Implementing Machine Translation of French Statements and Questions via Replacement and Movement Rules
by Andre Lee, CS 4744 Spring 2018

1. Introduction

This project was an exercise in using the Foma toolkit and movement rules we learned in computational linguistics class this semester to explore translating French statements and questions into proper English.

By using a corpus I was able to obtain data in the form of published French sentences and questions.

My methodology was to use the regular expression manipulation abilities of Foma to take in a French sentence and translate individual words and expressions into English. The file used is named translateWords.fst (Appendix B).

The second step in the translation process was using a Java program called Movement.java (Appendix C) to manipulate the strings output by my Foma program. Movement.java implements movement rules to translate the French grammar word order into a grammatically correct English sentence.

The movement rules I worked with were taken from my own observations as well as those outlined in articles referenced below. The information on individual movement rules contributed by them will be expanded upon in section 3.3.

Emphasis was placed on implementing the movement rules as the research question: *What movement rules should be used when translating from French to English, and in what environments do specific rules occurs?*

2. Explanation of Corpus Used

The corpus used in this project is called French Gigaword Third Edition , obtained properly via the University of Pennsylvania's Linguistic Data Consortium (catalog number LDC2011T10). The corpus contains detailed and well-catalogued newswire from Agence France-Presse and Associated Press French Service from 1994 to 2010.

The corpus was used to obtain sentences for data in the project. There are four categories of documents included: main articles "story", news summaries "multi", correspondence to news editors "advis" and "other". The sentences mainly come from articles ("story") published in 2010, as I tried to obtain vocabulary that was both more modern and more familiar to me. I focused on obtaining as many lexical categories as I could, rather than just expanding vocabulary size. As the corpus is newswire, sentences used are generally quite formal and well-detailed. As such and due to my own limits, thirty sentences were chosen in total: 20 declarative statements and 10 interrogative ones (Appendix A). The scope of the corpus also became a factor in limiting my machine translation to statements and questions. Due to the nature of publication, the newswire did not produce enough data in the realms of imperatives, exclamations or slang.

If additional work is put into the project, data could be obtained via a focus on maximizing the amount of syntactic structures rather than a focus on lexical categories or vocabulary size as detailed previously.

3. Implementation of Project

3.1 Comparison to Proposal

My original proposal for this project included a much larger scope and a more complicated methodology.

For reasons stated above, my project scope was narrowed to only statements and questions, and the words in the project lexicon were taken from fewer sentences. Final testing as a result was also more limited.

Methodology was largely unchanged except that the use of the lopar toolkit detailed in the proposal was not implemented in the final project. The use of lopar to compare my results to an English grammar designed in lopar to test correctness seemed extraneous. As a native English speaker, this step appeared unnecessary as it would be evident throughout the project if the grammar of

my results and tests were correct. Also, comparing the results of the machine translation to a real translation of the French sentence into English was assumed to be a much more realistic measure of correctness.

Additionally, the use of Foma in the proposal and in the project differed slightly. Words and small expressions were translated directly, categorized by usage and lexical category rather than translating word roots and applying morphological to translate usage. The latter approach would be more accurate by erasing some ambiguities in the word translation process. However, the former is more conducive to the movement rules which were the main focus of the project.

3.2 Replacement Rules

All replacement rules were created via Foma, using the file translateWords.fst, which can be found in Appendix B.

The online dictionary Word Reference was used for all word translations. It was chosen for its inclusion of French expressions and a multitude of different translations depending on usage and context. Due to the ambiguous nature of translating French words into English, if multiple translations exist, the Foma program would transform the French word into all of them separated by "/". For example, "de" can be commonly translated into the prepositions "of" or "from"(wordreference.com), so the Foma program outputs "of/from" as a result.

Overall, there are three functions directly applied to an input French sentence: preprocess, replace, post. The program is designed to be used with Foma in the command line. Applying down through all three functions will produce the desired output. Sometimes multiple outputs are created due to ambiguities of translation; in those cases, the latest output was used.

Foma was chosen for its regular expression manipulation abilities, but the project was complicated by a bug where a function cannot be applied to a regular expression modified directly by another function. Concatenating the functions into a single translation function would be more efficient, but is not possible due to this issue.

3.21 Preprocess

Preprocess, as the name implies, is the function used to preprocess a French sentence for translation. It is used to turn the entire sentence lowercase and isolate punctuation away from all letters, removing contractions and converting the French number system into the English one.

3.22 Replace

The replace function replaces individual French words and expressions with their English translations. It is a series of different functions that translate words based on lexical categories and usage. As they include multiple words, expressions are translated first. Nouns, verbs, adverbs, complementizers, WH-words, prepositions, etc. are translated directly, except those listed below.

A. Pronouns

Pronouns were divided into subject and object pronouns. Even though they can be ambiguous, pronouns can be translated directly. The special case is when pronouns come before the verb in a sentence. These are object pronouns and since the word "the" in French can be translated in "the" or "it" depending on this specific case, object pronouns are translated with the verbs to keep track of context.

B. Adjectives

In French, there is agreement for adjectives, meaning they have special endings depending on the gender and number denoted by the noun or subject the adjective is describing (wordreference.com). In the Foma program, the function to translate adjectives translates the base adjective directly and removing the agreement endings.

C. Determiners

Determiners are mainly translated directly in the Foma program. However, in English, determiners are less common as noun phrases denoting more general ideas do not use determiners. In French, these types of noun phrases usually use the determiner "the". To simplify the direct translation, this word is kept throughout the translation process.

Additionally, the phrase "of the" in French can also be translated into the determiner "some" in

certain contexts (wordreference.com). Since a lot is determined by meaning, “of the” in French only translates into “some” when coming after a preposition, as in both English and French, prepositional phrases contain a preposition followed by either an NP or DP.

3.23 Post

The function post is a postprocessing function for after all the French words have been translated.

A. Preposition Deletion

Double prepositions are made into single prepositions, since as stated above prepositional phrases generally do not have multiple prepositions. The main case this is accounting for is when translating a phrase containing a preposition followed by an infinitive. When translating an infinitive from French to English, the verb becomes “to” followed by the base form of the translated verb (wordreference.com). The preposition deletion is to delete this new preposition in contexts where it is not needed. Prepositions are also deleted when preceded by modal verbs and subject pronouns.

B. Days of the Week

When talking about the day an action occurs, French sometimes uses the word “the” where in English the word “on” is used. The replacement occurs in the function post.

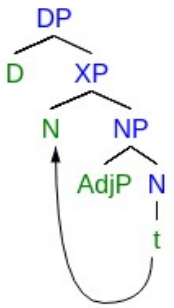
3.3 Movement Rules

The movement rules used in the project are only those that produce sentences that are acceptable as a native English speaker. The movement rules are implemented by a Java program Movement.java outlined in Appendix C.

The program turns the input sentence into an ArrayList of individual words. Using sets of words in the lexicon characterized by function, each word in the sentence can be assigned an additional type, making the implementation of movement rules possible.

3.31 Adjective Movement

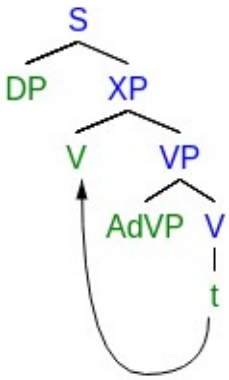
Adjectives in French generally come after the noun that they modify. Adjectives that come before the noun are identified by their meaning (associated with beauty, goodness, age or size).



The movement rule from English to French is that a trace is placed where the noun in a noun phrase is, and the noun itself raises itself as a head (Laenzlinger, 56). In the Java program, the function adjectiveMovement reverses this rule by moving the adjective to the front of the noun to recreate the English noun phrase.

3.32 Adverb Movement

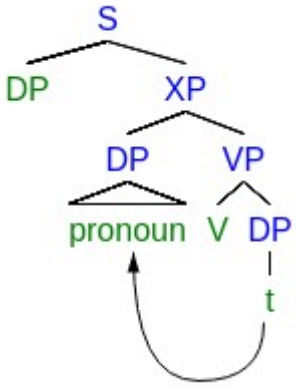
Adverbs in French experience different movement depending on what they are modifying. When modifying adjectives, they almost always precede the adjective (Pollock, 366). When modifying verbs, they come after. From English to French, there is movement of the verb in head raising to precede the adverb (Pollock, 367). However, in English, there are some adverbs that come after the verb they modify, most notably those that are time/timing related.



In the Java program, adverbMovement works similarly to adjectiveMovement by moving the adverb (if it normally functions in English as a preceding adverb) to the front of the verb to preserve the English verb phrase.

3.33 Object Pronoun Movement

Object pronouns are all clitics in French. When translating from English to French there is an option to move them from the determiner phrase after the verb to between the subject and the verb (Sichel). When this movement occurs, the word also changes form into a different set of object



pronouns. This has already been resolved in the Foma program as these forms translate into the same words in English.

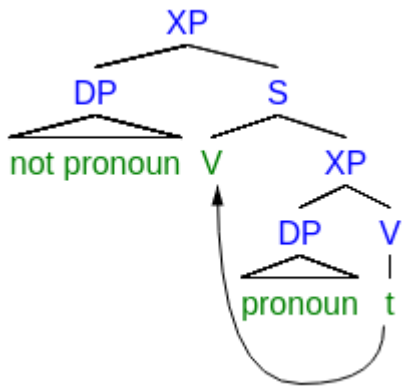
In the Java program, the function objectPronounMovement directly reverses this movement by moving the pronoun in front of the verb if the pronoun is found between the subject and the verb.

3.34 Movement in Questions

Questions in both English and French are generally in one of two camps: either requesting a specific piece of information using a WH-word (type 1), or requesting the truth value of a certain statement (type 2).

A. Verb Phrase Inversion

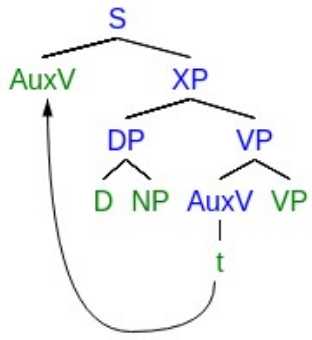
While both French and English use WH-movement and do-support to facilitate turning statements into questions, French uses verb phrase inversion, where the first verb is fronted to form a type 2 question (Pollock, 367). However, this only works if the subject is a pronoun. If the subject is any other type of determiner phrase, the determiner phrase is essentially doubled by adding another pronoun and moving the verb in front of that pronoun.



The pronounDeletion function of the Java program detects this type of structure in a sentence and removes the added pronoun to restore the original word order.

B. Auxiliary Verb Movement

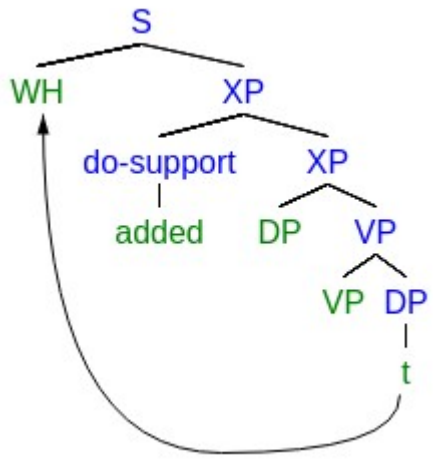
English has a version of verb phrase inversion as well, but only for auxiliary verbs, and unlike in French, works identically with pronoun DP's as well as non-pronoun DP's (Pollock, 368).



After running pronounDeletion on a question, if the first verb in the sentence is an auxiliary verb (in my data it usually is for questions, as articles usually use an auxiliary to add a sense of uncertainty to questions directed at the readers, or with a tense marker to reference specific timelines) then the verb is fronted to head the sentence. This is done via the auxVerbMovement function.

C. Do-Support

As a native English speaker, I have observed do-support in English questions can be used in two ways: added to the beginning of a sentence to form a type 2 question or added to the beginning of a statement prior to WH-movement to form a type 1 question.



The first do-support is the same as in French, using the expression "est-ce que" preceding a statement to ask a question. As this is simple mapping from a French word to an English translation this is done by the Foma program.

The second type of do-support is not present in French. The WH-movement is the same. The doSupport function in the java program adds do-support between the WH-word and the subject. Additionally, French sentences sometimes use the infinitive verb as a subject with WH-movement. This is not as present in English as it is more common to add a hypothetical subject "you" as well. For example, "why to take the cookie?" becomes "why *do you* take the cookie?".

4 Results

Results were based on how close the results of my machine translations of the sentences taken from the corpus were to actual human translations of them. I provided the translations using my own knowledge of the French language. The sets of sentences are listed in Appendix A.

Overall, a majority of the output sentences are identical or close to identical in meaning to the human translations. As to be expected, smaller phrases and simpler sentences were more accurate.

Result Level	Sentences
Identical in Meaning	19
Close to Identical in Meaning (contained small differences)	8
Clear Error (missing or wrong details)	3

4.1 Test Cases

These are sentences created using the lexicon for the project, but are different from the sentences in the data. They are randomly generated without a huge number of words. We will compare the results to the human translations:

1. Le roi Somalien va rester au journaliste.
Human Translation: The Somalian king is going to stay at the journalist.
Machine Translation Results: The Somalian king is going to stay at the journalist.
2. Le voix de Marcelo est nucléaire dans la matière.
Human Translation: The voice of Marcelo is nuclear in the matter.
Machine Translation Results: The voice of/from Marcelo is nuclear in the matters.
3. Doit-il trouver la soirée à Washington?
Human Translation: Must he find the evening at Washington?
Machine Translation Results: Must he/it find the evening to/at Washington?
4. Ils interpellent les passagers sur les métaux temporaires.
Human Translation: They address the passengers on the temporary metals.
Machine Translation Results: They address the passengers on the temporary metals.
5. Il a assassiné la féministe?
Human Translation: He has assassinated the feminist?
Machine Translation Results: Has he/it assassinated the feminist?

The test cases seem to have all passed. There are slight differences in word order that have little effect on the meanings of sentences.

4.2 Issues

Since some leeway was given regarding ambiguities for individual words, the results above might not be the best representation of how well my machine translations actually did.

Looking at the machine translated data sentences, determiners continue to be an issue as they change in different contexts in English that are very meaning dependent, as well as features like compound nouns that were not explored in the project. In longer sentences there also seem to be more errors as adjunct placement was not explored in the project and so made the machine translations very awkward and harder to understand.

When continuing to examine translations from French to English, topics like these should be expanded upon.

5. Conclusion

5.1 Project Summary

Overall, splitting the project into two halves, replacement and movement made it easier to understand many of similarities in the underlying structures of French and English.

Even though originally the project was focused on movement rules and the contexts in which they appear, a lot of the project also ended up devoted to replacement rules. There were many characteristics of function words in French that are not present in English, and it was a challenge finding ways to accommodate them. Some aspects did not end up accommodated and leeway was given on how accurately individual words were translated.

Movement rules were examined based on what was observed from the data sentences. Adjective movement, adverb movement and object pronoun movement were studied as applicable to both statements and questions, while auxiliary verb movement, pronoun deletion and do-support rules were specifically observed and implemented only with questions.

The results of my machine translation programs were tested against my original data sentences as well as randomly generated ones. The success rate for sentences was a little low at around 90% of sentences had close to perfect meaning conservation when translating from French to English. This percentage goes up as sentence size decreases and correctness appears to improve when considering the accuracy of translating phrases within sentences.

5.2 Possible Improvements

The clearest improvement to the project would be the inclusion of more data. There just were not a large number of sentences used; the corpus could have been used more effectively in finding diverse syntactic structures and vocabulary.

To improve the efficiency of a larger project, adding tags to words and phrases in the corpus would help sort through possible data. For this project such tags could not be implemented, but it would have been a valuable asset.

5.3 Future Ideas

To improve machine translations further, elements of machine learning could be added to the project. Data could also be taken from English sentences as well as French ones. Machine learning could identify syntactic contexts where new rules could be applied to correct the mistakes in this current project's machine translations. Things like compound nouns and null determiners could be implemented as a result.

Additionally, the scope of the project could be spread to include imperative or casual speech as well. The corpus' focus on formal and informative articles could be limiting at times, so adding new sources of data from telephone conversations or entertainment literature could expand what is possible using the project's implementation of replacement and movement rules.

6. References

Graff, David, Ângelo Mendonça, and Denise DiPersio. French Gigaword Third Edition LDC2011T10. DVD. Philadelphia: Linguistic Data Consortium, 2011.

Kellogg, Michael. *WordReference.com.* WordReference.com, 2018, wordreference.com. Accessed 25 Apr. 2018.

Laenzlinger, Christopher. "French Adjective Ordering: Perspectives on DP-Internal Movement Types." *Department de Linguistique*, Université de Genève, 2005. Accessed 16 May 2018.

Pollock, Jean-Yves. "Verb Movement, Universal Grammar, and the Structure of IP." *Linguistic Inquiry*, Vol. 20, No. 3, The MIT Press, 1989. pp. 365-424. Accessed 16 May 2018.

Sichel, Ivy. "Pronoun Movement and Attraction." *Proceedings of the Israeli Association for Theoretical Linguistics*, Hebrew University of Jerusalem, 2002. Accessed 16 May 2018.

APPENDIX A:

L'ONU et la France interpellent les candidats.

Real Translation: The United Nations and France address the candidates.

Project Results: The United Nations and France address the candidates.

Par la voix de son président Youssouf Bakayoko, la commission a réitéré dans la soirée qu'elle avait jusqu'à mercredi "pour rendre les résultats".

Real Translation: Through the voice of its president Youssouf Bakayoko, the commission reiterated in the evening that it had until Wednesday "to return the results".

Project Results: By the voice of/from his/her/its president Youssouf Bakayoko, the commission has reiterated in the evening what/that/than she/her/it had until Wednesday for/to render the results.

Le samedi, les corps enterrés d'un homme et d'une femme ont été retrouvés.

Real Translation: On Saturday, the buried bodies of a man and a woman were recovered.

Project Results: On Saturday, the buried bodies of/from a/an man and of/from a/an woman/wife have been recovered.

Les victimes avaient apparemment été assassinées à des dates distinctes.

Real Translation: The victims had apparently been assassinated at some distinct date.

Project Results: The victims had apparently been assassinated to/at some distinct dates.

Les dirigeants des 47 pays avaient pris en avril à Washington des engagements pour combattre le "terrorisme nucléaire".

Real Translation: The leaders of 47 countries had taken commitments in April at Washington to fight “nuclear terrorism”.

Project Results: The leaders of/from the 47 nation(s) had taken in/to April to/at Washington of/from the commitments for/to fight the nuclear terrorism.

L'euro se stabilisait face à la monnaie nippone, à 111,84 yens contre 112,01 yens vendredi soir.

Real Translation: The euro stabilized against the Japanese currency at 111.84 yen against 112.01 yen Friday night.

Project Results: The euro stabilized opposite the Japanese currency, to/at 111.84 yen against 112.01 yen Friday night.

Nous confirmons la mort de 3 citoyens américains à Ciudad Juarez.

Real Translation: We confirm the deaths of 3 American citizens at Cuidad Juarez.

Project Results: We/us confirm the death of/from 3 American citizens to/at Ciudad Juarez.

Leur rivalité a fait plus de 2.600 morts en 2009 et déjà quelque 2.500 cette année dans cette ville de 1,3 million d'habitants.

Real Translation: Their rivalry caused more than 2600 deaths in 2009 and already some 2500 this year in this city of 1.3 million people.

Project Results: Their rivalry has done/made more than 2,600 deaths in/to 2009 and already some 2,500 that year in that town/city of/from 1.3 million inhabitants.

Les pirates s'emparent d'un nombre croissant de navires en 2010.

Real Translation: The pirates capture an increasing number of ships in 2010.

Project Results: The pirates take possession of/from a/an increasing number of/from ships in/to 2010.

Un navire comorien, le Aly Zoulfecar, a été détourné par des pirates somaliens avec ses neuf membres d'équipage et ses 20 passagers alors qu'il faisait route vers Dar Es Salam, en Tanzanie, a annoncé mercredi la force navale antipiraterie de l'Union européenne, Atalante.

Real Translation: A Comorian ship, the Aly Zoulfecar, had been captured by Somalian pirates with its 9 crew members and its 20 passengers while en route to Dar Es Salam in Tanzania, announced on Wednesday the European Union antipiracy naval force Operation Atalanta.

Project Results: A/an Comorian ship, the Aly Zoulfecar, has been detained by some Somalian pirates with his/her/its new members of/from equipment and his/her/its 20 passengers while he/it did/made route/way towards Dar Es Salam, in/to Tanzania, has announced Wednesday the antipiracy naval force of/from the European Union, Operation Atalanta.

Le pape consacre la Sagrada Familia, symbole de la famille traditionnelle.

Real Translation: The Pope blessed the Sagrada Familia, symbol of the traditional family.

Project Results: The Pope blesses the Sagrada Familia, symbol of/from the traditional family.

Les métaux industriels profitaient aussi de l'affaiblissement du dollar.

Real Translation: Industrial metals also profited from the weakening of the dollar.

Project Results: The industrial metals also profited of/from the weakening of/from the dollar.

L'intellectuel Liu Xiaobo, à qui est remis de façon symbolique vendredi à Oslo le prix Nobel de la paix, incarne désormais aux yeux du monde, depuis sa prison, la lutte sans concession pour les idéaux démocratiques en Chine.

Real Translation: The intellectual Liu Xiaobo, to whom is returned the Nobel Peace Prize on Friday in symbolic fashion at Oslo, henceforth embodies to the eyes of the world, since his imprisonment, the uncompromising struggle for democratic ideals in China.

Project Results: The intellectual Liu Xiaobo, to/at who/whom/that is returned of/from symbolic fashion Friday to/at Oslo the Nobel prize of/from the peace, henceforth embodies to/at the eyes of/from the world, since his/her/its prison/imprisonment, the struggle without compromise for/to the democratic ideals in/to China.

Le ministre polonais de la Défense Bogdan Klich a indiqué jeudi qu'un accord sur le stationnement temporaire des avions américains de combat F-16 et des appareils de transport Hercules pourrait être trouvé entre Varsovie et Washington au 1er semestre 2011.

Real Translation: The Polish Minister of Defense Bogdan Klich indicated on Thursday that an agreement on the temporary stationing of F-16 American combat planes and Hercules transport aircraft could be found between Warsaw and Washington in the first half of 2011.

Project Results: The Polish minister of/from the defense Bogdan Klich has indicated Thursday what/that/than a/an agreement on the temporary parking of/from the American planes of/from combat F 16 and of/from the aircraft of/from transport Hercules could be found between Warsaw and Washington to/at the first half of the year 2011.

Selon la presse espagnole, le club madrilène chercherait également un latéral gauche pour concurrencer le Brésilien Marcelo.

Real Translation: According to the Spanish press, the Madridian club will also search for a left to compete with the Brazilian Marcelo.

Project Results: According to the Spanish press, the Madridian club will equally search for a/an left for/to compete with the Brazilian Marcelo.

Le roi Abdallah s'était envolé pour New York le 22 novembre et a été opéré deux jours plus tard pour une hernie discale compliquée par un hématome.

Real Translation: King Abdullah was flown to New York the 22nd of November and was operated on two days later for a slipped disc complicated by a hematoma.

Project Results: The king Abdullah was flown for/to New York the 22 November and has been operated on two days later for/to a/an slipped disc complicated by a/an hematoma.

Le trafic est quasiment redevenu normal jeudi.

Real Translation: Traffic was almost back to normal on Thursday.

Project Results: The traffic is almost back to normal Thursday.

M. Deighton a expliqué que des "ajustements" avaient été nécessaires.

Real Translation: Mr. Deighton explained that some “adjustments” had been necessary.

Project Results: Mr. Deighton has explained what/that/than some adjustments had been necessary.

Les forces de sécurité ont fermé les rues proches du poste de police, interdisant son accès aux journalistes.

Real Translation: The security forces closed the roads close to the police post, restricting its access to journalists.

Project Results: The forces of/from security have closed the roads close of/from the post of/from police, forbidding his/her/its access to/at the journalists.

Agent de la CIA, féministe enragée/amoureuse des musulmans, fondamentaliste chrétienne, lesbienne et mortellement éprise d'un homme, peut-on être tout cela à la fois?

Real Translation: CIA agent, enraged feminist/Muslim lover, Christian fundamentalist, lesbian and mortally in love with a man, can you be all that at the same time?

Project Results: Agent of/from the CIA, enraged feminist / lover of/from the Muslims, Christian fundamentalist, lesbian and mortally in love with a/an man, can one/you be all that
at the same time?

WikiLeaks aura-t-il le même effet en matière de diplomatie, de transparence et de liberté d'expression?

Real Translation: Will WikiLeaks have the same effect in matters of diplomacy, transparency and freedom of expression?

Project Results: Will WikiLeaks have the same effect in/to matters of/from diplomacy, of/from transparency and of/from freedom of/from expression?

Les Etats peuvent-ils parvenir à des finances publiques viables?

Real Translation: Can the states achieve sustainable public finances?

Project Results: Can the states achieve some sustainable public finances?

Qu'est-ce qui va se passer après?

Real Translation: What is going to happen after?

Project Results: What/that/than is going to pass/happen after?

Comment suivre les cours lorsqu'on est gravement malade et que l'on doit rester longtemps à la maison ?

Real Translation: How do you pursue courses when you are seriously sick and must rest at home for a long time?

Project Results: How do you pursue the courses when one/you is seriously sick and what/that/than one/you must rest a long time to/at home?

Peuvent-ils le faire seuls?

Real Translation: Can they do it alone?

Project Results: Can they do/make it alone?

Comment le stress affecte-t-il son attitude vis-à-vis de ses conseillers et ses prises de décision ?

Real Translation: How does the stress affect its attitude regarding its advisors and decision making?

Project Results: How do/does the stress affect his/her/its attitude regarding/opposite his/her/its advisors and his/her/its decision making?

Comment Cristina Fernandez de Kirchner gère-t-elle ses nerfs et son anxiété ?
Real Translation: How does Cristina Fernandez de Kirchner deal with her nerves and anxiety?

Project Results: How do/does Cristina Fernandez de Kirchner deal with his/her/its nerves and his/her/its anxiety?

Qu'est-ce qui a fait pencher la balance en votre faveur?

Real Translation: What tipped the balance in your favor?

Project Results: What/that/than has tipped the balance in/to your favor?

Mais où trouver ces 100 milliards?

Real Translation: But where to find the 100 billion?

Project Results: But where do you find these/those 100 billions?

APPENDIX B:

```
#preprocess sentence for usage
define toLowercase A -> a , B -> b , C -> c , D -> d , E -> e , F -> f , G -> g , H -> h , I -> i , J -> j , K -> k , L -> l , M -> m , N ->  n , O -> o , P -> p , Q -> q , R -> r , S -> s , T -> t , U -> u , V -> v , W -> w , X -> x , Y -> y , Z -> z;
define removeContractions [du|des] -> [de " " le] , [au|aux] -> [à " " le] || [" " | .#. ] _ [" " | .#.];
define digits ["0" | "1" | "2" | "3" | "4" | "5" | "6" | "7" | "8" | "9"];
define numbers [digits+ ["." digits+]^<2 ];
define punctuation2 """ -> [e " "] .o. "." -> " ." , "?" -> " ?" || _ .#.;
define punctuation1 "," -> "." , "." -> "," || digits+ _ digits+;
define punctuation3 "," -> " ," || _ " " .o. ["-t-" | "-"] -> " " .o. "/" -> " / " || ?* _ ?*;


#TRANSLATIONS

#expression translations
define Expressions1 [plus " " de] -> [more " " than] , [alors " " que] -> [while] , [selon] -> [according " " to] , [face " " à] -> [opposite] , [plus " " tard] -> [later] , [à " " la " " fois] -> [at "-" the "-" same "-" time] , [est " " ce " " qui] -> 0 , [vis " " à " " vis] -> ["regarding/opposite"] || [.#. | " "] _ [.#. | " "] .o. [se " " | ne " " ] -> 0 || [.#. | " "] _;
define Expressions2 [le] -> 0 || [que] " " _ " " [on] " " ;
define Expressions3 [ce " " [qui | que]] -> [what] || [.#. | " "] _ [.#. | " "];
define Expressions4 [de " " [le|la]] -> "some" || [par | de | dans | jusque | à | pour | en | contre | avec | vers | sans | sur | entre | depuis | que | qui | lorsque | pourquoi | où | comment] " " _ [.#. | " "];
define Expressions Expressions1 .o. Expressions2 .o. Expressions3 .o. Expressions4;

#individual word translations - verbs
define justVerbs [ [rendre] | [combattre] | [être] | [concurrencer] | [passer] | [suivre] | [rester] | [parvenir] | [faire] | [trouver] | [a] | [ont] | [interpellent] | [confirmons] | [emparent] | [consacre] | [incarne] | [peut] | [peuvent] | [va] | [est] | [doit] | [affecte] | [gère] | [avait] | [avaient] | [stabilisait] | [faisait] | [profitaient] | [pourrait] | [chercherait] | [aura] ];

define objectPronouns [le | la] -> [it], [les | leur] -> [them] || [.#. | " "] _ [.#. | " "] justVerbs [.#. | " "] .o. [[eux | elles] -> them , cela -> that];

define Infinitives
[rendre] -> "to render" , [combattre] -> "to fight" , [être] -> "to be" , [concurrencer] -> "to compete with" , [passer] -> "to pass/happen" , [suivre] -> "to pursue" , [rester] -> "to rest" , [faire] -> "to do/make" , [trouver] -> "to find" ,  [parvenir] -> "to reach" || [.#. | " "] _ [.#. | " "];

define ParticipleEndings [0 | e | s | "es"];
define Participles [réitéré [ParticipleEndings] ] -> [reiterated] , [enterré [ParticipleEndings] ] -> [buried] , [été [ParticipleEndings] ] -> [been] , [retrouvé [ParticipleEndings] ] -> [recovered] , [assassiné [ParticipleEndings] ] -> [assassinated] , [pris [ParticipleEndings] ] -> [taken] , [détourné [ParticipleEndings] ] -> [detained] , [annoncé [ParticipleEndings] ] -> [announced] , [remis [ParticipleEndings] ] -> [returned] , [indiqué [ParticipleEndings] ] -> [indicated] , [trouvé [ParticipleEndings] ] -> [found] , [envolé  [ParticipleEndings] ] -> [flown] , [opéré [ParticipleEndings] ] -> [operated "-" on] , [compliqué [ParticipleEndings] ] -> [complicated] ,
```

[redevenu [ParticipleEndings] " " normal] -> ["back-to-normal"] , [expliqué [ParticipleEndings]] -> [explained] , [fermé [ParticipleEndings]] -> [closed] , [épris [ParticipleEndings]] -> ["in-love with"] , [enragé [ParticipleEndings]] -> [enraged] , [fait " " pencher] -> [tipped] , [fait [ParticipleEndings]] -> ["done/made"] , [interdisant [ParticipleEndings]] -> [forbidding] || [.#. | " "] _ [.#. | " "];

define ConjugatedVerbs
[a] -> "has" , [ont] -> "have" , [interpellent] -> "address" , [confirmons] -> "confirm" , [emparent] -> "take-possession" , [consacre] -> "blesses" , [incarne] -> "embodies" , [peut | peuvent] -> "can" , [va] -> "is going to" , [est] -> "is" , [doit] -> "must" , [affecte] -> "affect" , [gère] -> "deal-with" , [avait | avaient] -> "had" , [stabilisait] -> "stabilized" , [faisait] -> "did/made" , [profitaient] -> "profited" , [pourrait] -> "could" , [chercherait] -> "will search for" , [aura] -> "will have" , [était] -> [was] || [.#. | " "] _ [.#. | " "];

define Verbs objectPronouns .o. Infinitives .o. Participles .o. ConjugatedVerbs;

#individual word translations - nouns
define subjectPronouns [elle] -> ["she/her/it"] , [il] -> ["he/it"] , [nous] -> ["we/us"] , [on] -> ["one/you"] , [ils] -> [they] || [.#. | " "] _ [" " | .#.];

define properNouns
[onu] -> ["United-Nations"] , [la " " france] -> [France] , [youssouf " " bakayoko] -> ["Youssouf-Bakayoko"] ,
[avril] -> [April] , [ciudad " " juarez] -> ["Ciudad-Juarez"] , [aly " " zoulfecar] -> ["Aly-Zoulfecar"] , [dar " " es " " salam] -> ["Dar-Es-Salam"] , [union " " européenne] -> ["European-Union"] , [tanzanie] -> [Tanzania] , [atalante] -> ["Operation-Atalanta"] , [sagrada " " familia] -> ["Sagrada-Familia"] , [liu " " xiaobo] -> ["Liu-Xiaobo"] , [chine] -> [China] , [bogdan " " klich] -> ["Bogdan-Klich"] , [varsovie] -> [Warsaw] , [new " " york] -> ["New-York"] , [novembre] -> [November] , [cristina " " fernandez " " de " " kirchner] -> ["Cristina-Fernandez-de-Kirchner"] , [m " . " deighton] -> "Mr.-Deighton" || [.#. | " "] _ [.#. | " "] .o. [mercredi] -> [Wednesday] , [samedi] -> [Saturday] , [vendredi] -> [Friday] , [jeudi] -> [Thursday] , [washington] -> [Washington] , [oslo] -> [Oslo] , [f " " 16] -> ["F-16"] , [hercules] -> [Hercules] , [marcelo] -> [Marcelo] , [abdallah] -> [Abdullah] , ["m."] -> ["Mr."] , [deighton] -> [Deighton] , [wikileaks] -> [WikiLeaks] , [cia] -> [CIA] || [.#. | " "] _;

define normalNouns
[candidat] -> [candidate] , [voix] -> [voice] , [président [0 | e]] -> [president] , [soirée] -> [evening] , [résultat] -> [result] ,
[corps] -> [bodies] , [hommes] -> [men] , [homme] -> [man] , [femmes] -> [women/wives] , [femme] -> ["woman/wife"] , [victime] -> [victim] ,
[engagement] -> [commitment] , [terrorisme] -> [terrorism] , [dirigeant [0 | e]] -> [leader] , [pays] -> ["nation(s)"] ,
[monnaies] -> [currencies] , [monnaie] -> [currency] , [yen [0 | s]] -> [yen] , [soir] -> [night] , [mort] -> [death] ,
[citoyen] -> [citizen] , [année] -> [year] , [rivalité] -> [rivalry] , [ville] -> ["town/city"] , [1er] -> [first] ,
[habitants] -> [inhabitants] , [nombre] -> [number] , [navire] -> [ship] , [membre] -> [member] , [équipage] -> [equipment] ,

[passager] -> [passenger] , [route] -> ["route/way"] , [pape] -> [Pope] , [symbole] -> [symbol] ,
[famille] -> [family] ,
[métaux] -> [metals] , [métal] -> [metal] , [affaiblissement] -> [weakening] , [façon] -> [fashion] ,
[prix] -> [prize] , [paix] -> [peace] ,
[yeux] -> [eyes] , [œil] -> [eye] , [prison] -> ["prison/imprisonment"] , [lutte] -> [struggle] ,
[concession] -> [compromise] ,
[idéaux] -> [ideals] , [idéal] -> [ideal] , [ministre] -> [minister] , [défense] -> [defense] , [accord] ->
[agreement] ,
[stationnement] -> [parking] , [avion] -> [plane] , [appareil [0 | s]] -> [aircraft] , [semestre] -> ["half-of-
the-year"] ,
[latéral [0 | s] " " gauche] -> left-side , [brésilien [0 | ne]] -> [Brazilian] , [roi] -> [king] , [hernie [0 | s] "
" discale] -> ["slipped-disc"] , [hématome] -> [hematoma] , [trafic] -> [traffic] , [sécurités] ->
[securities] , [sécurité] -> [security] , [rue] -> [road] ,
[poste] -> [post] , [accès] -> [access] , [journaliste] -> [journalist] , [féministe] -> [feminist] ,
[amoureuse | amoreux] -> [lover] ,
[musulman [e | 0]] -> [Muslim] , [fondamentaliste] -> [fundamentalist] , [lesbienne] -> [lesbian] ,
[effet] -> [effect] ,
[matière] -> [matters] , [diplomatie] -> [diplomacy] , [transparence] -> [transparency] , [liberté] ->
[freedom] , [état] -> [state] ,
[cours] -> [courses] , [conseiller] -> [counselor] , [prises " " de " " décision] -> ["decision-making"] ,
[nerf] -> [nerve] ,
[anxiété] -> [anxiety] , [faveur] -> [favor] , [milliard] -> [billion], [monde] -> [world] , [presse] ->
[press] , [jour] -> [day] , [ajustement] -> [adjustment] , [la " " maison] -> [home], [intellectuel] ->
[intellectual] || [.#. | " "] _ ;

define Nouns subjectPronouns .o. properNouns .o. normalNouns;

#individual word translations - adjectives
define AdjectiveEndings [e | s | "es" | "ne" | "nes" | "le" | "les" | 0];

define Adjectives
[distinct [AdjectiveEndings]]-> [distinct] , [nucléaire [AdjectiveEndings]]-> [nuclear] , [nippon
[AdjectiveEndings]] -> [Japanese] , [américain [AdjectiveEndings]] -> [American] , [croissant
[AdjectiveEndings]] -> [increasing] , [comorien [AdjectiveEndings]] -> [Comorian] , [somalien
[AdjectiveEndings]] -> [Somalian] , [[neuf | neuve] [AdjectiveEndings]]-> new , [naval
[AdjectiveEndings]] -> [naval] , [antipiraterie [AdjectiveEndings]] -> [antipiracy] , [industriel
[AdjectiveEndings]] -> [industrial] , [symbolique [AdjectiveEndings]] -> [symbolic] , [nobel
[AdjectiveEndings]] -> [Nobel] , [démocratique [AdjectiveEndings]]-> [democratic] , [polonais
[AdjectiveEndings]] -> [Polish] , [temporaire [AdjectiveEndings]] -> [temporary] , [espagnol
[AdjectiveEndings]] -> [Spanish] , [madrilène [AdjectiveEndings]] -> [Madridian] , [normal
[AdjectiveEndings]] -> [normal] , [proche [AdjectiveEndings]]-> [close] , [chrétien
[AdjectiveEndings]] -> [Christian] , [[tous | tout] [AdjectiveEndings]] -> [all] , [même
[AdjectiveEndings]] -> [same] , [publique [AdjectiveEndings]] -> [public] , [viable
[AdjectiveEndings]] -> [sustainable] , [malade [AdjectiveEndings]] -> [sick] , [seul
[AdjectiveEndings]] -> [alone] , [traditionnel [AdjectiveEndings]] -> [traditional] , [deux] -> [two] ,
[nécessaire [AdjectiveEndings]] -> [necessary] || [.#. | " "] _ [.#. | " "];

#individual word translations - adverbs

```
define Adverbs [apparemment] -> [apparently] , [déjà] -> [already] , [aussi] -> [also] , [désormais] -> [henceforth] , [également] -> [equally] , [quasiment] -> [almost] , [mortellement] -> [mortally] , [après] -> [after] , [gravement] -> [seriously] , [longtemps] -> ["a long time"] || [.#. | " "] _ [.#. | " "];


#the rest: determiners, complementizers, WH-words, prepositions, conjunctions
define finishing [ le | la | les ] -> [the] , [sa | son | ses] -> ["his/her/its"] , [leur | leurs] -> [their] , [un | une] -> ["a/an"] , [quelque] -> [some] , [ce | cette] -> [that] , [ces] -> ["these/those"] , [million " " de] -> [million] , [votre | vos] -> [your] , [par] -> [by] , [de] -> ["of/from"] , [dans] -> [in] , [jusque] -> [until] , [à] -> ["to/at"] , [pour] -> ["for/to"] , [en] -> ["in/to"] , [contre] -> [against] , [avec] -> [with] , [vers] -> [towards] , [sans] -> [without] , [sur] -> [on] , [entre] -> [between] , [depuis] -> [since] , [que] -> ["what/that/than"] , [qui] -> ["who/whom/that"] , [lorsque] -> [when] , [pourquoi] -> [why] , [où] -> [where] , [comment] -> [how] , [et] -> [and] , [ou] -> [or] , [mais] -> [but] || [.#. | " "] _ [.#. | " "];

define prepositions [by | "of/from" | in | until | "to/at" | "for/to" | "in/to" | against | with | towards | without | on | between | since | of | from | to | at | for];


#These are the functions directly called to translate sentences, in order
define preprocess toLowercase .o. removeContractions .o. punctuation1 .o. punctuation2 .o. punctuation3;
define replace Expressions .o. Verbs .o. Nouns .o. Adjectives .o. Adverbs .o. finishing;
define post [prepositions] -> 0 || [.#. | " " ] [prepositions | regarding "/" opposite] " " _ " " .o. " " -> 0 || " "_ .o. "the" -> "on" || _ " " [Sunday | Monday | Tuesday | Wednesday | Thursday | Friday | Saturday] .o. [to " "] -> 0 || [could | must | she "/" her "/" it | he "/" it | we "/" us | one "/" you | they | it] " " _ .o. ["reach to/at"] -> [achieve] || [.#. | " "] _ [.#. | " "];
```

APPENDIX C:

```
import java.util.*;

public class Movement {
	static Set<String> Pronouns, Nouns, AuxiliaryVerbs, Verbs, Adjectives, Adverbs, Determiners, WH_words;

	public static void main(String[] args) {
		//creating sets for use
		String[] pronouns = {"she/her/it", "we/us", "he/it", "one/you", "all", "they", "it"};
		String[] properNouns = {"United-Nations", "France", "Youssouf-Bakayoko", "Wednesday", "Saturday",
		"Friday", "Cuidad-Juarez", "Aly-Zoulfecar", "Dar-Es-Salam", "Tanzania", "European-Union",
		"April", "Washington", "Operation-Atalanta", "Pope", "Sagrada-Familia", "Liu-Xiaobo", "Oslo", "China",
		"Bogdan-Klich", "Thursday", "F-16", "Hercules", "Warsaw", "Brazilian", "Marcelo", "Abdullah", "New-York",
		"November", "Mr.", "Deighton", "CIA", "Muslims", "WikiLeaks", "Cristina-Fernandez-de-Kirchner"}
		String[] singularNouns = {"candidate", "voice", "president", "commission", "evening", "result", "body", "man",
		"woman/wife", "victim", "date", "leader", "terrorism", "commitment", "euro", "currency", "yen", "night",
		"death", "citizen", "rivalry", "year", "town/city", "million", "inhabitant", "pirate", "number", "ship", "member",
		"passenger", "route/way", "force", "symbol", "family", "metal", "weakening", "dollar", "intellectual", "fashion",
		"prize", "struggle", "compromise", "ideal", "peace", "eye", "world", "prison/imprisonment", "minister",
		"defense", "agreement", "plane", "transport", "half-of-the-year", "club", "left", "king", "day", "slipped-disc",
		"hematoma", "traffic", "adjustment", "security", "road", "post", "access", "journalist", "agent", "feminist",
		"lover", "fundamentalist", "lesbian", "effect", "diplomacy", "transparency", "freedom", "expression", "time",
		"home", "anxiety", "balance", "favor", "stress", "attitude"};
		String[] pluralNouns = {"candidates", "voices", "presidents", "commissions", "evenings", "results", "bodies",
		"men",  "victims", "dates", "leaders", "nation(s)", "commitments", "euros", "currencies", "nights", "deaths",
		"citizens", "rivalries", "years", "towns/cities", "millions", "inhabitants", "pirates", "numbers", "ships",
		"members", "equipment", "passengers", "forces", "symbols", "metals", "dollars", "intellectuals", "prizes",
		"ideals", "eyes", "ministers", "agreements", "parking", "planes", "combat", "aircraft", "press", "clubs", "days",
		"adjustments", "roads", "posts", "police", "journalists", "feminists", "matters", "states", "finances", "courses",
		"nerves", "billions", "counselors", "decision-making"};
		Pronouns = new HashSet<String>();
		Nouns = new HashSet<String>();
		for(String s : pronouns) Pronouns.add(s);
		for(String s : properNouns) Nouns.add(s);
		for(String s : singularNouns) Nouns.add(s);
		for(String s : pluralNouns) Nouns.add(s);


		String[] auxiliaryVerbs = {"has", "had", "have", "be", "is", "could", "will", "was", "can", "must"};
		String[] normalVerbs = {"render", "fight", "compete", "pass/happen", "pursue", "rest", "do/make", "find",
		"stabilized", "reach", "address", "confirm", "take-possession", "blesses", "profited", "embodies", "search",
		"close", "deal-with", "achieve", "affect"};
		String[] participles = {"reiterated", "buried", "been", "recovered", "assassinated", "taken", "done/made",
		"detained", "did/made", "announced", "returned", "indicated", "found", "flown", "operated-on",
		"complicated", "explained", "closed", "enraged", "tipped", "increasing", "forbidding", "going"};
		String[] adjectives = {"distinct", "nuclear", "Japanese", "American", "more", "Comorian", "Somalian", "new",
		"naval", "antipiracy", "traditional", "industrial", "symbolic", "Nobel", "democratic", "Polish", "temporary",
		"first", "Spanish", "Madridian", "two", "back-to-normal", "necessary", "Christian", "in-love", "at-the-same-
		time",  "same", "public", "sustainable", "sick", "long", "alone", "back-to-normal"};
		AuxiliaryVerbs = new HashSet<String>();
		Verbs = new HashSet<String>();
		Adjectives = new HashSet<String>();
		for(String s: auxiliaryVerbs){
			AuxiliaryVerbs.add(s); Verbs.add(s);
		}for(String s: normalVerbs) Verbs.add(s);
		for(String s: participles){
			Verbs.add(s); Adjectives.add(s);
		}for(String s: adjectives) Adjectives.add(s);


		String[] adverbsBefore = {"also", "henceforth", "equally", "almost", "mortally", "seriously"};
		Adverbs = new HashSet<String>();
		for(String s : adverbsBefore) Adverbs.add(s);
```

```
		String[] determiners = {"the", "his/her/its", "a/an", "some", "their", "that", "a", "your", "these/those"};
		Determiners = new HashSet<String>();
		for(String s : determiners) Determiners.add(s);
		String[] wh_words = {"what/that/than", "how", "where"};
		WH_words = new HashSet<String>();
		for(String s : wh_words) WH_words.add(s);

		//taking in user input and separating it into individual words
		ArrayList<Integer> indexes = new ArrayList<Integer>();
		ArrayList<String> words = new ArrayList<String>();
		Scanner reader = new Scanner(System.in);
		System.out.println("Enter a sentence:");
		String sentence = " " + reader.nextLine() + " ";
		for(int i = 0 ; i < sentence.length() ; i++){
			if(sentence.charAt(i) == ' ')
				indexes.add(i);
		}
		for(int i = 0 ; i < indexes.size() - 1; i++){
			words.add(sentence.substring(
		}
		adjectiveMovement(words);
		adverbMovement(words);
		objectPronounMovement(words);
		if(sentence.contains("?")){
			pronounDeletion(words);
			auxVerbMovement(words);
			doSupport(words);
		}
		System.out.println(postProcess(words));
}
	//fixes up the sentence to make it presentable
	static String postProcess(ArrayList<String> strings){
		String output = "";
		output += strings.get(0).substring(0, 1).toUpperCase() + strings.get(0).substring(1);
		for(int i = 1; i < strings.size(); i++){
			if(strings.get(i).equals(",") || strings.get(i).equals(".") || strings.get(i).equals("?"))
				output += strings.get(i);
			else output += " " + strings.get(i);
		}
		return output.replace('-', ' ');
	}

	//swaps the positions of two words in the sentence
	static void swap(ArrayList<String> strings, int num1, int num2){
		String s = strings.get(num1);
		strings.set(num1, strings.get(num2));
		strings.set(num2, s);
	}

	//moves adjective to front of noun and makes the resulting adjective phrase become single entity
	static void adjectiveMovement(ArrayList<String> strings){
		int index = 0;
		while(index < strings.size()){
			if(index > 0 && Adjectives.contains(strings.get(index)) && Nouns.contains(strings.get(index - 1))){
				if(index == strings.size() - 1 || !(strings.get(index + 1).equals("by"))){
					strings.set(index - 1, strings.get(index) + " " + strings.get(index - 1));
					strings.remove(index);
					Nouns.add(strings.get(index - 1));
					index = 0;
				}
			}index++;
		}
	}

	//moves adverbs to the front of verbs, leaves those in front of adjectives as-is
```

```
static void adverbMovement(ArrayList<String> strings){
	int adverbIndex = 0, verbIndex = 0;
	while(adverbIndex < strings.size()){
		if(Adverbs.contains(strings.get(adverbIndex))){
			if(adverbIndex == strings.size() - 1 || !(Adjectives.contains(strings.get(adverbIndex + 1)))){
				verbIndex = adverbIndex - 1;
				while(adverbIndex != 0 && verbIndex >= 0){
					if(Verbs.contains(strings.get(verbIndex))){
				strings.set(verbIndex, strings.get(adverbIndex) + " " + strings.get(verbIndex));
						strings.remove(adverbIndex);
						Nouns.add(strings.get(verbIndex));
						adverbIndex = 0;
					}verbIndex --;
				}
			}
		}adverbIndex++;
	}
}

//object pronouns behind a verb get moved to be after that verb
static void objectPronounMovement(ArrayList<String> strings){
	int index = 0;
	while(index < strings.size() - 1){
		if(index > 0 && Pronouns.contains(strings.get(index)) && Verbs.contains(strings.get(index + 1))
			&& (Pronouns.contains(strings.get(index - 1)) || Nouns.contains(strings.get(index - 1)))){
			swap(strings, index, index + 1);
		}index++;
	}
}


//these functions will be used only with questions
//in a question, removes placemarker pronouns that occur immediately after moved verbs
static void pronounDeletion(ArrayList<String> strings){
	int nounIndex = 0, pronounIndex = 0;
	while(nounIndex < strings.size() - 1){
		if(Nouns.contains(strings.get(nounIndex)) && Verbs.contains(strings.get(nounIndex + 1))){
			pronounIndex = nounIndex + 1;
			while(pronounIndex < strings.size() - 1){
				if(Pronouns.contains(strings.get(pronounIndex)) &&
					Verbs.contains(strings.get(pronounIndex - 1))){
					strings.remove(pronounIndex);
					return;
				}pronounIndex ++;
			}return;
		}nounIndex ++;
	}
}

//when necessary, moves the auxiliary verb in a question to the front of the question
static void auxVerbMovement(ArrayList<String> strings){
	int index = 0;
	while(index < strings.size()){
		if(index > 0 && AuxiliaryVerbs.contains(strings.get(index)) &&
			(Pronouns.contains(strings.get(index - 1)) || Nouns.contains(strings.get(index - 1)))){
			String verb = strings.get(index);
			strings.remove(index);
			strings.add(0, verb);
			return;
		}
		if(Verbs.contains(strings.get(index))){
			return;
		}index ++;
	}
}
```

```
	//adds do-support after WH-words in questions
	static void doSupport(ArrayList<String> strings){
		int index = 0;
		while(index < strings.size() - 1){
			if(WH_words.contains(strings.get(index)) && (Nouns.contains(strings.get(index + 1))
			|| Pronouns.contains(strings.get(index + 1)) || Determiners.contains(strings.get(index + 1)))){
				strings.add(index + 1, "do/does");
				return;
			}
			if(WH_words.contains(strings.get(index)) && strings.get(index + 1).equals("to")){
				strings.set(index + 1, "do you");
				return;
			}
			if(WH_words.contains(strings.get(index))) return;
			index ++;
		}
	}
}
```