

Why??

Problem:

- Lack of information and documentation about resources is, in the e-science paradigm, a very critical issue
- Non documented resources **don't exist!!**

The Map as an answer to start to fill this gap, but also:

- To encourage the needed **“change in culture”**
 - Each researcher must become aware of the importance of his/her **personal engagement in documenting resources**
 - A task as important as creating new resources and not an accessory to be disregarded
 - As the necessary **service to the whole community** that we all have to provide (take care of the forest and not only of your tree!)

How many LRs?

Submissions: 1288



Language Resource forms: 1994

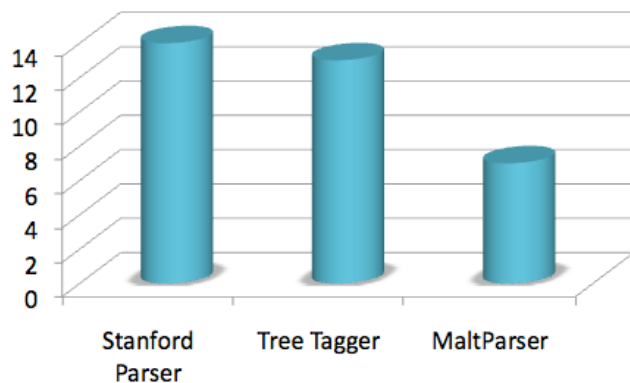
Corpora: 785
Lexicons: 289

Tagger/Parser: 181
Annotation tool: 134

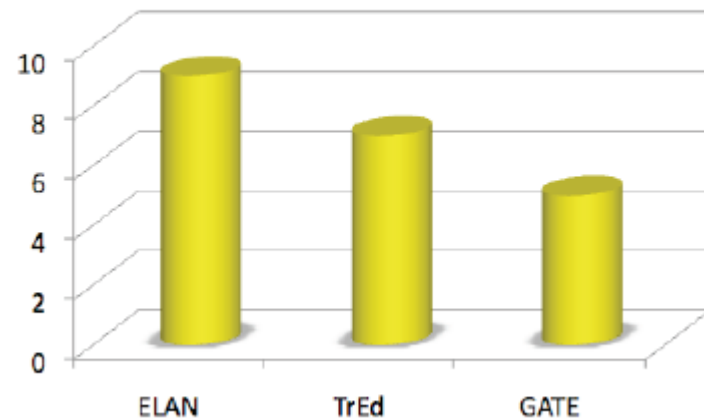
Ontology: 73
Evaluation data: 40

Annotation Guidelines: 35
...

Taggers and Parsers



Annotation Tools



Languages:

170 !!

- But obviously ...



image courtesy of Wordle (<http://www.wordle.net>)

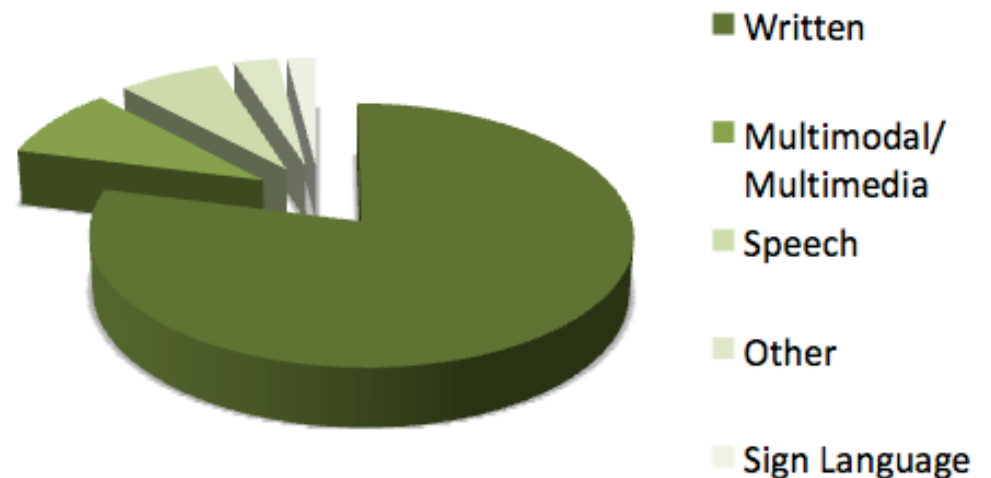


Languages & Modality

Monolingual vs. multilingual resources



Modality



Applications



Most frequent applications



- IE, IR
- MT, SpeechToSpeech Translation
- Knowledge Discovery and Representation
- Language Modelling
- Acquisition
- Other

Application	%	Citations
Information Extraction, Information Retrieval	16	224
Machine Translation, SpeechToSpeech Translation	12	171
Knowledge discovery and representation	7	93
Language Modelling	5	78
Acquisition	5	78
Emotion Recognition	4	63
Discourse	4	60
Named Entity Recognition	4	56
Word Sense Disambiguation	4	53
Question Answering	3	48
Speech Recognition	3	47
Dialogue	3	47
Document Classification, Text categorisation	3	42
Web services	3	37
Semantic Web	3	36
Language Identification	2	33
Text Mining	2	33
Part-of-Speech Tagging	2	25
Sign Language Recognition/Generation	2	23
Textual Entailment	1	19
Parsing	1	19
Speech Synthesis	1	17
Natural Language Generation	1	15
Summarisation	1	12

Most used Resources for some Applications

For MT

- Moses
- Europarl
- GIZA++
- JRC Acquis
- BLEU
- Google Translate
- SRILM

For IE-IR

- Arabic Treebank
- ACE corpora
- Open CV
- Wikipedia
- GATE
- TREC d

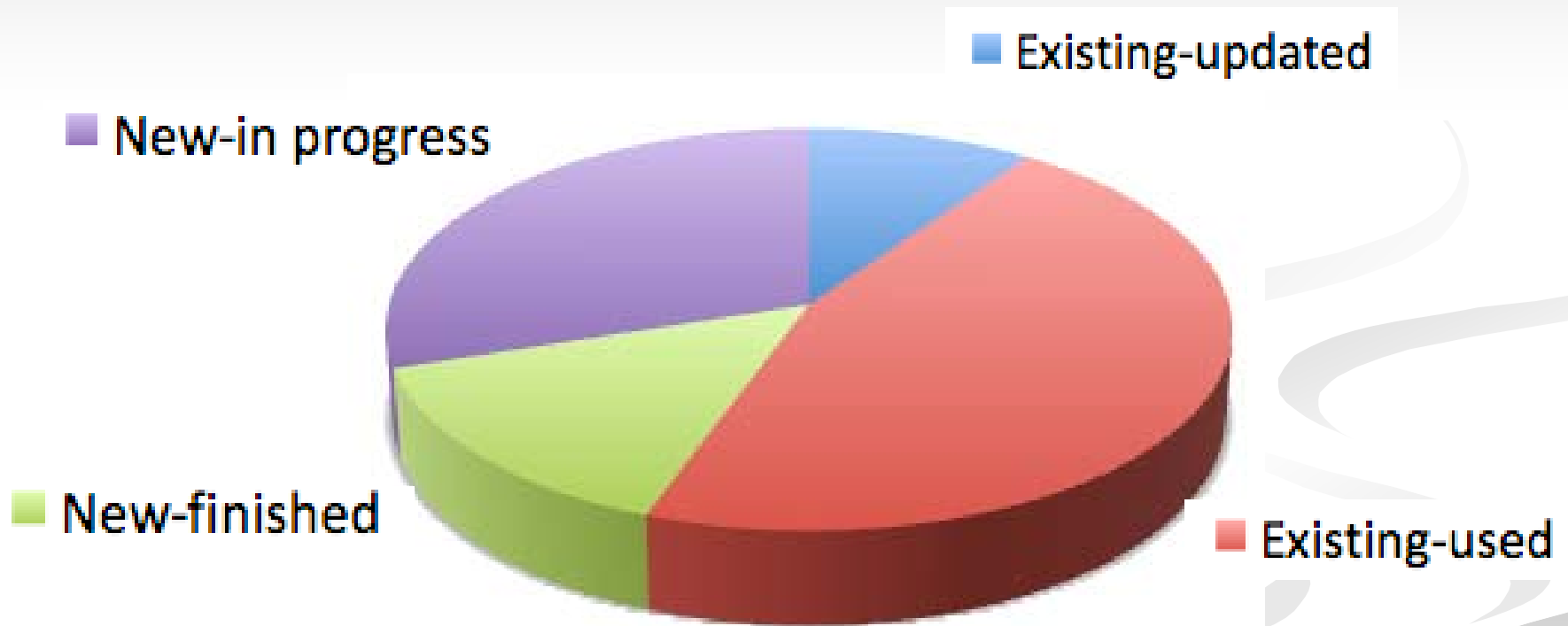
For Knowledge Discovery

- WordNet
- Wikipedia

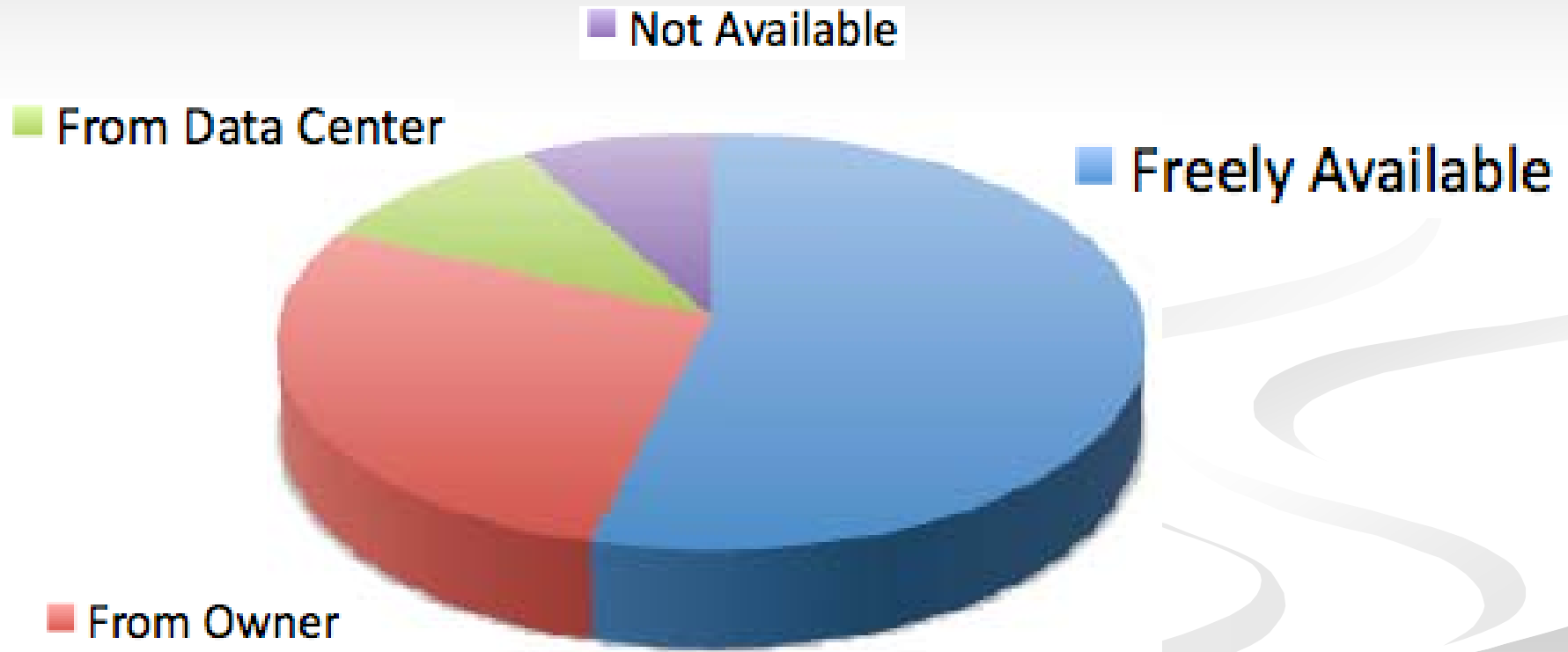


New Resources

New: 44%



Availability



The wide majority of resources (54%) are freely available

Simple idea

↔ Great Potential??

Many uses, in addition to being **an information gathering tool**

- It may become a great instrument **for monitoring the evolution of the field** (useful also for funders), if applied in different contexts and times
- It can be seen as a huge joint effort, the beginning of an even larger **cooperative action** not just among few leaders but among all the researchers
- It can thus become also an **“educational” means** towards the broad acknowledgment of the need of meta-research activities with the active involvement of many
- It is also instrumental in introducing the **new notion of “citation of resources”** that could provide an award and a means of scholarly recognition for researchers engaged in resource creation

Next steps??

→ Available to all

- I am proud of the fact that the Map is already being adopted in other contexts
- The *Language Resources and Evaluation Journal* will also make use of it, as announced in the new Special Issue *LREC 2008: Selected papers* (freely accessible to all LREC participants)
- And quite recently also **COLING 2010** has decided to make use of the Map
- The Map is taking its first steps outside home!



Dissemination

- The Map couldn't happen without the contribution of all: it will be necessary to continue
 - We will take actions
- Present at the **EC Village**
- First analyses on: <http://www.resourcebook.eu/>
- **More will be added:**
 - Everyone could add new analyses
- There will be the possibility **to add info on new LRs** also independently from Conferences!

Next steps??

- **Check intersection** of LRs in the Map with those of **ELRA & LDC catalogues**
- **Integration** with existing catalogues
- **Source of information for META-SHARE**
- ...

Next steps?? First suggestions

- Discuss its adoption in COCOSDA meetings/
workshops
- ...