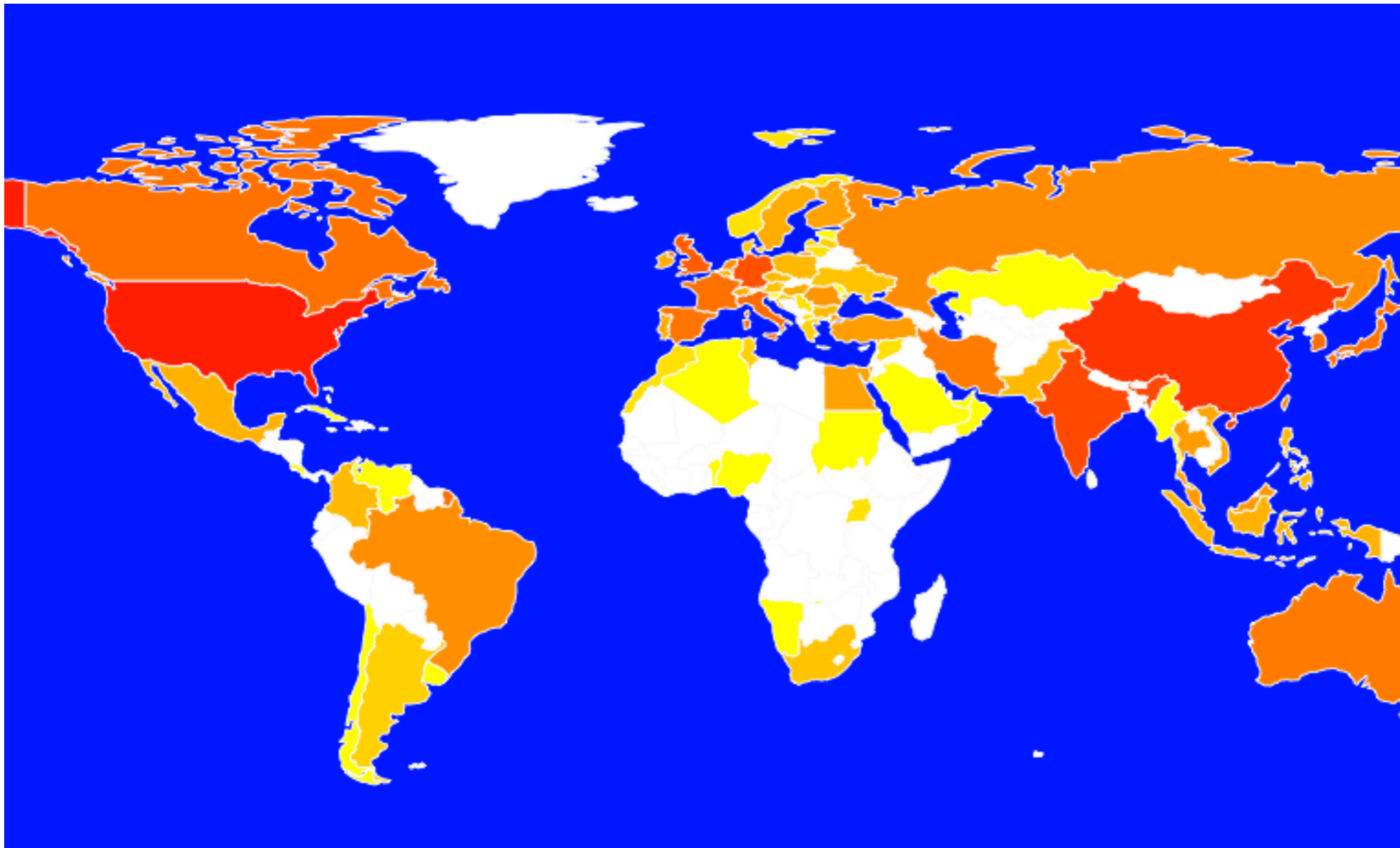


FrameNets in "Other" Languages

A Quick Introduction

Collin Baker
Multilingual FrameNet Tutorial
LREC 2016,
Portorož, Slovenia



ICSI FrameNet Users by Country

FrameNets in Other Languages

Name	Institution	Leading figure
Spanish FN	UA Barcelona	Carlos Subirats
SALSA	Saarland U	Manfred Pinkal
Japanese FN*	Keio U, Tokyo U	Kyoko Ohara
Chinese FN	Shanxi U, Taiyuan	Liu Kaiying, Li Ru
Swedish FN*	U Gothenburg	Lars Borin
FN Brasil	UF de Juiz de Fora	Tiago Torrent
French FN*	multiple	Marie Candito
Hebrew FN	Ben Gurion U	Michael Elhaddad
Korean FN	KAIST	Key-Sun Choi
Arabic FN	UAE U	Andrew Gargett

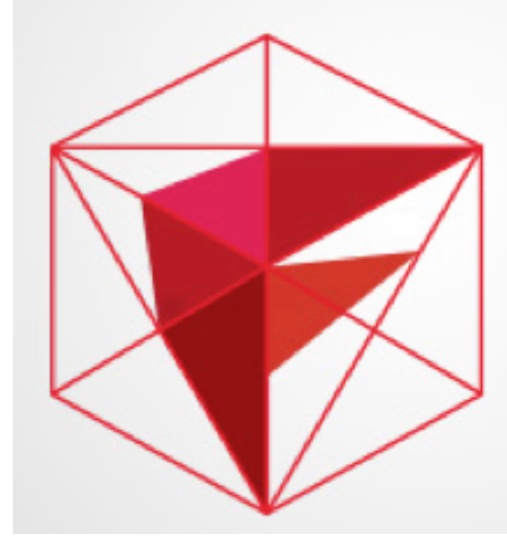
Spanish FrameNet

(Subirats 2009)

- Created a new balanced corpus, mainly New World Spanish, and their own POS tagger, large tagset
- Manual annotation, frame by frame lexicographic annotation, following Berkeley closely,
- Generally English frames were OK, some differences e.g. verbs of motion (verb framed vs. satellite framed)

<http://sfn.uab.es>

FrameNet Brasil



- PIs: Maria Margarida Martins Salomão and Tiago Timponi Torrent (Salomão *et al.* 2013)
- At Universidade Federal de Juiz de Fora, Minas Gerais, Brazil
- Projects:
 - Building FrameNet for Brazilian Portuguese
 - Copa (2014) website for Soccer World Cup with frames/LU in EN, ES & BrPT
 - m.knob (2016) FN mobile app & website for Summer Olympics
 - Creating construction for Brazilian Portuguese

<http://www.ufjf.br/framenetbr-eng/>

LU counts by project

Chinese FN	3,947
FN Brasil (BPT)	251
FN Brasil (Copa)	1,125
Japanese FN	3,392
SALSA (DE)	1,826
Spanish FN	1,269
Swedish FN	33,183
ICSI FN	13,235

Numbers are not reliable, do not cite!

Counts by POS

	Spanish	Swedish	Japanese	English
N	272	24,736	2,043	5,348
V	856	5,398	908	5,080
Adj	99	3,229	133	2,320
Adv	16	201	89	201
Other	26	216	221	420
Tot LUs	1,269	33,780	3,394	13,189
Anno.s	10k	13k?	73k	~200K

Numbers are not reliable, do not cite!

Variations in approaches

- Lexicographic vs. corpus-based annotation
- General coverage vs. specialized domain
- Manual vs. automatic
 - Projection from English lexicon
 - Projection from English annotation with translation
 - ASRL in English, in target lang.
- Relation to existing lexical resources

Manual Annotation

- Spanish FN: created own balanced corpus, mainly New World Spanish, own NLP tools, annotation tools adapted from ICSI FN to Spanish, frame-by-frame lexicographic annotation.
- SALSA (Burchart *et al.* 2006): Used existing parsed corpus, contracted for new annotation tool based on parse trees. Created many partial frames as needed.

Other FN Building Methods

- Projection from English FN
 - (S. Padó 2007 Ph.D. thesis, Padó and Lapata 2009) Cross-lingual annotation projection
 - Swedish (R. Johansson & Nugues 2005)
- Starting from a corpus
 - SALSA, Korean FN
- Based on existing lexical resource
 - Swedish FN (Borin *et al.* 2010), Korean FN
- For special domains
 - Kictionary (Schmidt 2008), World cup