

A2

Exercise

Our group attempted parts of speech tagging on The Little Prince in the original French translation. In order to construct the confusion matrices for individual tags and calculate cohen's κ , R code was used.

The link to the lpp.ch4.fr.group3 is:

<https://docs.google.com/a/georgetown.edu/spreadsheets/d/1dp2bDpxuDs7Adr0rmFu0hftjAeJXQQAAa9WPH3grufA/edit?usp=sharing>

code to build confusion matrix and cohen's κ

```
cm = read.csv("confusion_matrix_POS.csv")
POS_cm = function(p1, p2){
  cm_pi = cm[, c(p1,p2)]
  levels(cm_pi[,p1]) = c(levels(cm_pi[,p1]),
                        levels(cm_pi[,p2])[-which(levels(cm_pi[,p2]) %in% levels(cm_pi[,p1]))])
  levels(cm_pi[,p2]) = c(levels(cm_pi[,p2]),
                        levels(cm_pi[,p1])[-which(levels(cm_pi[,p1]) %in% levels(cm_pi[,p2]))])

  confusion_matrix = table(cm_pi)
  confusion_matrix = confusion_matrix[order(rownames(confusion_matrix)),
                                     order(colnames(confusion_matrix))]

  print("Confusion Matrix")
  print(confusion_matrix)
  p_0 = sum(as.character(cm_pi[,p1])==as.character(cm_pi[,p2]))/nrow(cm_pi)
  p_e = sum(rowSums(confusion_matrix)/nrow(cm_pi)*(colSums(confusion_matrix))/nrow(cm_pi))
  #P(A=X) = 0 therefore P(A=X)*P(B=X) = 0

  print("Agreement rate")
  print(p_0)
  print("Cohen's Kappa")
  return((p_0-p_e)/(1-p_e))
}
```

Mohammad vs Arif

```
POS_cm("Mohammad","Arif")

## [1] "Confusion Matrix"
##           Arif
## Mohammad  ADJ ADP ADV AUX CONJ DET NOUN NUM PART PRON PROPN PUNCT SCONJ
##           0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
```

```

##      ADJ      0  5  0  0  0  0  0  0  0  0  0  0  0  0  0
##      ADP      0  0 13  0  0  0  0  1  0  0  0  0  0  0  0
##      ADV      1  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##      AUX      1  0  0  0  2  0  0  0  0  0  0  0  0  0  0
##      CONJ     0  0  0  0  0  1  0  0  0  0  0  0  0  0  0
##      DET      0  0  0  0  0  0 14  0  0  0  2  0  0  0  0
##      NOUN     1  0  0  0  0  1  0 14  1  0  0  5  0  0  0
##      NUM      0  0  0  0  0  0  2  0  1  0  0  0  0  0  0
##      PART     0  0  0  0  0  0  0  0  0  2  0  0  0  0  0
##      PRON     0  0  0  0  0  0  1  0  0  0  5  0  0  0  0
##      PROPEN  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##      PUNCT    0  0  0  0  0  0  0  0  0  0  0  0 12  0  0
##      SCONJ    0  0  0  0  0  1  0  0  0  0  2  0  0  1  0
##      VERB     1  0  0  0  0  0  1  0  0  0  0  0  0  0  0
##      X        0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##      Arif
## Mohammad VERB  X
##              0  5
##      ADJ      0  0
##      ADP      0  0
##      ADV      0  0
##      AUX      1  0
##      CONJ     0  0
##      DET      0  0
##      NOUN     0  0
##      NUM      0  0
##      PART     0  0
##      PRON     0  0
##      PROPEN  0  0
##      PUNCT    0  0
##      SCONJ    0  0
##      VERB     8  0
##      X        0  0
## [1] "Agreement rate"
## [1] 0.7428571
## [1] "Cohen's Kappa"
## [1] 0.7128532

```

There were five blanks from my POS tags compared to Mohammed's. This is because Arif was unable to determine the proper tags for these words in French. As expected both of us agreed on the PUNCT tags. The use of PRON seems to be the biggest difference between us. Of the 9 that Arif tagged as PRON, only five agreed with Mohammed. Also, Mohammed never tagged PROPENs while Arif tagged 5 words as proper-nouns. This could be due to differences in what we both consider to be names in French. Some difficulty in the name could be due to the multiple parts of speech that make up a name. Another interesting case, Arif marked all blank values with X while Mohammed left them blank. Based on Landis and Koch interpretation of κ , the matching between Mohammed and Arif is substantial. The main reason for the differences could be because Mohammad has a significant amount of experience in the French language from living in France for a while whereas Arif took 2 semester of French nearly three years ago with no real indepth use.

Mohammad vs Max

```

POS_cm("Mohammad", "Max")

## [1] "Confusion Matrix"
##           Max
## Mohammad  ADJ ADP ADV AUX CONJ DET NOUN NUM PART PRON PROPN PUNCT SCONJ
##           5   0   0   0   0   0   0   0   0   0   0   0   0   0
##    ADJ      0   4   0   0   0   0   0   0   0   0   0   1   0   0
##    ADP      0   0  14   0   0   0   0   0   0   0   0   0   0   0
##    ADV      0   0   0   0   0   0   0   0   0   1   0   0   0   0
##    AUX      0   0   0   0   4   0   0   0   0   0   0   0   0   0
##    CONJ      0   0   0   0   0   1   0   0   0   0   0   0   0   0
##    DET      0   0   0   0   0   0  11   0   0   0   5   0   0   0
##    NOUN      0   0   0   0   0   0   0  17   0   0   0   5   0   0
##    NUM      0   0   0   0   0   0   0   0   3   0   0   0   0   0
##    PART      0   0   0   0   0   0   0   0   0   2   0   0   0   0
##    PRON      0   0   0   0   0   0   0   0   0   0   6   0   0   0
##    PROPN     0   0   0   0   0   0   0   0   0   0   0   0   0   0
##    PUNCT     0   0   0   0   0   0   0   0   0   0   0   0  12   0
##    SCONJ     0   0   0   0   0   0   0   0   0   0   0   0   0   4
##    VERB      0   0   0   0   0   0   0   0   0   0   0   0   0   0
##           Max
## Mohammad VERB
##           0
##    ADJ      0
##    ADP      0
##    ADV      0
##    AUX      0
##    CONJ      0
##    DET      0
##    NOUN      0
##    NUM      0
##    PART      0
##    PRON      0
##    PROPN     0
##    PUNCT     0
##    SCONJ     0
##    VERB     10
## [1] "Agreement rate"
## [1] 0.8857143
## [1] "Cohen's Kappa"
## [1] 0.872418

```

Overall, there was a very high raw agreement rate between Max and Mohammad. The two main categories of disagreement came from tagging PROPN and tagging PRON. In the case of PROPN, Mohammad never tagged any of the words as PROPN, which likely stems from a disagreement in what is considered a name, especially if the name contains common words/nouns (such as Congres International d'Astronomie). This was the largest source of disagreement. In the other major source of disagreement, Max tagged several words as PRON, where Mohammad tagged those same words as DET. This disagreement seems to stem in how groupings of le/la/l' + verb were tagged. Max treated the le/la/l' as pronouns of direct objects of their adjacent verbs, while Mohammad treated these as determiners. The last point of disagreement stemmed from the word alors where Max tagged it as PART, but noted that he was unsure. Looking at the data with Max and Nick, it is likely that Max was mistaken on this tagging, given his uncertainty. The Landis and Koch interpretations of Kappa suggest that the matching between Mohammad and Max is substantial.

Mohammad vs Nick

```

POS_cm("Mohammad","Nick")

## [1] "Confusion Matrix"
##      Nick
## Mohammad  ADJ ADP ADV AUX CONJ DET NOUN NUM PART PRON PROPN PUNCT SCONJ
##      5    0    0    0    0    0    0    0    0    0    0    0    0    0
##      ADJ    0    4    0    0    0    0    0    0    0    0    0    1    0    0
##      ADP    0    0   12    0    0    0    1    0    0    1    0    0    0    0
##      ADV    0    0    0    1    0    0    0    0    0    0    0    0    0    0
##      AUX    0    0    0    0    4    0    0    0    0    0    0    0    0    0
##      CONJ    0    0    0    0    0    1    0    0    0    0    0    0    0    0
##      DET    0    2    0    0    0    0   13    0    0    0    1    0    0    0
##      NOUN    0    0    0    0    0    0    0   17    1    0    0    4    0    0
##      NUM    0    0    0    0    0    0    1    0    2    0    0    0    0    0
##      PART    0    0    0    0    0    0    0    0    0    2    0    0    0    0
##      PRON    0    0    0    0    0    0    0    0    0    0    6    0    0    0
##      PROPN   0    0    0    0    0    0    0    0    0    0    0    0    0    0
##      PUNCT   0    0    0    0    0    0    0    0    0    0    0    0   12    0
##      SCONJ   0    0    0    0    0    1    0    0    0    0    0    0    0    3
##      VERB    0    0    0    0    0    0    0    0    0    0    0    0    0    0
##      Nick
## Mohammad VERB
##      0
##      ADJ    0
##      ADP    0
##      ADV    0
##      AUX    0
##      CONJ    0
##      DET    0
##      NOUN    0
##      NUM    0
##      PART    0
##      PRON    0
##      PROPN   0
##      PUNCT   0
##      SCONJ   0
##      VERB   10
## [1] "Agreement rate"
## [1] 0.8761905
## [1] "Cohen's Kappa"
## [1] 0.8615057

```

Mohammad and Nick have an excellent agreement rate and an excellent Cohens Kappa as well. Their disagreements come on the easily controversial tags. For example, on PROPN they have no agreement; however this appears to be due to how Nick tagged the organization names in the text, opting to have every word in the organization name be a proper noun. Mohammad took the alternative approach which is to tag every word as its proper part of speech independent of whether it is part of a name. Their other disagreements were on similar types of issues where the confusion is easily explained. They had minor disagreement on ADP and DET, though in French this is a common confusion because of terms like *de* which can in some contexts mean from and in others be used to express that one has some of something. Additionally there was one disagreement where Nick said that something was a CONJ but Mohammad put SCONJ; both recognized the conjunction yet disagreed on whether it was subordinate. Overall, their agreement is excellent by any

of the proposed metrics for assessing a Cohen Kappa.

Nick vs Arif

```
POS_cm("Nick", "Arif")

## [1] "Confusion Matrix"
##      Arif
## Nick  ADJ  ADP  ADV  AUX  CONJ  DET  NOUN  NUM  PART  PRON  PROP  PUNCT  SCONJ
##      0   0   0   0   0   0   0   0   0   0   0   0   0   0
## ADJ    0   4   0   0   0   0   0   0   0   0   2   0   0   0
## ADP    0   0  11   0   0   0   1   0   0   0   0   0   0   0
## ADV    1   0   0   0   0   0   0   0   0   0   0   0   0   0
## AUX    1   0   0   0   2   0   0   0   0   0   0   0   0   0
## CONJ   0   0   0   0   0   1   0   0   0   0   1   0   0   0
## DET    0   0   1   0   0   0  14   0   0   0   0   0   0   0
## NOUN    1   0   0   0   0   1   0  13   1   0   0   1   0   0
## NUM    0   0   0   0   0   0   1   0   1   0   0   1   0   0
## PART    0   0   1   0   0   0   0   0   0   2   0   0   0   0
## PRON    0   0   0   0   0   0   2   0   0   0   5   0   0   0
## PROP    0   1   0   0   0   0   0   1   0   0   0   3   0   0
## PUNCT   0   0   0   0   0   0   0   0   0   0   0   0  12   0
## SCONJ   0   0   0   0   0   1   0   0   0   0   1   0   0   1
## VERB    1   0   0   0   0   0   1   0   0   0   0   0   0   0
## X       0   0   0   0   0   0   0   0   0   0   0   0   0   0
##      Arif
## Nick  VERB  X
##      0   5
## ADJ    0   0
## ADP    0   0
## ADV    0   0
## AUX    1   0
## CONJ    0   0
## DET    0   0
## NOUN    0   0
## NUM    0   0
## PART    0   0
## PRON    0   0
## PROP    0   0
## PUNCT   0   0
## SCONJ   0   0
## VERB    8   0
## X       0   0
## [1] "Agreement rate"
## [1] 0.7333333
## [1] "Cohen's Kappa"
## [1] 0.7043741
```

Nick and Arif have fairly good agreement rate. The majority of their disagreement comes from the fact that Arif omitted tags where he was completely unable to decipher the part of speech and he put in X for blank words where Nick simply left these blank. On the majority of tags they do have good agreement with confusion coming on the objectively confusing words. For example, the word *le/la/l* preceding verbs was a point of contention with Arif labeling it a DET and Nick labeling it a PRON. This makes sense since in those

situations le/la/l takes the place of the object, but normally le/la/l is a determiner. Another agreement issue, which is common to many of the annotation differences, is what words are PROPNS. With organization names and names containing numbers there is much room for disagreement about which tag is correct. Other disagreements were minor such as AUX vs VERB and CONJ vs SCONJ. Some amount of error can certainly be attributed to the fact that neither annotator is a native speaker. Overall, by Landis and Kochs guidelines the pair had substantial agreement and by Fleiss standards the pair had almost excellent agreement.

Nick vs Max

```
POS_cm("Nick", "Max")

## [1] "Confusion Matrix"
##           Max
## Nick      ADJ ADP ADV AUX CONJ DET NOUN NUM PART PRON PROPN PUNCT SCONJ
##           5  0  0  0  0  0  0  0  0  0  0  0  0  0
## ADJ       0  4  0  0  0  0  0  0  0  0  2  0  0  0
## ADP       0  0 12  0  0  0  0  0  0  0  0  0  0  0
## ADV       0  0  0  0  0  0  0  0  0  1  0  0  0  0
## AUX       0  0  0  0  4  0  0  0  0  0  0  0  0  0
## CONJ      0  0  0  0  0  1  0  0  0  0  0  0  0  1
## DET       0  0  1  0  0  0 11  0  1  0  2  0  0  0
## NOUN      0  0  0  0  0  0  0 17  0  0  0  0  0  0
## NUM       0  0  0  0  0  0  0  0  2  0  0  1  0  0
## PART      0  0  1  0  0  0  0  0  0  2  0  0  0  0
## PRON      0  0  0  0  0  0  0  0  0  0  7  0  0  0
## PROPN     0  0  0  0  0  0  0  0  0  0  0  5  0  0
## PUNCT     0  0  0  0  0  0  0  0  0  0  0  0 12  0
## SCONJ     0  0  0  0  0  0  0  0  0  0  0  0  0  3
## VERB      0  0  0  0  0  0  0  0  0  0  0  0  0  0
##           Max
## Nick      VERB
##           0
## ADJ       0
## ADP       0
## ADV       0
## AUX       0
## CONJ      0
## DET       0
## NOUN      0
## NUM       0
## PART      0
## PRON      0
## PROPN     0
## PUNCT     0
## SCONJ     0
## VERB     10
## [1] "Agreement rate"
## [1] 0.9047619
## [1] "Cohen's Kappa"
## [1] 0.8945254
```

Between Nick and Max there were six categories for disagreement, however each category only had minor disagreements. The biggest source of disagreement was in the tagging of PRON, where on four occasions,

Max tagged the word as PRON and Nick did not. Nick tagged possessive pronouns such as *sa* or *son* as ADJ, while Max tagged them as PRON. Similar to the situation with Mohammad and Max, there was some disagreement in groupings of *le/la/l'* + verb, where Max always tagged the *le/la/l'* as PRON, while Nick sometimes tagged this as DET (and in other contexts as PRON). The two also had a disagreement regarding SCONJ and CONJ in one instance, but both still saw the word as some kind of conjunction (differing only in which type of conjunction). There was an instance of disagreement for names, in which Nick considered an ADP within a name to actually be a PART, while Max considered it an ADP. Nick noted that this decision was influenced by the fact that the ADP within the name should not be parsed without the surrounding PROPON was thus a PART. There was also a disagreement in tagging *alors*, where Max tagged it as PART and Nick tagged it as ADV. Max noted that he was unsure of this tag, and given Mohammad's tagging of *alors* as ADV, it is likely that Max was incorrect in this tagging. Despite these many minor disagreements, Max and Nick had the highest agreement rate, at 90.476%, with a similarly high Cohen's Kappa suggesting that the matching between Nick and Max is substantial. This could possibly be attributed to both participants having similar backgrounds in French and English.

Arif vs Max

```
POS_cm("Arif", "Max")

## [1] "Confusion Matrix"
##           Max
## Arif      ADJ ADP AUX CONJ DET NOUN NUM PART PRON PROPON PUNCT SCONJ VERB
##           0  0  0  1  0  0  1  0  1  0  0  0  0  1
## ADJ       0  4  0  0  0  0  0  0  0  0  1  0  0  0
## ADP       0  0 13  0  0  0  0  0  0  0  0  0  0  0
## AUX       0  0  0  2  0  0  0  0  0  0  0  0  0  0
## CONJ      0  0  0  0  1  0  1  0  0  0  0  0  1  0
## DET       0  0  1  0  0 11  0  2  0  4  0  0  0  1
## NOUN      0  0  0  0  0  0 13  0  0  0  1  0  0  0
## NUM       0  0  0  0  0  0  1  1  0  0  0  0  0  0
## PART      0  0  0  0  0  0  0  0  2  0  0  0  0  0
## PRON      0  0  0  0  0  0  0  0  0  7  0  0  2  0
## PROPON    0  0  0  0  0  0  1  0  0  0  4  0  0  0
## PUNCT     0  0  0  0  0  0  0  0  0  0  0 12  0  0
## SCONJ     0  0  0  0  0  0  0  0  0  0  0  0  1  0
## VERB     0  0  0  1  0  0  0  0  0  0  0  0  0  8
## X         5  0  0  0  0  0  0  0  0  0  0  0  0  0
##           Max
## Arif      X
##           0
## ADJ       0
## ADP       0
## AUX       0
## CONJ      0
## DET       0
## NOUN      0
## NUM       0
## PART      0
## PRON      0
## PROPON    0
## PUNCT     0
## SCONJ     0
## VERB     0
```

```
##      X      0
## [1] "Agreement rate"
## [1] 0.752381
## [1] "Cohen's Kappa"
## [1] 0.7260686
```

Like the case with Mohammad vs Arif, Arif used X values for blanks words whereas Max Omitted them. The two categories with majority disagreements were DET and PRON. While a majority of DET tags were agreed upon, Max disagreed and labels some of these as PRON. This could be do to cases where *le/la/l* was used in place of an object, but the POS tag guide generally states that this is a DET. One of the starkest difference is the agreement of tags that are SCONJ. Max and Arif disagree in most cases. This could be due to the difference in French between CONJ and SCONJ. Based on Landis and Koch interpretation of κ , the matching between Arif and Max is substantial. The main reason for the differences could be difference in lnaguage exposure between Arif and Max. Based off of discussions during the tagging, Max has had years of study in the French language in addition to a fluent Parent.