

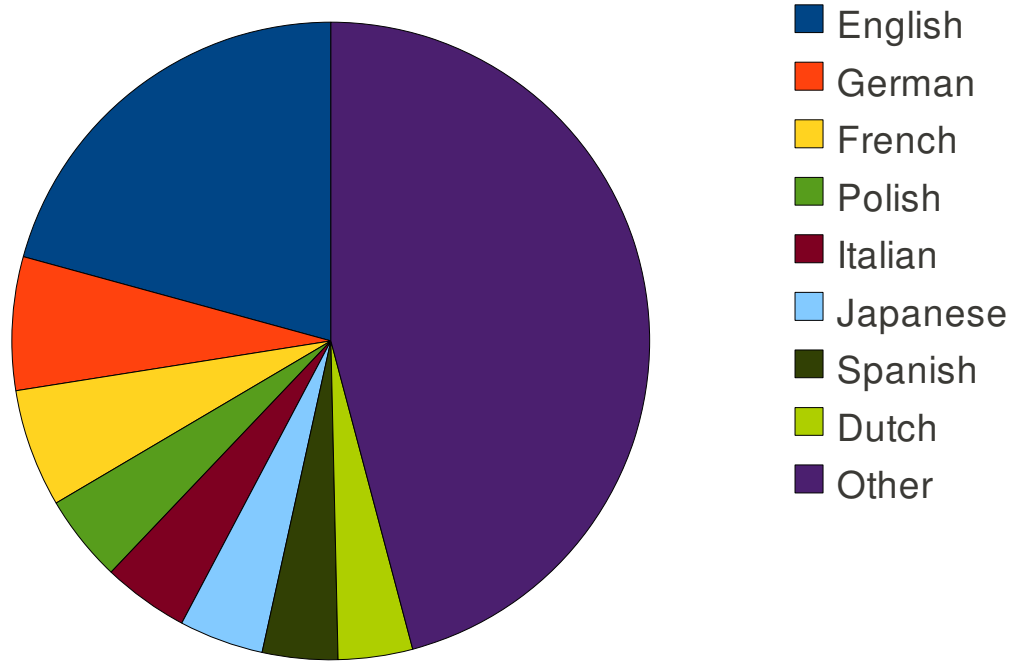
Reciprocal Enrichment between Wikipedia and Machine Translators

OpenMT-2 project

Mikel Iturbe
Wikimania 2010
Gdańsk, Poland

languages in
wikipedia

Distribution of wikipedia articles by language



Less than 1% of
languages have
more than 50% of
articles

Can we ease good
article creation?

How can we boost
article creation in
minority
languages?

OpenMT-2 project

<http://ixa.si.ehu.es/openmt2/>

What is it?

EHU, UPC and
Basque wikipedians

Funded by the
Spanish
government

Free



Hybrid Machine Translation and advanced evaluation system

Hybrid?

Rule-based MT

+

Statistical post-editing

The aim: To teach the existing MT to correct its own mistakes when translating

Using wikipedia

How?

(1)

Translate using
rule-based
Matxin-Opentrad

<http://opentrad.com/>

100 long articles

es  eu

(2)



Correct Basque
output manually

(3)

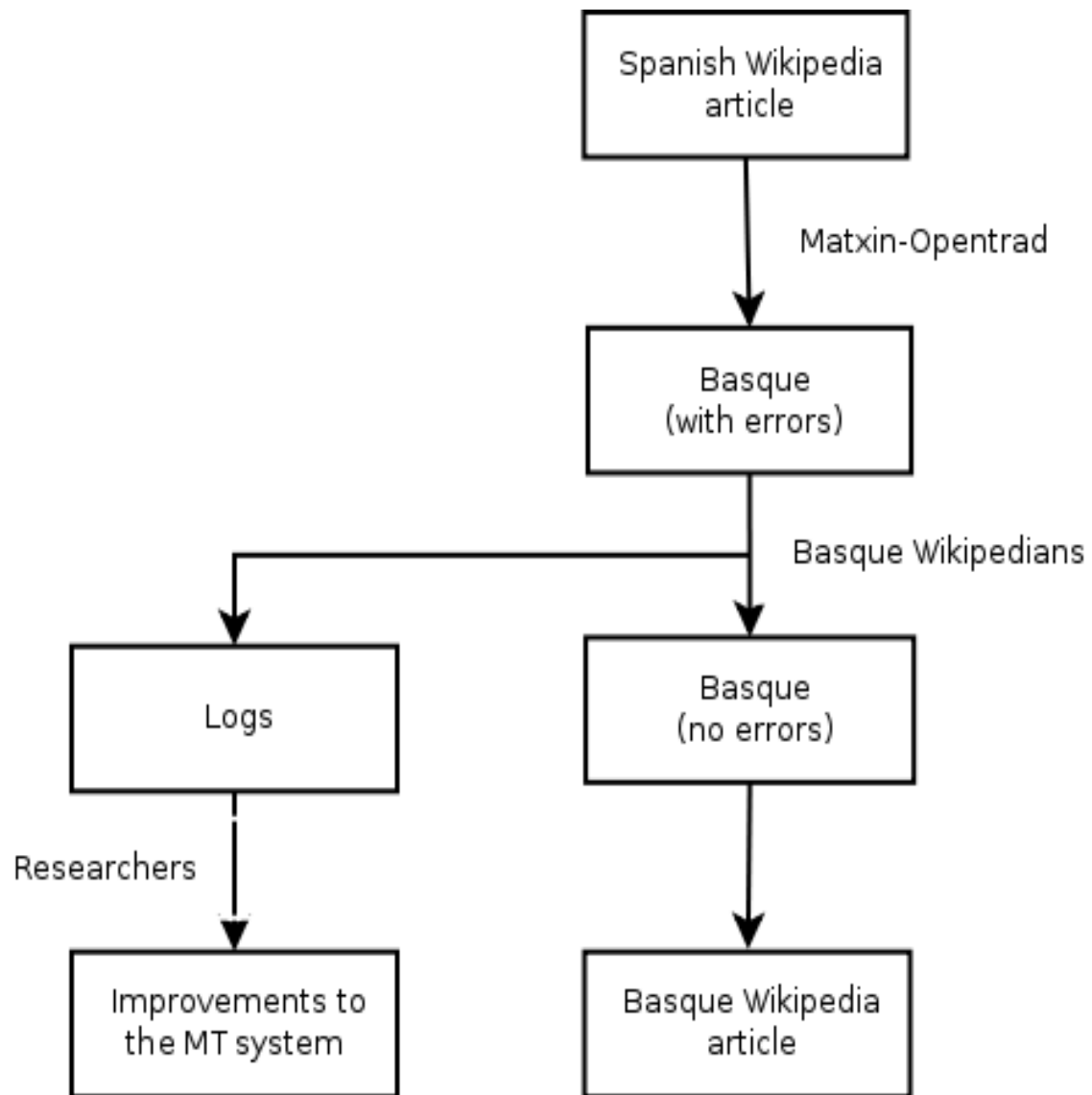
Analyze logs



(4)



Make
improvements to
the MT system



Final test and results

Tools

Google translator toolkit

Specific help for wikipedia ✓
Not Free Software ✗✗



OmegaT

<http://omegat.org>

Suitable to do the job



Free software



What's in?

100 new and good
articles for the
Basque Wikipedia

Provide research
material

Walk towards a MT
system that can be
used in our wikipedia

Thank you.

Aurélio A. Heckert ([source](#)), David Vignoni ([source](#)),
Wilfredor ([source](#)), Tango project & Arkanosis ([source](#))
, OmegaT project ([source](#))

Image credits

e-mail: mikel@hamahiru.org

User page: <http://eu.wikipedia.org/wiki/Lankide:Janfri>

Address: <http://hamahiru.org/media/wikimania2010.pdf>

contact

Text licensed under
cc-by-sa 3.0

images maintain their original licenses