Executive summary

Deliverable D4.1.1 reports on the work carried out in **WP4: Structure selection.** The objective of WP4 is "*to design and implement the first phase of the machine translation process. This involves defining the basic pattern of each sentence (in terms of structure) in the target language, by making use of the limited-size bilingual corpus to define correspondences*".

WP4 is divided into two tasks, *T4.1: Design and implementation of the structure selection module*, which "*focuses on developing a module which uses pattern recognition principles to determine for a given source language sentence the best-matching structure, by utilising the information contained in a small parallel corpus, and correspondingly transform it*" and *T4.2: Optimisation of module-specific parameters.*

In the PRESEMT architecture the Structure selection module is the only module that has direct access to bilingual resources, namely the aligned bilingual corpus and a bilingual dictionary. The key idea is to account for major structural differences between source (SL) and target (TL) languages based on the linguistic patterns found in the bilingual corpus. To this end, two methods are under investigation, based on dynamic programming and the use of a synchronous grammar respectively. Both aim at providing the appropriate TL structure for a given ISS in terms of phrase order.

A series of prototypes has been implemented that exemplify different variants of the Structure selection module.

The deliverable has the following structure: Section 2 provides a concise summary of the key design features of the PRESEMT approach with respect to the Structure selection module. Section 3 discusses the properties of the aligned bilingual corpus as the major linguistic resource for the module. Section 4 discusses dynamic programming as a method to exploit the contrastive linguistic knowledge encoded in the aligned bilingual corpus, while Section 5 presents synchronous grammars as an alternative methodology. Section 6 concludes the deliverable by outlining the parameters of the Structure selection module that are expected to undergo optimisation.