

P A R R S  M E

Translation issues group

Mika Hämäläinen, Jelena Mitrović, Aleksandar Petrovski, Mafīss Rikters, Beatriz Sánchez Cárdenas, Katalin Simkó, Georgios Zakis

Supervisors: Jörg Tiedemann, Agata Savary

2nd Training School, 27 June - 1 July, La Rochelle



Idiomatic expressions

- ▶ Pairs of idioms with the same compositional meaning but different idiomatic meaning
- ▶ Building a database of Hungarian-English idioms
- ▶ Similar to False friends in some respect
- ▶ Mostly an issue of recognizing as mwe



Idiomatic expressions

► Word for word translation, but different idiomatic meaning e.g.

► HUN: *Helyén van a szíve*

Place.sg.sup is.pres.3sg the heart.sg.nom.poss.3sg

His heart is in the right place/Brave

► EN: *His heart is in the right place*

Has good intentions

► GR: *Η καρδιά μου είναι στη θέση της*

The heart poss.1sg is in_the position poss.3sg

My heart is in its place/ I am relieved



Using computational rhetoric resources for MT

- ▶ Ontology of Rhetorical Figures for Serbian – formal representation of 98 figures of speech
- ▶ Using WordNet enriched with a *semantic relation* related to the Simile rhetorical figure (successfully used for *detection of irony in Twitter*, along with antonymy and synonymy relations)

Crven kao mak “Red as Poppy”

Red is *SpecificOf* Poppy

Poppy is *SpecifiedBy* Red



Simile in different languages

- ▶ Άσπρο σαν το γάλα – Beo kao mleko – “As white as milk”
- ▶ Άσπρο σαν το χιόνι – Beo kao sneg – “As white as snow”
- ▶ Άσπρο σαν το πανί – Beo (bled) kao krpa – “As white (pale) as a kitchen cloth” –negative sentiment, meaning that someone is very pale, very sickly looking.



Easy as...

Serbian: “Prosto ko pasulj” As easy as cooking beans

Macedonian: “Prosto kako grav” - | | -

English: “Piece of cake”; “Easy peasy”; “As easy as taking candy from a baby”



Easy as...

Spanish: “Más fácil que comerse un huevo” Easier than eating an egg
“Es pan comido” It's bread eaten

French: “C'est du gateau” It's a cake

Hungarian: “Gyerekjáték” Child's play

Greek: “Κάτι είναι παιχνιδάκι” Child's game



Possible Solutions (1)

- ▶ Vector models could be useful for Machine translation
- ▶ Syntactic models for mapping
- ▶ Explicitation and Implicitation techniques



Possible Solutions (2)

- ▶ XML markup in input language – saying that a rhetorical figure should be treated differently
- ▶ Using parallel ontologies
- ▶ Adding aligned translations of the rhetorical figures to the training data of the SMT system



Merci!
Ευχαριστώ!
Thank you!
Hvala!
Blagodaram!
Kiitos!
Köszönöm!
Paldies!
Gracias!