# Statut 2

Ce qui est fait :

* Première version de l’application
* Ajout d’un nettoyage de données par stemming
* Amélioration du nettoyage de données (ponctuation)
* Ajout d’un apprentissage continue du programme par l’ajout de propositions des utilisateurs
* Ajout d’un menu basique.

Les difficultés :

* Expressions et mots par émotion insuffisante (je n’ai pas avancé sur ce sujet depuis le statut 1)
* La régression logistique ne convient pas pour mon traitement de mes données. Chaque mot ou expression à un ID qui est envoyés dans la régression logistique. Mais l’algorithme fait une sorte de rapprochement de chaque numéro qui peut fausser le résultat final, exemple :
  + 1 : mot1 : cat1 ; 2 : mot2: cat1 ; 3 : mot3: cat1 ; 4 : mot4: cat1 ; 5 : mot5: cat2 ; 6 : mot6: cat2 ; 7 : mot7 : cat2, 8 : mot8 : cat1,
  + Le mot4 pourrait être de catégorie 2 au lieu de 1 dû à la proximité du nombre (id) 5, 6 et 7
  + Le mot8 sera forcément de catégorie 2 alors qu’il est de 1, car selon son integer (id) il est plus proche de 7, 6, 5 que de 1, 2,3, 4
  + Existe-t-il une régression logistique inverse ? au lieu de prédire X, il prédit Y.

Les solutions :

* Trouver d’autres listes de mots par émotion
* Essayer de faire la reconnaissance des émotions en anglais (moins de difficulté au niveau du vocabulaire et des genres, peut-être plus facile de trouver des apis, etc.)
* Recherche de nouveaux algorithmes.

La prochaine fois :

* Essayer de trouver des listes de mots par émotion en anglais et français
* Faire un algorithme pour l’anglais (si les autres problèmes réglés)
* Recherche d’un algorithme plus adéquate à mes besoins (travail prioritaire)