Business Insight Report

Estelle Eteki – MSBA 3

Cameroon is a country in Central Africa with about 24 million habitants. The country is divided in ten regions; the north west and the south west compose the anglophone area. The latter is covered by about 5 million habitants. In October 2016, lawyers went on strike, and organized a peaceful march to protest against the failure of the justice system to use Common Law and the translation of the code OHADA (Code of the Organization for the Harmonization of Business Law in Africa) in English. The government's repressive approach towards them lead to a series of strikes from different sectors. The region lived through about three months internet shutdown, general strike and strong fight between police forces and population. Three years later, the region is still politically instable; secessionists groups have emerged, and the government has not been able to bring back peace in the country. The articles, used to perform text mining analysis and provide insights, are reports from International Crisis Group in 2017 and 2019. It helps us to dive through the crisis from the beginning till end of 2019 in terms of causes and what measures have been put in place so far.

Methodology: Our analysis has been inspired by multiple frameworks of data mining on R based on articles from the International Crisis Group. To elaborate more on our paper, complimentary information about specific events happening within the country is provided by external research. The external research was based only on the words provided by the analysis.

The frameworks used to elaborate this paper, will help to dive through the chronology, from the beginning until the end of the crisis. We will be able to understand what the roots are, the actions of the government to bring back peace and the population's reactions by 2019. The first framework used is the frequency histogram. The appendix A displays graphs representing the frequency of words used in both articles. The most frequent words used across both articles are crisis, anglophone, government. We easily understand that there is anglophone crisis going on in Cameroon and the government is trying to fix issues. Moreover, anglophones also complain about government's management towards English speaking people. The latter feel marginalized as their language. Furthermore, Buea is the epicenter of the crisis since it is the city where the march protest took place back in 2016. Since the crisis started, the anglophone population asked for an effective decentralization to balance power across the country. The separatist movement started a couple of months, it became a conflict between secessionist and armed forces; and population started moving. In September 2019, under the instructions of the president head of the state, a major national dialogue took place from September 30 to October 4 between the government and some opposition parties to discuss about different recommendations to completely end the crisis. Digging down from one article to another, we have a deeper understanding of the chronology and how the situation evolves in two years.

Secondly, appendix A presents another graph showing the histogram across each article independently. The icg_2017 emphasizes more the roots of the problem. Back in 2016, marches started in Bamenda and Buea, and were spread out throughout the northwest and the southwest regions. Teachers went on strike in Bamenda. To be able to write a report accurately, The International Crisis group conducted interviews with anglophone leaders to deeply understand their revendications. The francophone considered mostly the crisis pointless because they believe that the revendications were unfounded. On the other hand, the icg_2019 shows the situation of the crisis at the end of 2019. It intensified and there is strong fight between separatists and armed forces, and there is no more security in the region. The president ordered a national dialogue at the end of the year to find sustainable solutions with opposition parties to the crisis.

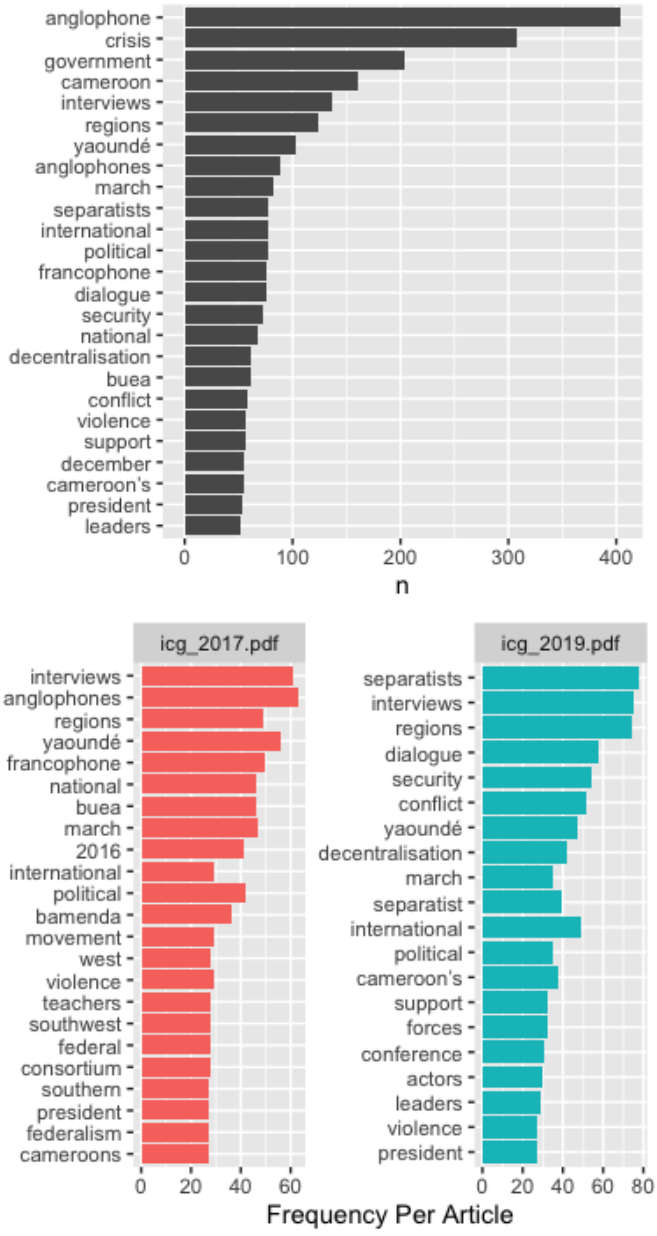
The tf-idf, standing for term frequency-inverse document, is a framework which help to understand which better represents an article even though it is not the most used one. Computing this framework on our article help us to dig deeper in the situation of the country in those two years. The appendix B is a graph resulting from the computation of this framework to both articles. Starting by icg_2017, some words with high tf-idf are Ahidjo, unification, cpdm and reunification. Those words brought us back to 1961. In a referendum held in February 1961, Southern Cameroons chose to join Cameroon while Northern Cameroon chose to join Nigeria (the Francophone territory was independent since January 1960). The political leaders who led to reunification that influenced this choice were Muna and Jua. After reunification on October 1, 1961, Ahidjo was the federal president and Foncha vice president of the country. In October 2016, lawyers went on strike in 2016. After the repression of the police towards their peaceful march, many sectors followed the movement as students at Buea university on November. To ease people's mind, CPDM (political party of the president) tried to organize a rally in Bamenda. However, it just leads to severe confrontations and four people's death. Moreover, oil is one of the great sectors in the economy of the country. Anglophones were also complaining about low percentage of their peers in the industry.
On the other hand, words with high tf-idf in icg_2019 are separatists, refugees, Abuja and idps (referring to internally displaced people). The crisis became more intensified throughout the time, and the movement of separatists rose. Some federalists joined separatists. It affected the education system by the closure of schools and the threatening of population. Many children have not been to school for at least one year. The conflict leads to approximately 530,000 IDPs and 35,000 refugees in Nigeria (mostly women and children). Despite the national dialogue organized by the president in November 2019, the crisis in northwest and southwest region is still live and far from ending.

To conclude this paper, we compute a sentiment analysis on the 2019 article (appendix C). The sentiments perceived by the author are both highly negative and positive. On the negative side, we have sentiments as repression, disapproval, poor, killing. It summarizes the situation in those regions where people are moving because they are threatened or don't feel secured anymore. They have been experiencing kill on a daily basis. The economy of the region is suffering, and poverty is all over the region. Moreover, federalists or even separatists are frustrated because their complaints are ignored. On another hand, the positive sentiments are liberation, faith, encouraging and peaceful. There is hope and faith in population that peace is going to come back in the region. The government has finally decided to take actions towards the crisis and has a way to show it off, the president ordered the liberation of 289 out of about 1,000 detained Anglophone activists. Moreover, a disarmament and demobilization for separatists has been created in 2019.
Diving deeper throughout the chronology of the anglophone crisis, it comes out that anglophone population feel marginalized asked for decentralization and federalism to balance power across francophones and anglophones. However, government actions have not satisfied population striking and the separatists rose on the region. We firmly recommend to the Cameroonian government to take actions on decentralization as it's one on the revendications they claim both on 2017 and 2019. Legislatives and municipal elections took place in the country (and the concerned region) on February 9, 2020. May it be a real sign of peace coming back in the region.

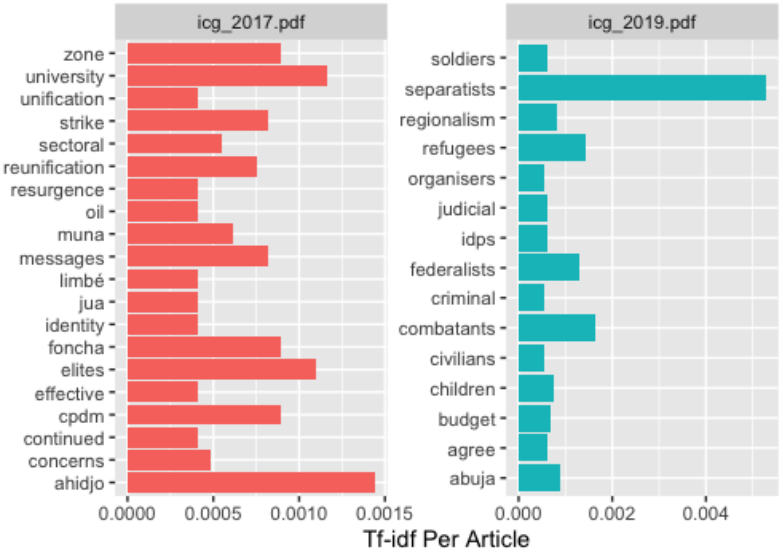
References

International Crisis Group (August 2, 2017). *Cameroon's Anglophone Crisis at the Crossroads*.
International Crisis Group (May 2, 2019). *Cameroon's Anglophone Crisis: How to Get to Talks?*

Appendix A Frequency Histogram



Appendix B Tf-Idf per Article



Appendix C Sentiment Analysis in 2019

R Code

```
library(tidyverse)
library(tidytext)
library(dplyr)
library(pdftools)
library(tm)
library(stringr)
library(ggplot2)
library(reshape2)
library(wordcloud)
library(ggraph)
library(igraph)
library(magrittr)
library(rvest)
library(scales)

setwd("/Users/estelle_eteki/Documents/Hult Business School/Spring Semester/Text Analytics/Individual Assignment/Pdf")
files <- list.files(pattern = "pdf$")
opinions <- lapply(files, pdf_text)

corp <- Corpus(URISource(files),
               readerControl = list(reader = readPDF))

lapply(opinions, length)

my_df <- tidy(corp)

junk_list <- data.frame(
  word = c("1","iufld","wkh","ri","ulvlv","dqg","5hsruw","2017", "2018", "dphurrq",
           "lq", "qjorskrqh", "uhqfk", "3djh", "2019", "dydlodeoh", "6", "0d", "ruphu",
           "dovr", "urxs", "wr", "ulhilqj", "8", "ruhljq", "dw", "iulfd", "wzr", "urvvurdgv",
           "7donv", "rz", "ig", "glylvlrq", "xjxvw", "hw"),
  lexicon = "junk"
)

junk_list2 <- data.frame(
  word = c("cameroon","cameroon's", "anglophone", "crisis", "government", "october",
           "january", "december"),
  lexicon = "junk"
)
my_df_token <- my_df %>%
  unnest_tokens(word, text) %>%
  anti_join(stop_words) %>%
  anti_join(junk_list) %>%
  count(word, sort = TRUE)

icg17 <- my_df %>%
  unnest_tokens(word, text) %>%
  filter(id == "icg_2017.pdf") %>%
```

```
  anti_join(stop_words) %>%
  anti_join(junk_list) %>%
  count(word, sort = TRUE)

icg19 <- my_df %>%
  unnest_tokens(word, text) %>%
  filter(id == "icg_2019.pdf") %>%
  anti_join(stop_words) %>%
  anti_join(junk_list) %>%
  count(word, sort = TRUE)

my_df2 <- my_df %>%
  unnest_tokens(word, text) %>%
  anti_join(stop_words) %>%
  anti_join(junk_list) %>%
  group_by(id)

my_df_freq <- my_df %>%
  unnest_tokens(word, text) %>%
  anti_join(stop_words) %>%
  anti_join(junk_list) %>%
  count(word, sort = TRUE)

ggplot(my_df_freq %>% filter(n>=40),aes(x=reorder(word, n),y=n))+
  geom_col()+
  geom_bar(stat="identity")+
  xlab(NULL)+
  coord_flip()

ggplot(my_df_freq %>% filter(n>=50),aes(x=reorder(word, n),y=n))+
  geom_col()+
  geom_bar(stat="identity")+
  xlab(NULL)+
  coord_flip()

my_df %>%
  unnest_tokens(word, text) %>%
  anti_join(stop_words) %>%
  anti_join(junk_list) %>%
  anti_join(junk_list2) %>%
  group_by(id) %>%
  count(word, sort = TRUE) %>%
  top_n(20) %>%
  ungroup %>%
  ggplot(aes(x=reorder(word, n),y=n, fill=id))+
  geom_col(show.legend=FALSE)+
  labs(x=NULL, y="Frequency Per Article")+
  facet_wrap(~id, ncol=2, scales="free")+
  coord_flip()

# Tf_idf
```

```
my_df_token <- my_df %>%
  unnest_tokens(word, text) %>%
  anti_join(stop_words) %>%
  anti_join(junk_list) %>%
  group_by(id) %>%
  count(word, sort = TRUE)

total_words <- my_df_token %>%
  group_by(id) %>%
  summarize(total=sum(n))

my_df_join <- left_join(my_df_token, total_words)

my_df_tf_idf <- my_df_join %>%
  bind_tf_idf(word, id, n)

my_df_tf_idf <- my_df_tf_idf %>%
  arrange(desc(tf_idf))

my_df_tf_idf17 <- my_df_tf_idf %>%
  filter(id == "icg_2017.pdf") %>%
  arrange(desc(tf_idf))

my_df_tf_idf19 <- my_df_tf_idf %>%
    filter(id == "icg_2019.pdf") %>%
  arrange(desc(tf_idf))
#what can we say about these words?

##############
# looking at the graphical apprach:
my_df_tf_idf %>%
  arrange(desc(tf_idf)) %>%
  mutate(word=factor(word, levels=rev(unique(word)))) %>%
  group_by(id) %>%
  top_n(15) %>%
  ungroup %>%
  ggplot(aes(word, tf_idf, fill=id))+
  geom_col(show.legend=FALSE)+
  labs(x=NULL, y="Tf-idf Per Article")+
  facet_wrap(~id, ncol=2, scales="free")+
  coord_flip()

# Sentiment analysis

frequencies_token_afinn <- my_df %>%
  unnest_tokens(word, text) %>%
  anti_join(stop_words) %>%
  anti_join(junk_list)%>%
  inner_join(get_sentiments("afinn")) %>%
  summarise(mean(value))
```

```
frequencies_token_bing <- my_df %>%
  unnest_tokens(word, text) %>%
  anti_join(stop_words) %>%
  anti_join(junk_list)%>%
  inner_join(get_sentiments("bing")) %>%
  count(word,sentiment)%>%
  count(sentiment)

frequencies_token_nrc <- my_df %>%
  unnest_tokens(word, text) %>%
  anti_join(stop_words) %>%
  anti_join(junk_list)%>%
  inner_join(get_sentiments("nrc")) %>%
  count(word,sentiment)%>%
  count(sentiment)


# visualising setiments
icg17 %>%
  inner_join(get_sentiments("bing")) %>%
  count(word, sentiment, sort=TRUE) %>%
  acast(word ~sentiment, value.var="n", fill=0) %>%
  comparison.cloud(colors = c("grey20", "gray80"),
                   max.words=200, fixed.asp=TRUE, scale=c(0.8,0.8), title.size=1, rot.per=0.25)

icg19 %>%
  inner_join(get_sentiments("bing")) %>%
  count(word, sentiment, sort=TRUE) %>%
  acast(word ~sentiment, value.var="n", fill=0) %>%
  comparison.cloud(colors = c("grey20", "gray80"),
                   max.words=200, fixed.asp=TRUE, scale=c(0.8,0.8), title.size=1, rot.per=0.25)
my_df_token %>%
  inner_join(get_sentiments("bing")) %>%
  count(word, sentiment, sort=TRUE) %>%
  acast(word ~sentiment, value.var="n", fill=0) %>%
  comparison.cloud(colors = c("grey20", "gray80"),
                   max.words=200, fixed.asp=TRUE, scale=c(0.8,0.8), title.size=1, rot.per=0.25)
```