Reconciling Parameterization, Configurability and Optimality in Natural Language Generation via Multiparadigm Programming

Jorge Marques Pelizzoni, Maria das Graças Volpe Nunes

Universidade de São Paulo, Instituto de Ciências Matemáticas e de Computação
Av. do Trabalhador São Carlense, 400. CEP 13560-970. São Carlos – SP – Brasil
`{jorgemp, gracan}@icmc.usp.br`

Abstract. This paper focuses on how multiparadigm – namely, constraint, object-oriented and higher-order – programming can be drawn upon not only to specify multiparameterized linguistic realization engines but also and above all to rationalize their configuration into full-fledged generation modules for specific language-application pairs. We describe Manati, one such engine whose instantiations render linguistic form to conceptual/semantic directed hypergraphs, and point out how its constraint-based concurrent architecture entails collaboration and interleaving so as to allow the definition and optimization of global quality measures.

1 Introduction

Natural Language Generation (NLG) refers to rendering linguistic form to input in a non-linguistic representation. As pointed out by e.g. Reiter & Dale [13], Cahill & Reape [2], Paiva [11], this can be a very complex task involving processing both linguistic (e.g. lexicalization, aggregation and referring expression generation) and otherwise (content selection and layout planning). In this paper, we are exclusively concerned with the linguistic aspect of generation, herein referred to as *linguistic realization*.

A range of linguistic realization work has been reported on so far in the literature varying in scope and depth. Nonetheless, it is a rare work that focuses on configurability issues, especially in a multiparameterization scenario. By *configurability* we mean ease of configuration, or rather, instantiation of required parameters in a disciplined, manageable, friendly manner. *Parameter*, in turn, refers to any blank whatsoever that should be filled in so as to make a generic solution into a full-fledged linguistic realization component. Possible parameters are grammars, lexicons, strategies, heuristics, etc.

In fact, there is far more usual to be material either detailing an isolated solution (an instantiation of a single parameter, e.g. Eddy [7]) or sketching a complex solution (i.e. a multiparameterized one, e.g. Stone & Doran [16]) – in either case, usually with not much regard to the discipline of instantiation. This communication takes a complementary path and attempts to focus on (i) the case of abstracting away a reusable