Online and Free! Ten Years of Online Machine Translation: Origins, Developments, Current Use and Future Prospects

Federico Gaspari   *John Hutchins*

F.Gaspari@manchester.ac.uk   *w_john_hutchins@yahoo.co.uk*

Origins of Online MT

- Minitel service from 1988

  – 22 lines of text per minute, charge of $1.20 per page

  – potentially accessible to 4.5 million users in France (Ryan, 1987: 100)

- CompuServe started testing in 1992

  – "CompuServe's basic goal for MT is to provide draft-quality translation […] We suspect that there is a market for low-cost translations, even if the quality is less than ideal" (Harrison, 1992: 11)

- Babel Fish launched on AltaVista on December 9, 1997

  – Systran had offered online translations of webpages since 1996

Attractiveness

- To users

  - free of charge, instantaneous, anything accepted, many languages

- To vendors

  - "sprat to catch mackerel", i.e. hope users will buy 'more advanced' commercial software

- To web designers

  - Cheap multilingual localisation

- Potentially to MT researchers

  - 'real' language, major impact on the Internet community -- so why neglected?

Use of AltaVista Babel Fish (between 1998 and 2000)

Number of translations processed per day between May 98 and Oct. 00

Span of 29 months
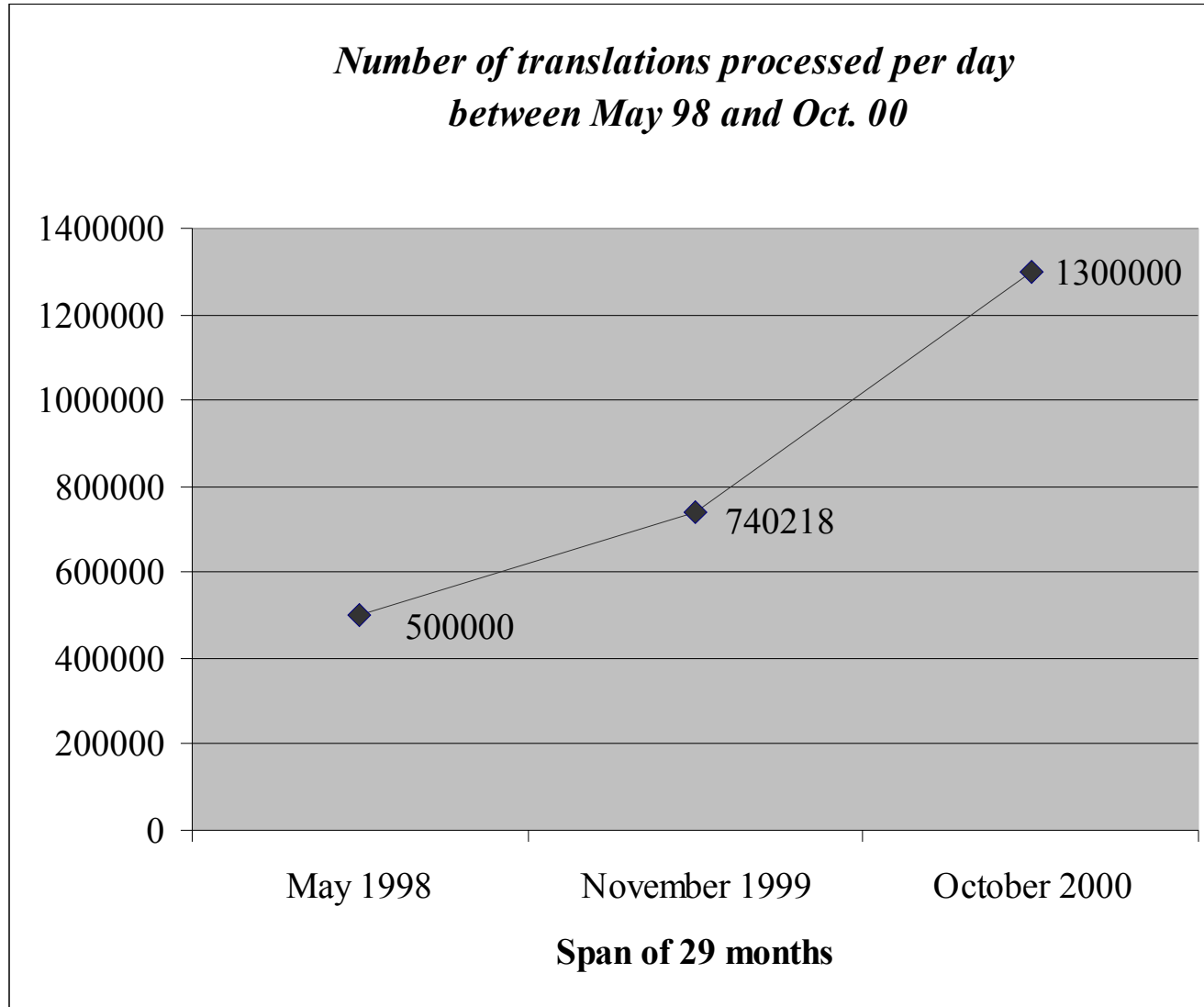
Based on Yang & Lange (1998) and Yang & Lange (2003)

Use and Misuse

- Most popular languages
  - English/Spanish, English into local native language

- Input
  - taboo words, 'adult content', email, chatroom postings

- Purposes
  - assimilation, dissemination, language learning, entertainment

- Trying out MT

  - proverbs, idioms, 'difficult' (trick) expressions
  - back-forth translations (not new! Apocryphal "The spirit is willing…")

- In the main, users seem unaware of MT limitations
  - based on assumptions about the infinite power of computers?

Online MT and the MT Community

- Impact on commercial vendors

  - hope for increased sales of products

  - add-on services: longer texts, post-editing, specialized dictionaries

  - customization and repackaging

- Impact on MT research

  - more research on 'ungrammatical' language (email, chat, etc.) -- not much evidence yet!

- Legal implications

  - liability for mistranslations or illegal use of online MT: first flagged up by Westfall (1996), still seem thorny unresolved issues

Translators and Online MT

- Do they use it?

    - very little, about 5% of freelance translators is reported in the literature

- How?

    - search for up-to-date terminology

    - produce rough draft which gets polished

- Possibility that HT may increase when users of online MT realise need for quality translation and become more aware of limitations of MT

    - but usually websites of professional translators and translation agencies make a mockery of online MT, showing howlers to suggest it is useless

Language pairs: FreeTranslation (since 1999)

Most popular language pairs based on a sample of 1 million requests to a single server (approx. 86%)
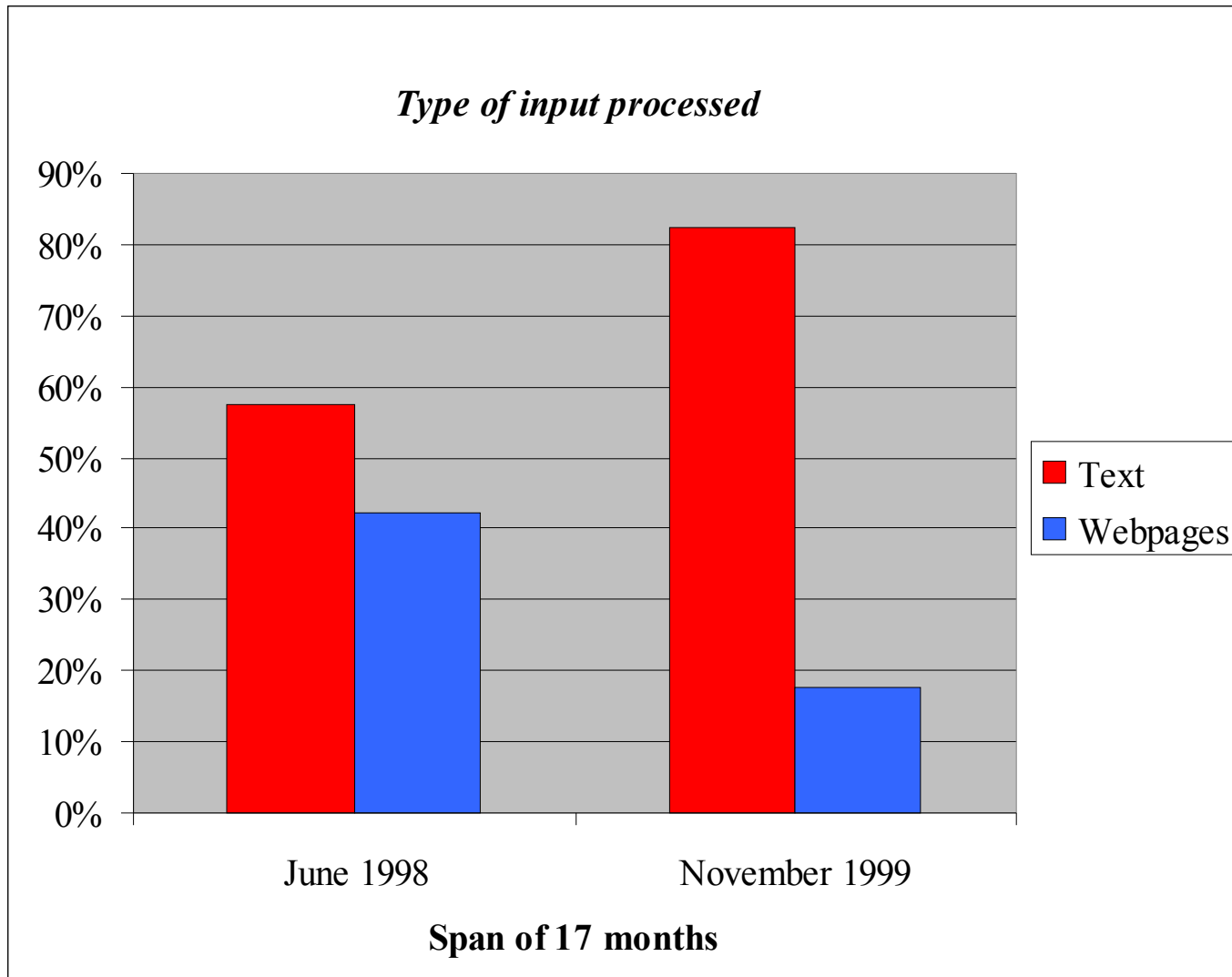

- ES>EN; 22.17%
- EN>FR; 11.85%
- FR>EN; 7.31%
- DE>EN; 5.09%
- EN>DE; 4.91%
- EN>ES; 34.83%
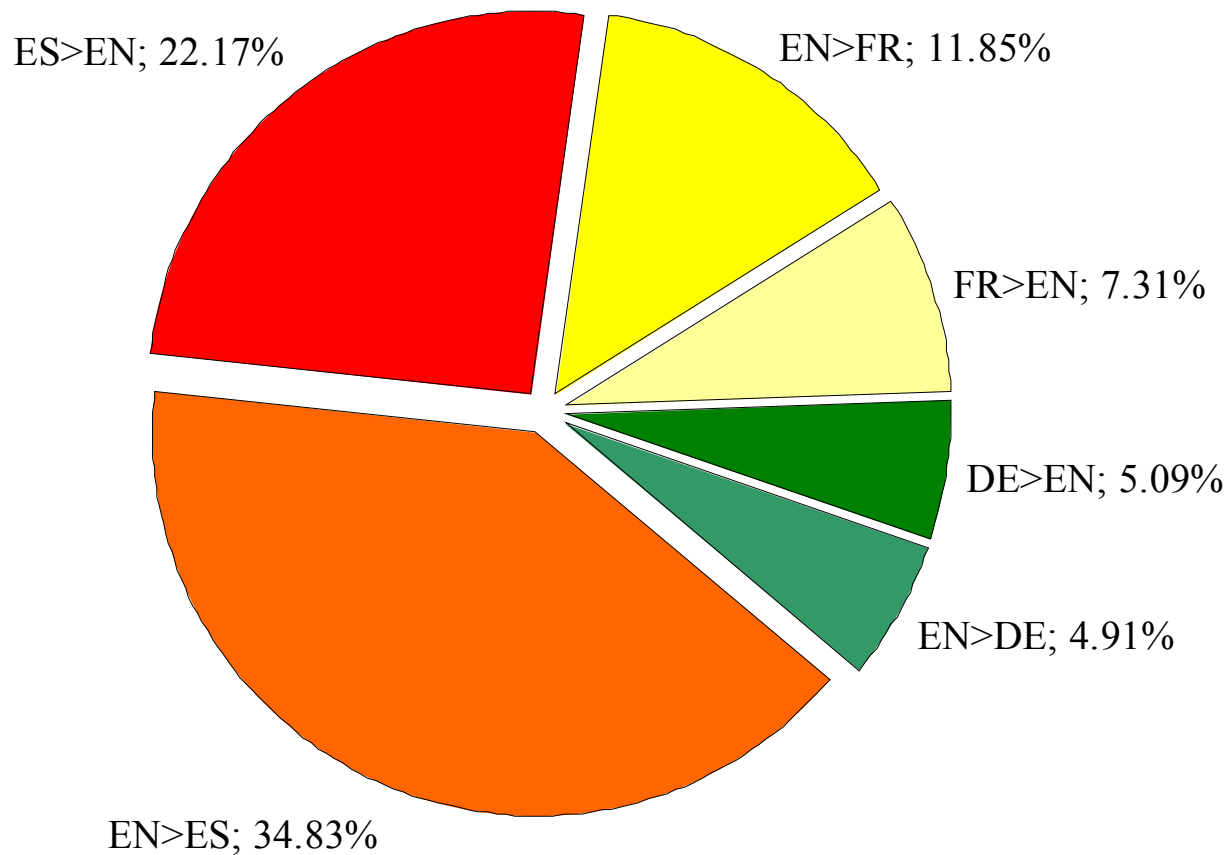
Source: Jay Marciano (FreeTranslation)
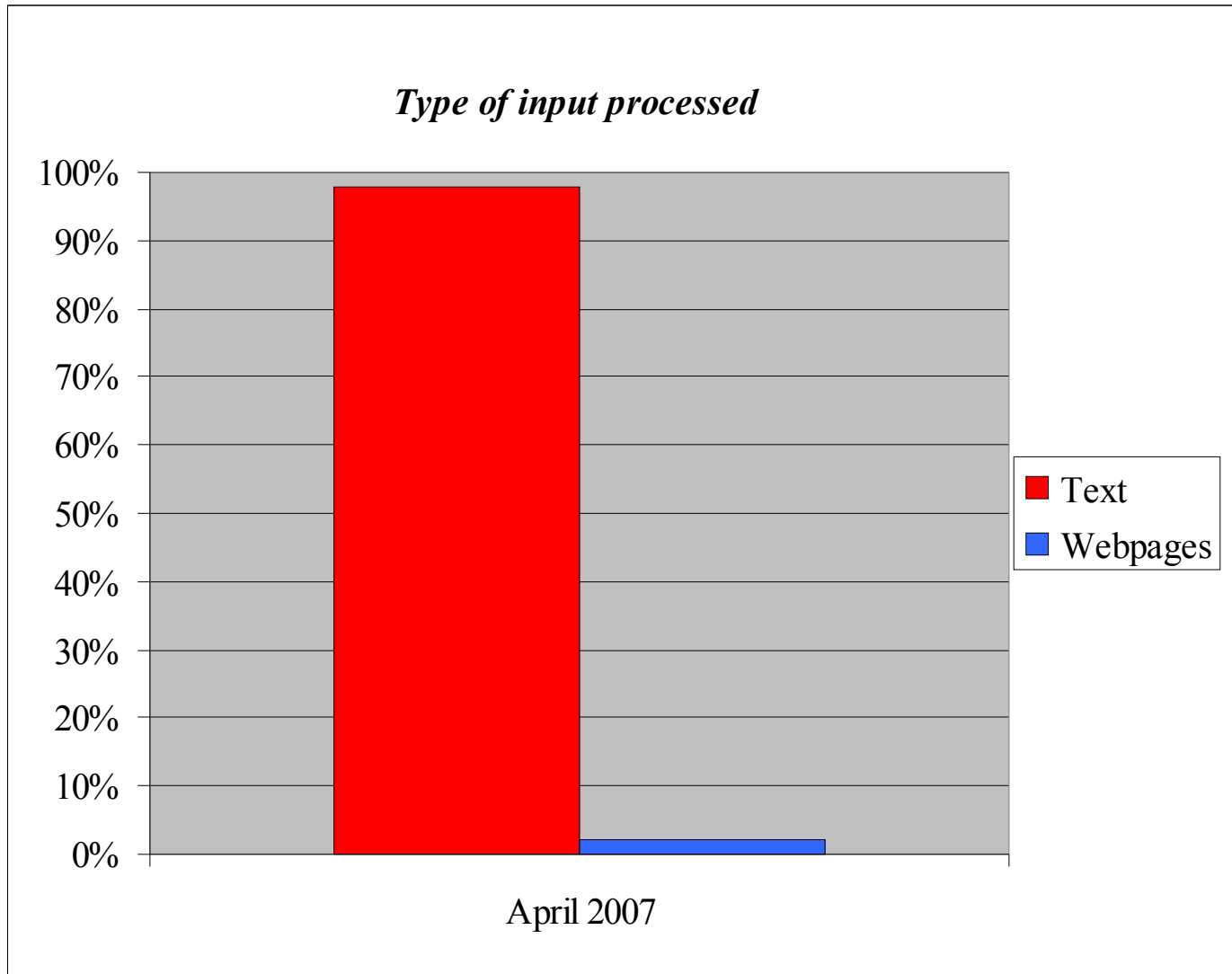
Current Use of Online MT: text vs URL

- Input: words, texts, URLs
  - still mainly single words, short phrases, few long texts, and relatively few webpages
  - is online MT used as electronic dictionaries by users who already have some knowledge of the source (and/or target) language and require only occasional assistance?
  - are users looking for help with only very short text extracts?
  - volume increase in 10 years -- little incentive to improve systems
- Why is translation of webpages so low compared to text?
  - 17% or less, with ratio of text translations constantly increasing
- Why so little text as opposed to single words/phrases?
  - over 50% just one or two words; average length less than 15/20 words

# Use of AltaVista Babel Fish (two days in 98 and 99)

Type of input processed

Legend: Text (red), Webpages (blue)

Categories: June 1998, November 1999

Span of 17 months

Based on Yang & Lange (1998) and Yang & Lange (2003)

Use of Yahoo! Babel Fish (April 2007)

Source: Raymond Flournoy (Yahoo! Babel Fish)

Use of FreeTranslation (1999 to 2006)

Number of translations processed per day between Dec. 99 and Sep. 06
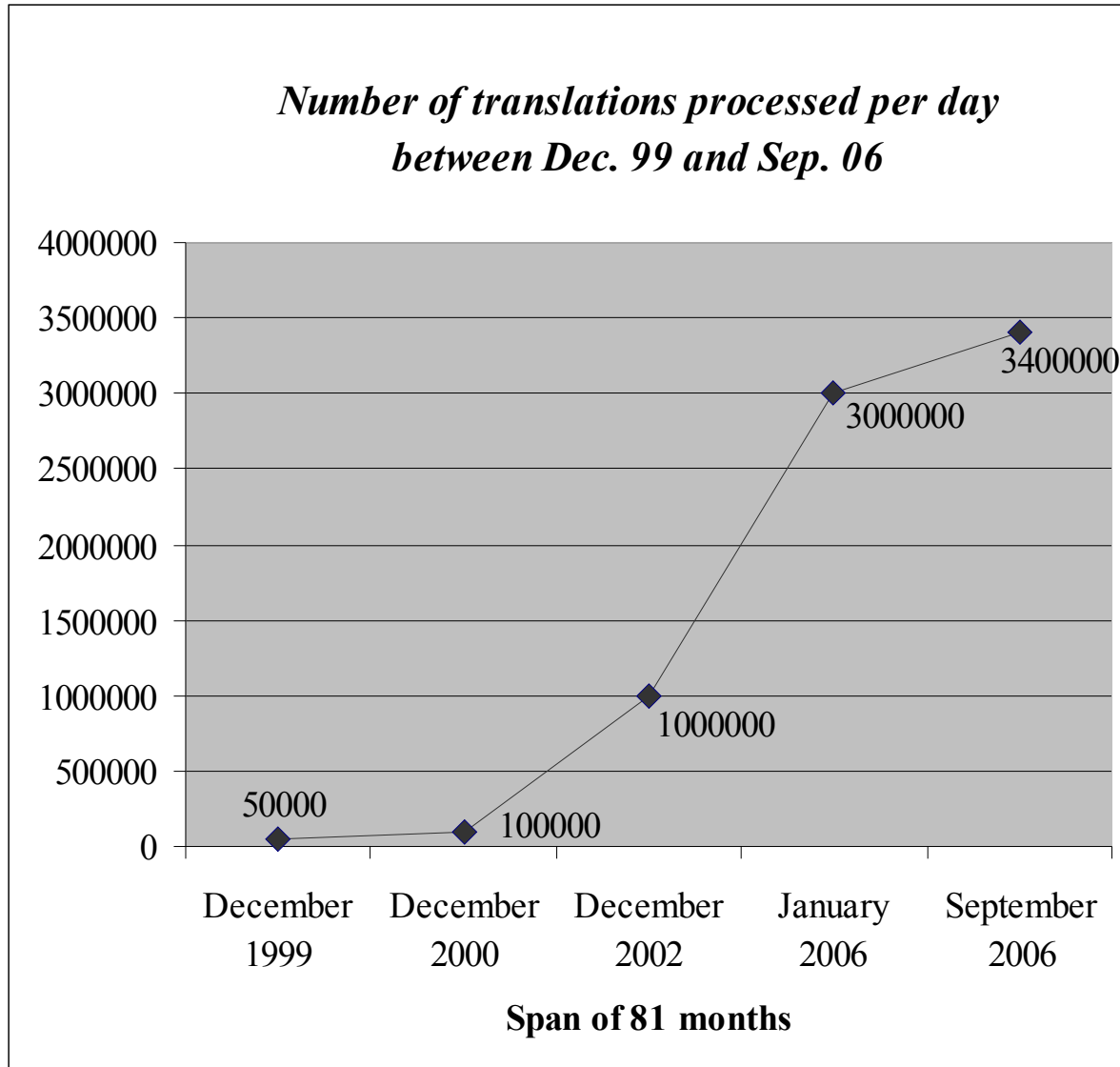
Source: Jay Marciano (FreeTranslation)

Issues for Evaluation

- Does online MT satisfy users?

  - for what purposes?

- What are the backgrounds of users? How much of the source and target languages do they know?

- What do they dislike most?

  - vocabulary omissions

  - no disambiguation

  - bad grammar and syntax

- What developments would they like?

  - which new languages?

  - special domains? (e.g. medical texts, news texts)

Future Prospects and Needs

- Since we know so little it is difficult to predict future, but we may expect:
  - increased frustration with slow progress of MT
  - increased assumptions that online MT is the 'best' that MT can do

- What is needed:
  - more data from providers about usage and user feedback
  - surveys of users' needs, opinions, preferences and expectations
  - robust comparative evaluations of systems (online, PC and research)
  - promote good practice in the use of MT within monolingual sites
  - clear, jargon-free disclaimers on the use of online MT services

Future challenges

- For researchers:

  - More attention to 'ungrammatical' (casual, conversational, non-literary, etc.) language for text online MT

- For developers:

  - Systems linked to ('embedded with') other online NLP applications (e.g. summarization, information retrieval/extraction)

  - Systems for less common languages

  - Systems for specialised domains (e.g. medical and news texts)

- For researchers and developers:

  - Online MT for spoken language

Thank you!

Federico Gaspari

F.Gaspari@manchester.ac.uk

John Hutchins

w_john_hutchins@yahoo.co.uk

Online and Free! Ten Years of Online Machine Translation: Origins, Developments, Current Use and Future Prospects