

Sujet : Traduction de la parole

Contacts : H  l  ne Maynard, Alexandre Allauzen
(prenom.nom@limsi.fr)
01 69 85 80 49 - 06 70 81 15 93

11 avril 2013

La traduction de la parole est un des d  fis soci  taux majeurs dont l'objectif est de permettre la communication parl  e entre personnes ne parlant pas la m  me langue. Le d  fi est   galement scientifique puisqu'il se situe    la jonction de deux domaines de recherche que sont la reconnaissance automatique de la parole (ASR) et la traduction automatique (MT). Si ces deux domaines ont connu dans les derni  res d  cennies des progr  s consid  rables gr  ce    l'introduction de mod  les statistiques, il n'existe pas encore de solution satisfaisante pour leur couplage.

Le sujet de th  se propos   vise    imaginer et mettre en oeuvre des strat  gies permettant de mieux faire interagir reconnaissance de la parole et traduction automatique statistique.

Dans la situation classique, les mod  les statistiques utilis  s par les deux syst  mes - ASR et MT - sont appris ind  pendamment,    partir de donn  es de diff  rentes natures. Les mod  les de reconnaissance de la parole sont entra  n  s sur des donn  es mono-lingues (transcriptions manuelles de donn  es de parole), alors que les mod  les de traduction utilisent des donn  es multi-lingues (corpus parall  les) qui g  n  ralement sont issus de donn  es purement textuelles (langage   crit). Ainsi les corpus utilis  s en MT se distinguent de diff  rentes mani  res des corpus en ASR : ce sont rarement des transcriptions de parole et lorsque c'est le cas il s'agit toujours transcriptions exactes, et dans lesquelles les disfluences qui caract  risent l'oral sont gomm  es. De m  me, les mod  les de traduction sont entra  n  s sur des corpus parall  les qui supposent une segmentation en unit  s correspondant classiquement    des phrases (une quarantaine de mots au plus), alors que les donn  es orales se pr  sentent comme des flux de mots, s  par  s   ventuellement par des trames de silence.

Le couplage le plus   l  mentaire pour traduire de la parole consiste    donner en entr  e du syst  me de traduction la s  quence de mots la plus probable issue du syst  me de reconnaissance de la parole, sans se soucier de l'absence d'ad  quation entre les mod  les qu'utilisent ces deux syst  mes.

Deux mani  res d'am  liorer ce couplage simpliste ont jusqu'   pr  sent   t   explor  es dans la litt  rature :

- permettre au module ASR de produire plusieurs hypoth  ses (n-meilleures solutions, treillis, r  seaux de confusions) laissant ainsi au syst  me de traduction la possibilit   de choisir la meilleure hypoth  se pour produire la traduction la plus probable.

- modifier la partie source des corpus d’entraînement des modèles de traduction pour la rendre compatible avec la sortie - supposée exacte - produite par le système de reconnaissance de la parole.

Dans cette thèse nous proposons d’explorer plus généralement et en profondeur les possibilités de couplage, qui peuvent être :

- Adapter les modèles de traduction aux sorties ASR :
 - entraîner des modèles de traduction sur des données de parole (transcriptions exactes) et les combiner avec des modèles entraînés sur des données textuelles,
 - mais aussi étudier la possibilité d’y inclure des transcriptions automatiques, de façon à intégrer la connaissance des erreurs produites par le système ASR dans le modèle de traduction,
 - prendre en compte le fait que le système de reconnaissance de la parole utilise un lexique (vocabulaire) limité.
- Adapter les modèles ou les sorties ASR pour les rendre compatibles avec les entrées attendues par le modèle de traduction :
 - explorer les interfaces par treillis ou réseaux de consensus,
 - transformer les transcriptions automatiques pour les rendre plus proches des transcriptions exactes attendues dans le modèle MT.

Plus généralement la modélisation de l’oral dans les systèmes de traduction pose des questions fondamentales, renvoyant notamment à la question des unités et de la segmentation, mais aussi à la nature mêmes des modèles de traduction qui, contrairement aux modèles de langues bien adaptés aux phénomènes locaux présents dans la parole spontanées, doivent intégrer des dépendances à long terme.

Enfin dans l’idéal, les modèles utilisés par les deux systèmes devraient être optimisés de manière conjointe. Etant donnée la taille des modèles utilisés, ceci constitue un défi scientifique et pratique considérable. Pour cet axe de recherche, il est possible d’explorer des représentations unifiées comme les transducteurs à états finis afin de poser le problème dans un cadre théorique commun.

Afin d’appréhender tous les enjeux et toutes les problématiques soulevés par la traduction de la parole, il est proposé d’analyser les erreurs qui peuvent être produites par un couplage simple afin d’évaluer l’importance des différents problèmes cités précédemment.

Le (la) doctorant(e) sera intégré(e) au groupe Traitement du Langage Parlé (TLP) du LIMSI-CNRS, qui réalise des recherches au plus haut niveau dans les deux domaines ASR et MT, et plus généralement en apprentissage statistique.

Il (elle) bénéficiera d’un environnement dynamisé par de nombreux projets lui permettant de mettre en place des expériences sur de larges corpus, dans domaines variés (sous-titrage et traduction de bulletins d’informations, de séries télévisées et films, de conférences et cours, et dialogues multilingues) et pour différents couples de langues.