

A Comparative Evaluation of UNL Participant Relations using a Five-Language Parallel Corpus

Brian Murphy and Carl Vogel

Brian.Murphy@cs.tcd.ie*, Vogel@cs.tcd.ie

Department of Computer Science, University of Dublin, Trinity College

Abstract. In this paper we describe a manual case study in interlingual translation among five languages. Taking the UN Declaration of Human Rights in Chinese, English, German, Irish and Spanish, we annotated the five texts with a common interlingual logical form. We then studied four inventories of semantic roles (developed for both theoretical and NLP applications), including a subset of UNL's relations, and evaluated their suitability to describe the predicate-argument relationships found in the annotation. As a result, we make some suggestions for possible additions to the UNL relations, and propose that some of the existing relations be conflated or redefined.

1 Introduction

The work described here is part of a feasibility study on the use of semantic roles in interlingua-based machine translation. Our objective was to see if any set of semantic roles could give a description of verb-predicate relationships across a range of languages that would form an adequate basis for automatic generation.

The languages chosen were those that the authors have some working knowledge of (English, Chinese, German, Irish and Spanish), and include widespread and minority languages, both well and less-studied. The corpus used is the UN Declaration of Human Rights [1], a short text covering a broad range of topics in many languages (see Sect. 2).

From the literature on roles we selected four inventories (of which UNL's relations is one) that we considered to be well-enough developed for the annotation of unrestricted text. These inventories ([2,3,4,5] detailed in Sect. 4) were also chosen to be representative both theoretically and in terms of application to tasks such as machine translation and information retrieval.

After aligning the five language versions of the corpus, we manually annotated each article of the text with a language-neutral logical form (effectively a prototype interlingua) following the guidelines described in Sect. 3.1. The main part of the work then involved applying each of the role inventories in turn to the logical form and determining whether they satisfied three key criteria: coverage, differentiation and lack of ambiguity (Sect. 5). In other words, one should

* Supported by the TCD Senior Lecturer's Broad Curriculum Fellowship and Enterprise Ireland

© J. Cardeñosa, A. Gelbukh, E. Tovar (Eds.)

Universal Network Language: Advances in Theory and Applications.

Research on Computing Science 12, 2005, pp. 64–76.