BabelDr: A Web Platform for Rapid Construction of Phrasebook-Style Medical Speech Translation Applications

BOUILLON, Pierrette, et al.
BabelDr: A Web Platform for Rapid Construction of Phrasebook-Style Medical Speech Translation Applications

Pierrette Bouillon (FTI), Hervé Spechbach (HUG), Sophie Durieux-Paillard (HUG), Johanna Gerlach (FTI), Sonia Halimi (FTI), Patricia Hudelson (HUG), Manny Rayner (FTI), Irène Strasly (FTI), Nikos Tsourakis (FTI)

Context

52% of patients at HUG (Geneva University Hospitals) are foreign; 12% don’t speak French

Google translate is not a solution:
- not reliable and precise enough
- most important languages for HUG not covered (Tigrinya, Romani, etc.)

Solution: BabelDr
- Common project UNIGE/FTI-HUG
- a web-based platform for building medium-size speech translators in medical domains
- Reliable, easy to add new languages, easy to deploy

Financed by: «Fondation privée des Hôpitaux universitaires de Genève» and Innogap/Mimosa grants

The application

Examination-tool for medical subdomains: semantic coverage consists of a prespecified set of questions, but users can use a wide variety of surface forms when speaking to the system

Translation steps:
- Rule-based recognition of the source question (Nuance)
  Est-ce que vous avez de la température ?
- Rule-based translation of the source to source canonical with synchronized CFG
  Avez-vous de la fièvre ?
- Human translation from source canonical to target sentence

Data will be used to build a robust recognizer and a classifier

Development platform and steps

Developed with Regulus Lite
- The set of canonical questions is defined by HUG
- The questions are transformed in synchronized CFG by experts (UMLS, etc.)

| Utterance | Source $avez_vous ? $souffrez_vous de $ressentez_vous ? $irradient (ce que|qu')) $votre_douleur $irradie_elle $vers_organe ? | Target/french la douleur au ventre irradie $vers_organe ? |
|-----------|---------------------------------------------------------------------------------------------------------------|-----------------------------------------------|

The canonical questions are translated by translators

<table>
<thead>
<tr>
<th>Utterance</th>
<th>Source ?$vos_douleurs $irradient $vers_organe ?</th>
</tr>
</thead>
</table>

The system is compiled on the web

Content

- 3000 questions collected for now
- 6 domains
- 4 languages in development
- Size of the 6 domains:

<table>
<thead>
<tr>
<th>Domain</th>
<th>Words</th>
<th>Expansions</th>
<th>Number of Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>abdominal</td>
<td>11642</td>
<td>8’000’000</td>
<td>620</td>
</tr>
<tr>
<td>poitrine</td>
<td>1572</td>
<td>6’000’000</td>
<td>551</td>
</tr>
<tr>
<td>tête</td>
<td>1734</td>
<td>16’000’000</td>
<td>604</td>
</tr>
<tr>
<td>accueil</td>
<td>684</td>
<td>2’277</td>
<td>110</td>
</tr>
<tr>
<td>suivi</td>
<td>881</td>
<td>236’000</td>
<td>217</td>
</tr>
<tr>
<td>habitudes</td>
<td>684</td>
<td>2’514</td>
<td>66</td>
</tr>
</tbody>
</table>