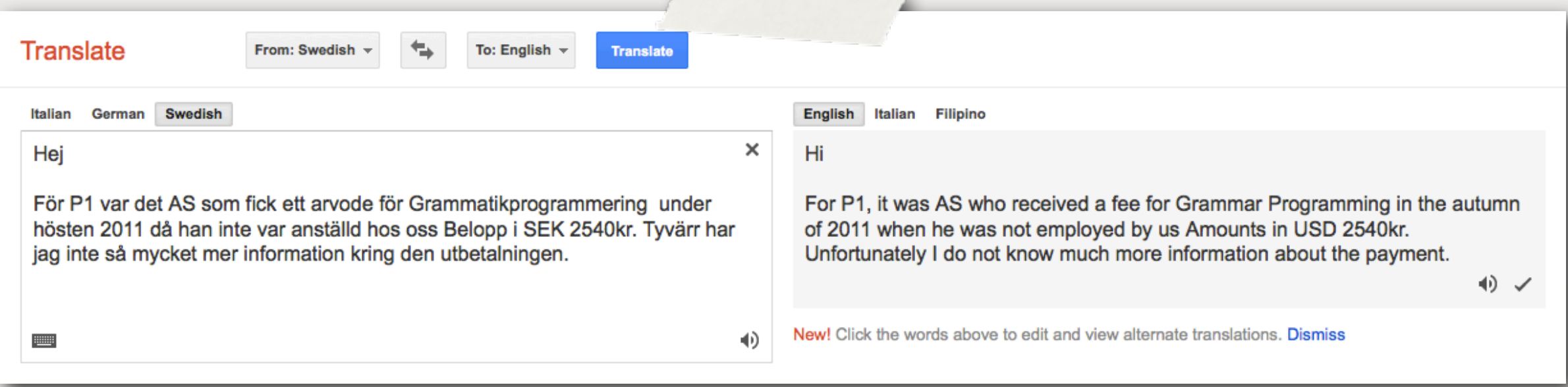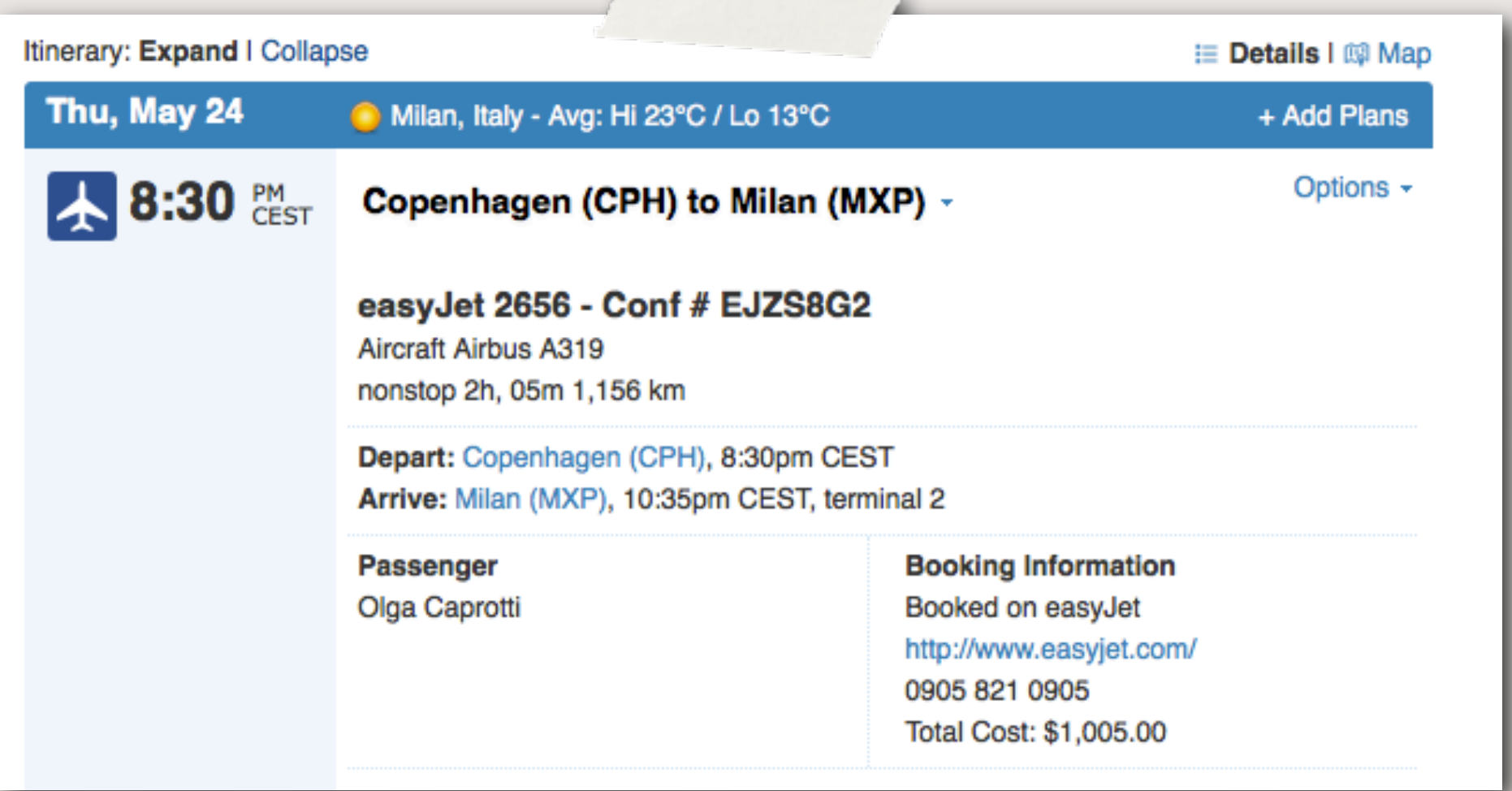Many web-sites today are transparently using web services based on statistical machine translation to provide multilingual versions. Wide coverage comes at the cost of some translation errors. It is obvious that in case of e-business, quality demands the correct translation of currencies.

MOLTO case studies are characterized by the usage of a *restricted language* since it is impossible to combine large coverage with high precision in automatic translation. Adoption of high-quality restricted language translations has been hindered by:

❖ development cost for a new domain or language,
❖ learning curve for authoring texts in a restricted language.

MOLTO tools radically decrease the effort of developing restricted language translators by entrusting the linguistic aspects to the Grammatical Framework (GF) libraries.: a grammar author needs few basic skills to be able to add a new language.

Feature	Current	Projected
Languages	up to 7	up to 15
Domain size	100's of words	1000's of words
Robustness	none	open-text
Cost/domain	months	days
Cost/language	days	hours
Learning (grammarians)	weeks	days
Learning (authors)	days	hours

MOLTO is also exploring the two-way *interoperability of grammars* with *Semantic Web conceptual models* (ontologies) and *hybrid models of combining rule-based*

translation systems with statistical machine translation. Data sets, available in a machine readable form, like the Resource Description Framework, or the Web Ontology Language, are used to construct a knowledge representation infrastructure that can be queried in a meaningful way using natural languages.

Combination approaches aim to integrate grammar-based and SMT models in a hybrid, robust MT system to improve quality in unconstrained text translation. Variants under consideration include soft integration in which phrase pairs or tree fragment pairs, generated by GF, are integrated as discriminative probability models in a phrase-based SMT system. The testbed application for this research activity is information retrieval from patents in the pharmaceutical domain.

GF	Une utilisation selon la revendication 3, dans laquelle le médicament séparé est administré at the same time as…
SMT	Utilisation selon la revendication 3, dans laquelle le médicament séparée est administré en même temps que…
HI	Une utilisation selon la revendication 3, dans laquelle le médicament séparé est administré en même temps que…
SI0.5	Utilisation selon la revendication 3, dans laquelle le médicament séparé est administré en même temps que…
Ref.	Utilisation selon la revendication 3, dans laquelle le médicament séparé est administré en même temps que…

MOLTO is a European project in the computational linguistic field to develop tools for real-time multilingual translation of web documents with high quality. The approach is based on the Grammatical Framework used in combination with techniques from the semantic web and from statistical machine translation. MOLTO aims to provide technology which can simultaneously tackle issues arising from the following scenario for real-time machine translation of web documents: production of translations in several languages, maintainance of consistency between localized versions in spite of asynchronous collaborative authoring with frequent edits, and grammatically and stylistically flawless text. The number of languages to be covered in the translations is 15, including 12 of the 23 official languages of the European Union: Bulgarian, Danish, Dutch, English, Finnish, French, German, Italian, Polish, Romanian, Spanish, and Swedish. The 3 non-EU languages are Catalan, Norwegian, and Russian.

MOLTO

non multa, sed multum

Multilingual Online Translation

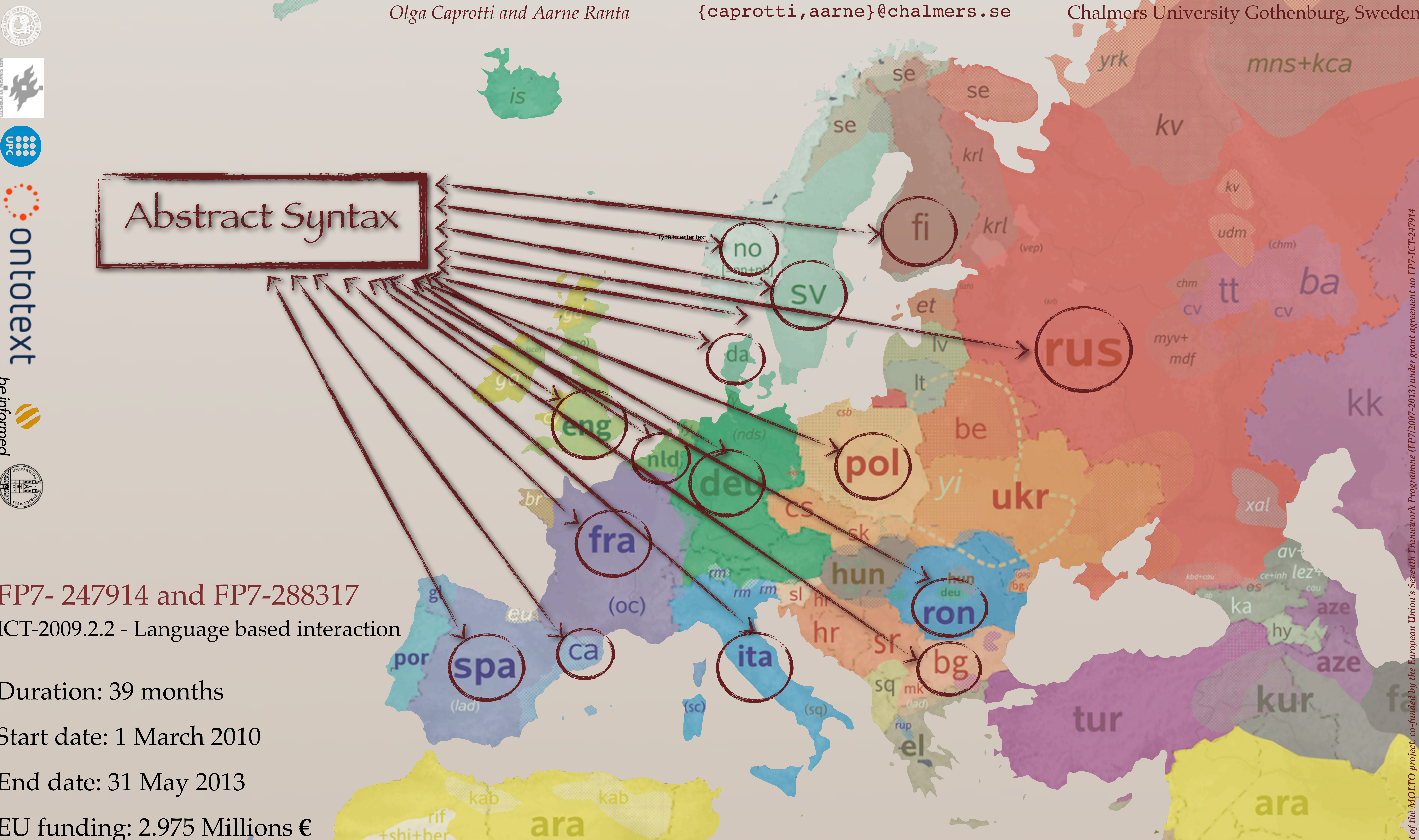molto-project.eu
twitter.com/moltoproject.eu

Olga Caprotti and Aarne Ranta   {caprotti,aarne}@chalmers.se   Chalmers University Gothenburg, Sweden

Abstract Syntax

Deep evaluation of hybrid architectures: simple metrics correlated with human judgments. Labaka, Gorka, Díaz De Ilarraza Arantza, España-Bonet Cristina, Sarasola Kepa, and Marquez Lluís, International Workshop on Using Linguistic Information for Hybrid Machine Translation, 11/2011, Barcelona, Spain, p.50-57, (2011)
Annual Public Report M24. Caprotti, Olga et al., 03/2012, Number D X.2, University of Gothenburg, (2012)
The GF Mathematics Library. Saludes, Jordi, and Xambó Sebastian, Proceedings First Workshop on CTP Components for Educational Software (THedu'11), 02/2012, Volume Electronic Proceedings in Theoretical Computer Science , Number 79, Wroclaw, Poland, p.102–110, (2011)
Typeful Ontologies with Direct Multilingual Verbalization. Angelov, Krasimir A., and Enache Ramona, Controlled Natural Languages Workshop (CNL 2010), Marettimo, Italy, (2011)
The Patents Retrieval Prototype in the MOLTO project. Chechev, Milen, Gonzàlez Meritxell, Màrquez Lluís, and España-Bonet Cristina, WWW2012 Conference, Lyon, France, (2012)
Multilingual Online Generation from Semantic Web Ontologies. Dannélls, Dana, Enache Ramona, Damova Mariana, and Chechev Milen, WWW2012, 04/2012, Lyon, France, (2012)
Using GF in multimodal assistants for mathematics. Archambault, Dominique, Caprotti Olga, Ranta Aarne, and Jordi Saludes, 02/2012, Digitization and E-Inclusion in Mathematics and Science 2012, (2012)
A Hybrid System for Patent Translation. R. Enache, C. España-Bonet, A. Ranta, L. Màrquez, Proceedings of EAMT 2012, May 2012, Trento Italy.
GF Eclipse Plugin: an IDE for grammar development in GF. John J. Camilleri, Krasimir Angelov, Proceedings of EAMT 2012, May 2012, Trento Italy.