped up all 16 articles ranging from 2003.dev1 to 2003.laps. Clusters 1 and 2 are a bit more diversified. I was still interested to find where the education articles had collected. It turns out that they piled into cluster number 5, around the concepts of teacher, education, and school. One notes from Table 7 that there were only 3 articles published in English educational journals. We can infer that the content of these journals, at least for these three articles, is very tightly related and distinct from the rest of the sample.

Table 11 displays the dendrogram for Portuguese texts (Legend in Table 10). Judging by the way the texts have coalesced, one would have to say that the content of respective Portuguese journal publications is much more consistent than in the case of English MST articles. As a reminder, we only saw 3 distinguishable journal pockets in English print.

FIGURE 11. Dissimilarity Distributions in Corpora

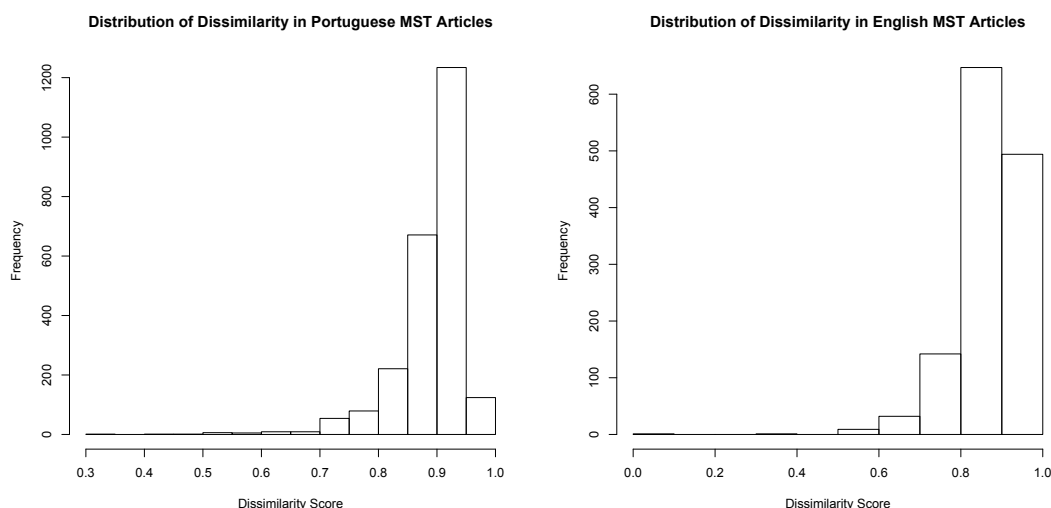


TABLE 7. English Journal Abbreviations for Cluster Dendrogram

Journal Name	Abbreviation	# Articles
Agriculture Journals	agr	7
Development Journals	dev	4
Education Journals	edu	3
Geography Journals	geo	3
Health Journals	hea	3
Women's Studies	fem	1
"Journal of Latin American Geography"	jlag	1
"Latin American Perspectives"	lap	10
"Latin American Politics and Society"	laps	3
Other	oth	10
Other "Latin American" Journals	olj	2
Political Journals	pol	3

"quotations" indicates title of a specific journal

The Portuguese texts seem to cluster closer to 6 (fem, hea, edu, ced, psy, oth, and soc). In similar fashion, I "ten cut" the Portuguese dendrogram. The distribution is in Table 12. Megacluster 1 is an eclectic titled group of journals focusing on school and pedagogy. It is curious then, that the education journals at the midpoint did not get incorporated into this cluster as well. Instead, they cluster closely with the Journal "Cedes." With a little

TABLE 8. Cluster Dendrogram for English MST Articles

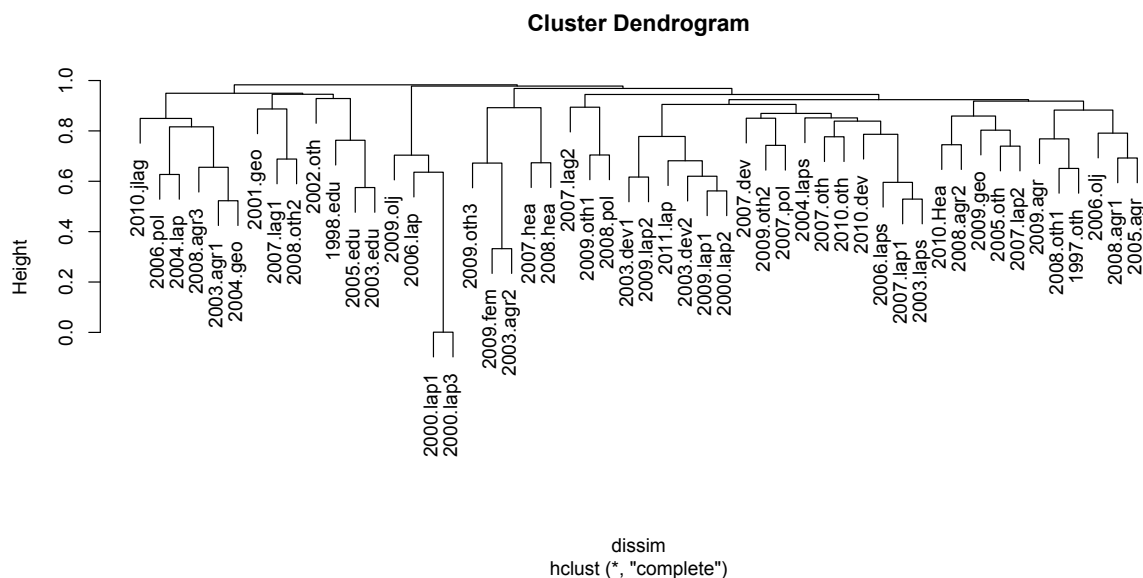


TABLE 9. Distribution of Dendrogram cut into 10 Clusters

Cluster Number	1	2	3	4	5	6	7	8	9	10
Number of Articles	6	6	3	16	3	5	5	4	3	1

research, it turns out that Cedes is a popular education journal and that if you "eight cut" the dendrogram, all of the education journals coalesce into one giant of 27 articles.

## 5. DISCUSSION AND CONCLUSION

At the outset of this paper I stated two specific hypotheses. First, I claimed that education was a pressing issue in Brazil and that both journals, in Portuguese and in English, would expend a great deal of effort to understand this issue. The data I collected refutes this hypothesis for the most part. Education, generally, seems to be much more important in Portuguese print than in English. The cluster analysis shows that Portuguese-language scholars are publishing on it in a variety of disciplines and that the theme is widespread, as evident from defined searches. When education is studied in English language journals, there is evidence that the research is concentrated on macro-perspective approaches; approaches focusing on education in relationship to policy and the economy.

This point relates more closely to the second hypothesis I offered. I stated that English publications would have a tendency to focus more attention on macro-political aspects of

TABLE 10. Portuguese Journal Abbreviations for Cluster Dendrogram

Journal Name	Abbreviation	# Articles
Agriculture Journals	agr	1
Anthropology Journals	anth	1
"Capa"	capa	3
"Cedes"	ced	4
Education Journals	edu	9
Geography Journals	geo	2
Health Journals	hea	5
Literature Journals	lit	3
Other	oth	22
Psychology Journals	psy	7
Science Journals	sci	3
Sociology Journals	soc	5
Women's Studies	fem	5

"quotations" indicates title of a specific journal

the movement and that Portuguese publications would be more qualitative and contextualized. This hypothesis seems to be supported based on my empirics. English publications have a concentration on macro-systems and an abstraction likely due to their isolation from the social milieu in which the MST is operating. Portuguese publications seem more in touch with the micro-level characteristics of the movement. They seem to be interested in how the MST is running schools and the nuances of this experience. It seems qualitative, personal and subjective, probably because the scholars who write in Portuguese live with, around or near the movement.

In conclusion, the tools of content analysis have indeed shed light on the subject of MST publications. In this paper two hypotheses were raised and responded to, although, the virtues of content analysis truly seem to lie in the qualitative and exploratory continuum. Scholars compile such vast quantities of information, yet, despite our sophisticated forms of organization, we have difficulty knowing what information is specifically out there and how exactly to retrieve it. The tools of this method seem to offer new ways of recovering reported conclusions and synthesizing seemingly disparate works to make insightful inferences.

During the course of this research several themes were discovered that merit further scholarly attention. First, it would be interesting to know how the discipline of Psychology is dealing with the MST in Portuguese print. Psychology journals are prolific in regards to the MST and cluster analysis shows that their focus is well honed. It would be interesting to discover what themes are operating in this field. Next, I am curious to know why so many sociologists write on the MST, but none that I know of actually publish MST articles in sociology journals. This is a curious finding. Finally, I think that network analysis could shed some light on how these articles and publishers are related. It would be interesting to



TABLE 11. Cluster Dendrogram for Portuguese MST Articles

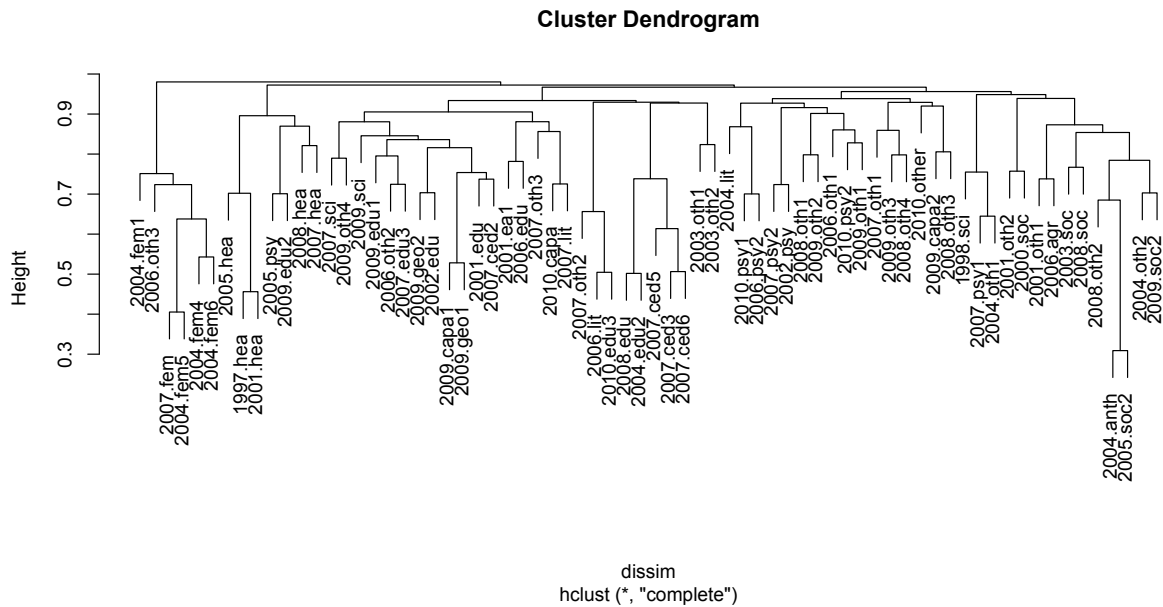


TABLE 12. Distribution of Dendrogram cut into 10 Clusters

Cluster Number	1	2	3	4	5	6	7	8	9	10
Number of Articles	17	10	9	6	3	6	7	7	3	2

see if these scholars are working in collaboration and what the web of academic relationships looks like.

FIGURE 12. 'Pam' Cluster Diagrams of English Articles

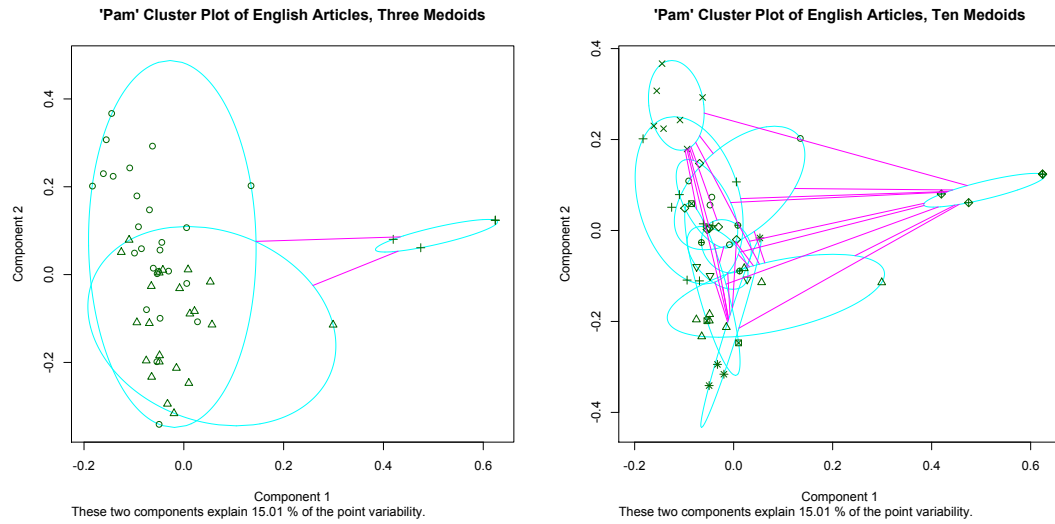


FIGURE 13. 'Pam' Cluster Diagram of Portuguese Articles

