Executive summary

Deliverable D2.1 reports on the work carried out within tasks *T2.1: System design & architecture* and *T2.2: Definition of the system modules* of **WP2: System specifications**. It illustrates the PRESEMT system architecture and provides a detailed account of the constituent modules in terms of their mode of operation, the input they receive and the output they yield, their interfaces and the workflow between them.

The PRESEMT system will have a tripartite structure:

1. **Pre-processing stage**, which involves the compilation of resources needed for the MT system to perform.
2. **Main translation engine**, which constitutes the 'translation proper' component of the system.
3. **Post-processing stage**, which covers the user modifications to the system output and the subsequent adaptation of the system to this input.

Each stage encompasses a series of modules, as depicted in the following table:

Pre-processing stage: 4 modules	Main translation engine: 4 modules	Post-processing stage: 2 modules
Corpus creation & annotation module	Structure selection module	Post-processing module
Phrase aligner module	Translation equivalent selection module	
Phrasing model generator	Optimisation module 1	User adaptation module
Corpus modelling module	Optimisation module 2	

The **Corpus creation & annotation module** entails the compilation and annotation of monolingual and bilingual corpora to be utilised by the Main translation engine. The **Phrase aligner module**, operating on a bilingual corpus, performs text alignment at a phrasal level within a language pair, on the basis of which the **Phrasing model generator** elicits a phrasing model of the source language. The **Corpus modelling module** concerns the creation of a target language model involving semantic-type and statistical information extracted from a monolingual corpus.

The **Structure selection module** and the **Translation equivalent selection module** constitute the core part of the Main translation engine, being responsible respectively for defining the optimal structure and finding the best translation for a given source language text to be translated. These two modules are each enhanced by an **Optimisation module**, which optimises the values of the parameters they employ.

The **Post-processing module** provides the user with a GUI for modifying the system translation output; these modifications are then collected by the **User adaptation module**, so that the system can adapt to user preferences.

The deliverable has the following structure: Section 2 gives an account of requirements regarding machine translation. Section 3 provides a general overview of the system architecture by enumerating and briefly describing the system modules, while the subsequent sections 4 to 6 outline in detail the specifications per system module. Section 7 presents some first thoughts on parallelisation of processing. Section 8 summarises the PRESEMT system modules and their function, while Section 9 provides the guidelines for the linguistic resources.