



## NATURAL LANGUAGE – MP2 – GROUP 4

### GROUP IDENTIFICATION AND CONTRIBUTION'S ESTIMATE:

Daniel Pereira – 89425 – 50%	Tiago Barroso – 89549 – 50%
------------------------------	-----------------------------

In this project we worked at the same time together, so each member had an estimated contribution of 50%.

### OPTIONS TAKEN:

We concluded it was necessary for two extra transducers to be created:

- **dois\_pontos** - Given a word “e” in the input, puts “:” in the output. This is necessary for the transducers **text2num** and **rich2text**.
- **preencher\_minutos** - Simply puts “:00” in the output. This is necessary for the transducer **lazy2num**.

For the transducers in point 2, we used operations from the OpenFst library with the basic transducers from point 1. Two of the transducers from point 2 used an extra transducer. Here is how each transducer is made:

- **text2num** - Concatenates **horas** with **dois\_pontos**, then concatenates the result with **minutos**.
- **lazy2num** - Union of **text2num** with the concatenation of **horas** with **preencher\_minutos**.
- **rich2text** - Concatenates the projections of **horas** and **dois\_pontos**, then concatenates the result with the union of **quartos** and **meias**.
- **rich2num** - In this case, we chose to do a composition of transducers **rich2text** and **lazy2num**, this allows for expressions like “*duas e meia*” to be converted to “*hh:mm*”. We then use the operation union with **lazy2num** and the result of the previous composition, allowing for all types of expressions to be transformed to “*hh:mm*”
- **num2text** - For this transducer, all it was necessary was an inversion of **text2num**, as the name suggests. The use of weights makes it so that the output always has the words “*hora(s)*” e “*minuto(s)*”

### COMMENTS ON THE SOLUTION:

- ➔ We opted to recognize the written version of numbers, such as “*vinte e cinco*”, instead of the spoken one, respectively “*vinte cinco*”.
- ➔ Our solution in cases where there is ambiguity, for example in the transducer **lazy2num** with the input “*vinte e tres*”, chooses the output “*23:00*” instead of “*20:03*”.