

1. Descripción del proyecto final (a realizarse en grupos de hasta 2 personas)

Cada grupo deberá armar un sistema de TA en las siguientes etapas:

1. Armar un sistema baseline → sys1
2. Traducir el corpus de test
3. Optimizar parámetros con mert → sys2
4. Traducir el corpus de test
5. Mejorar el sistema de acuerdo a lo propuesto mas abajo o con alguna estrategia propuesta por ustedes mismos (en este caso consultarlo conmigo antes de implementarlo) → sys3
6. Evaluar sys1, sys2 y sys3 usando métricas automáticas y humanas
7. Analizar los errores producidos en las traducciones, reflexionar sobre su origen y proponer posibles soluciones o mejoras al sistema para corregir estos errores

2. Archivos a entregar

1. Textos paralelos utilizados, salida de cada sistema, resultado de las evaluaciones para cada sistema
2. Un informe con al menos: una sección donde se documenta el proceso detallado en el punto 1 (qué herramientas usaron, qué mejoras hicieron, etc), una sección de discusión sobre los resultados obtenidos, su correlación con la evaluación realizada, y el análisis de los errores encontrados y una sección donde se propongan algunas soluciones a los errores/problemas encontrados y/o formas de mejorar el sistema.

3. Recursos a utilizar

Se puede elegir entre estos dos paquetes de traducción

- Cunei (TA hibrida basada en ejemplos y estadística) - <http://www.cunei.org/doc/http://www.cunei.org/about/phillips-ebmt09.pdf>
- Moses (TA estadística) - <http://www.statmt.org/moses/>

Para el armado de sys3, dependiendo del paquete elegido en el punto anterior, las mejoras propuestas son las siguientes pero se aceptan otras ideas:

- Cunei: implementar anotaciones semánticas (ej nombres de personas, ubicaciones, ...) Ver <http://www.cs.jhu.edu/~ccb/publications/scale-2009-report.pdf> cap 3
- Moses: entrenar un modelo factorizado con al menos un factor adicional a la “surface form” (ej POS tags o lemmas) Ver <http://www.statmt.org/moses/?n=Moses.FactoredTutorial>

Textos paralelos

- Se pueden usar cualquiera de los corpora disponibles en <http://urd.let.rug.nl/tiedeman/OPUS/index.php> o en <http://wt.jrc.it/lt/Acquis/> o cualquiera de los conjuntos compilados para los workshops desde 2005 <http://www.statmt.org/>
- Utilizar una cantidad considerable de texto de entrenamiento y desarrollo, para evaluación pueden traducir muchas frases (ej 500) y evaluar con humanos una parte representativa (digamos unas 50+)

Métricas (usar al menos una humana y una automática, mientras más se apliquen mejor)

- Automáticas: Word Error Rate, Position-independent ER, General Text Matcher <http://nlp.cs.nyu.edu/GTM/gtm-1.4-source.tgz>, Translation ER, BLEU/NIST <http://kheafield.com/code/scoring.tar.gz>
- Humanas: pueden hacer una evaluación comparativa, evaluar alguna aspecto en particular de la traducción (gramaticalidad, fluidez, adecuación, traducción de nombres propios, etc) o proponer su propia métrica (esta opción es fuertemente recomendada)