Supporting Non-Expert Translators with Language Services

Ari Hautasaari
Kyoto University
arihau@ai.soc.i.kyoto-u.ac.jp

The 1st Inter-University Symposium on Field Based Design
Hong Kong Baptist University – Kyoto University
Ph.D Student Forum
30.3.2013

Contents

Ari Hautasaari – Department of Social Informatics, Kyoto University

Topic 1: Wikipedia Article Translation by Non-Experts

- Activity Theory approach to analyze **translation work by individual non-experts** in online environments.
- Activity Theory approach to analyze the **effects of Language Service support**.

Topic 2: Tool Design Based on Activity Analysis

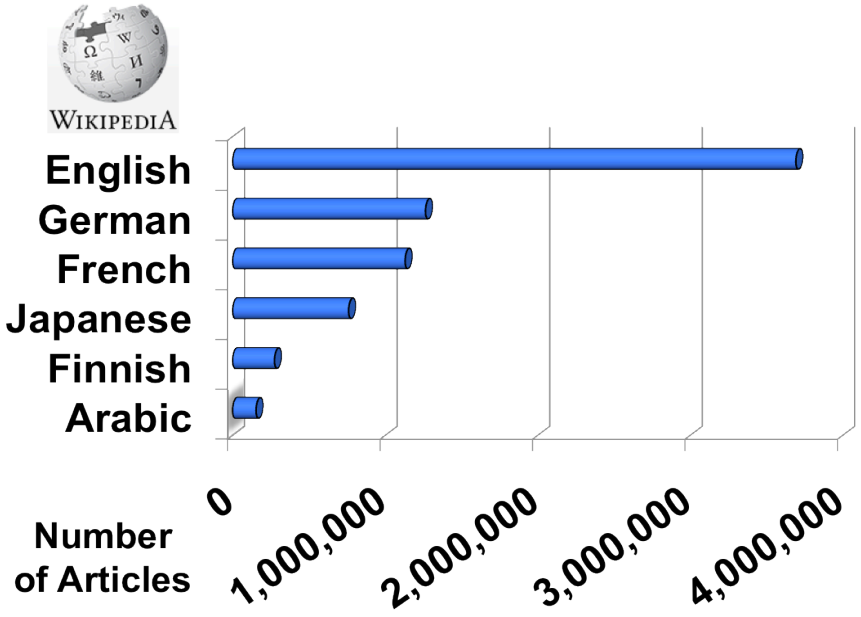
- Illustrate how **Activity Analysis** can be harnessed to derive design implications for novel tool design.

Ari Hautasaari – Department of Social Informatics, Kyoto University

CASE: Wikipedia Article Translation

Wikipedia is the largest online encyclopedia available.

However, there are large discrepancies in the number of articles and active users between different language versions.



Wikipedia translation activities aim to make information equally available in all languages by translating Wikipedia articles.

Ari Hautasaari – Department of Social Informatics, Kyoto University

Motivation

- Non-experts, or casual users, are an untapped resource for Wikipedia article translation.
  - However, very little previous research exists on **how non-experts translate articles** in practice.

- Wikipedia articles are **linguistically complex**, especially in the case of conceptual articles, making the mechanical translation process difficult even for experts [Yasseri et al., 2012].

- Very few tools are currently aimed at **supporting non-expert translators**, none of them in Wikipedia (e.g., “games-with-purpose –approaches”)

Ari Hautasaari – Department of Social Informatics, Kyoto University

Research Questions and Experiment Setting

RQ1: What are the characteristics of translation as a task for non-expert Wikipedia article translators?

RQ2: What are the information needs of non-expert translators?

RQ3: How are the information needs resolved in an online environment?

- 15 **English to Chinese** translators
- 3 English articles (no version in Chinese Wikipedia / Baidu Baike)
  - Akan Volcanic Complex, Crowdsourcing, Ying Huang
- Each participant was assigned to one article
- Task:
  - **Translate English language article to Chinese in an online environment**
  - 5 translations of each of the 3 articles
- Collected ~18 hours of **video data** for analysis

Activity Analysis

Ari Hautasaari – Department of Social Informatics, Kyoto University

- Activity Analysis aims to provide a detailed description of **human activity** conducted within the constraints and resources of the real-world setting.
- Activity Analysis consists of a qualitative study, an observation phase, and a detailed analysis of activity patterns based on Activity Theory [e.g., Bardram and Doryab, 2011].

Analysis Framework

Activity – Identifying all the activities included in the task based on the motive and object of work.
Action – All actions performed as a part of an activity.
Operations – Manual operations performed as a part of an action.
Context – Context of each action including the time of action occurrence during an activity, and the artifacts used to mediate the action.
Actors – Individuals conducting the activities.

Activity Analysis of Wikipedia Article Translation

Ari Hautasaari – Department of Social Informatics, Kyoto University

Example of a coding scheme

Activity	Action	Operations	Context	Actors
Translate Article A	Translate sentence by hand	-Type in Chinese sentence -Delete Chinese word(s) -Add Chinese word(s) -Copy words in target article -Paste words in to target sentence	-**Tool:** Text document -**Timestamp:** 12 minutes 30	Translator A1
Translate Article A	Search word in a dictionary	-Type word to search field -Copy word from source article -Paste copied word to search field -Paraphrase word in search field -Click “OK” -Read results	-**Tool:** Dict.cn online dictionary -**Timestamp:** 13 minutes 30 seconds	Translator A1
Translate Article C	Translate word in machine translator	-Type word to source field -Copy word from source article -Paste copied word to source field -Paraphrase word in source field -Click “OK” -Read results	-**Tool:** Google Translate online machine translator -**Timestamp:**5 minutes 22 seconds	Translator C5

Annotation scheme for **Actions** in Wikipedia Article Translation by non-experts

- Based on the *goal* of each identified action
- Created iteratively in collaboration multiple annotators
  - Tested for accuracy by reliability annotation of a sample video with perfect agreement.

C – Information Search

B – Sentence Translation

A – Tool Management

	Action ID	Action Description
C	14	Refine machine translation input
	13	Machine translate words/phrases
	12	Refine word/phrase in dictionary search
	11	Search word/phrase in dictionary
	10	Refer to knowledge resources (Wikipedia, online articles, publications…)
	9	Scan search engine results
	8	Refine search engine search terms
	7	Search engine query
B	6	Post-edit machine translated sentence in target article
	5	Compare machine translation result to original text
	4	Machine translate full sentences
	3	Read own translation (proofread)
	2	Translate sentence
	1	Read source article
A	-1	Open resources (text editor, web pages…)
	-2	Create target language page

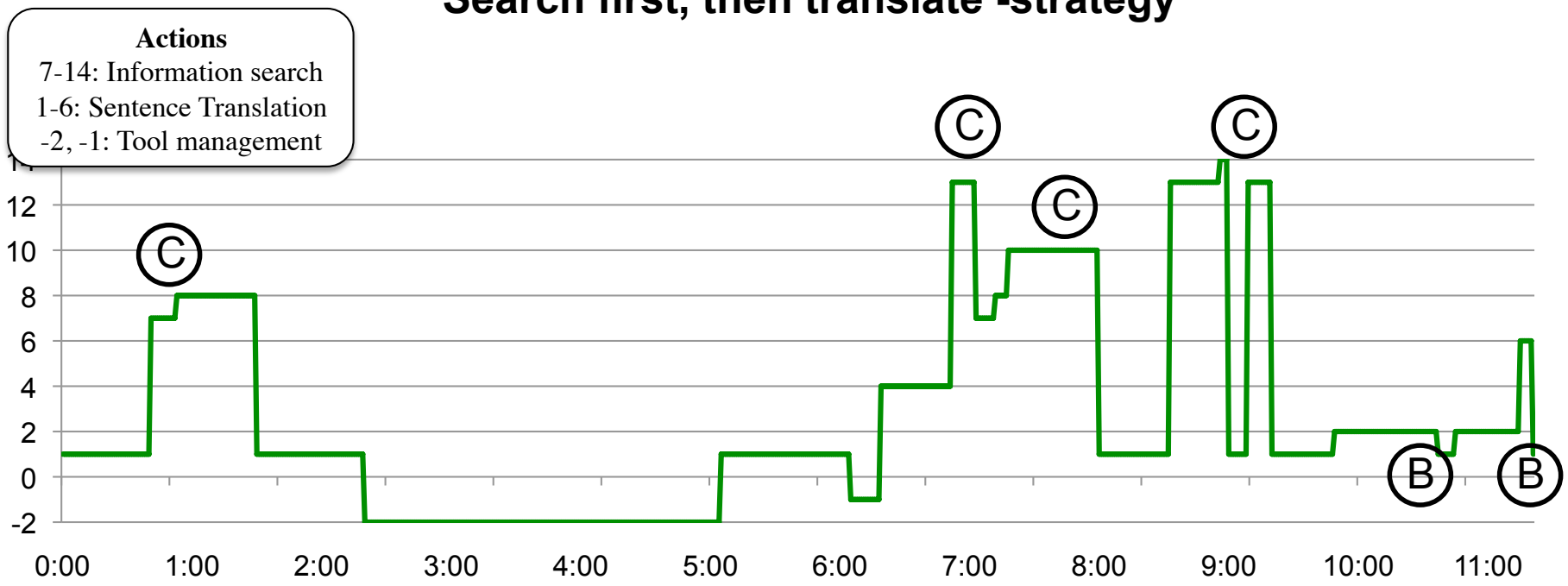
How do Non-Experts Translate Wikipedia Articles?

Activity Analysis of Wikipedia Article Translation

Ari Hautasaari – Department of Social Informatics, Kyoto University

Three different translation strategies using online resources

Search first, then translate -strategy



(B) Translate sentence by hand

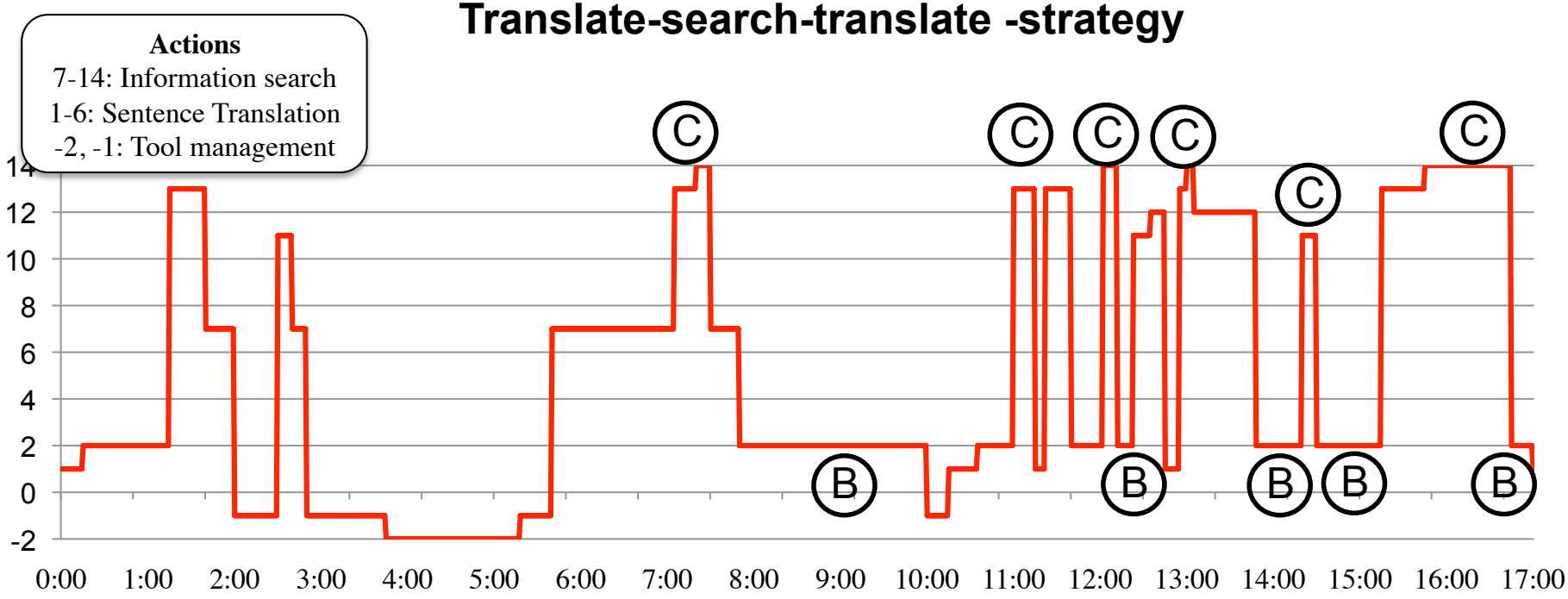
(C) Search words in dictionaries/Google/etc.

Activity Analysis of Wikipedia Article Translation

Ari Hautasaari – Department of Social Informatics, Kyoto University

Three different translation strategies using online resources

Translate-search-translate -strategy



Ⓑ Translate sentence by hand

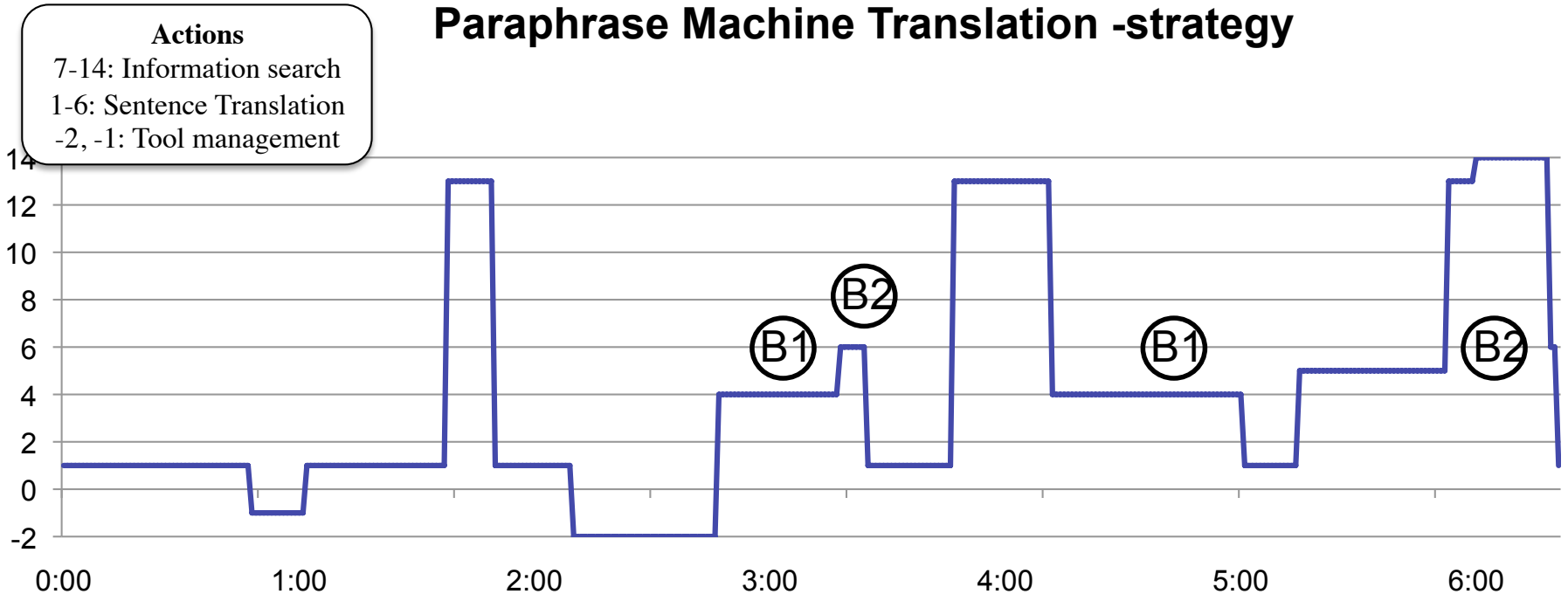
Ⓒ Search words in dictionaries/Google/etc.

Activity Analysis of Wikipedia Article Translation

Ari Hautasaari – Department of Social Informatics, Kyoto University

Three different translation strategies using online resources

Paraphrase Machine Translation -strategy



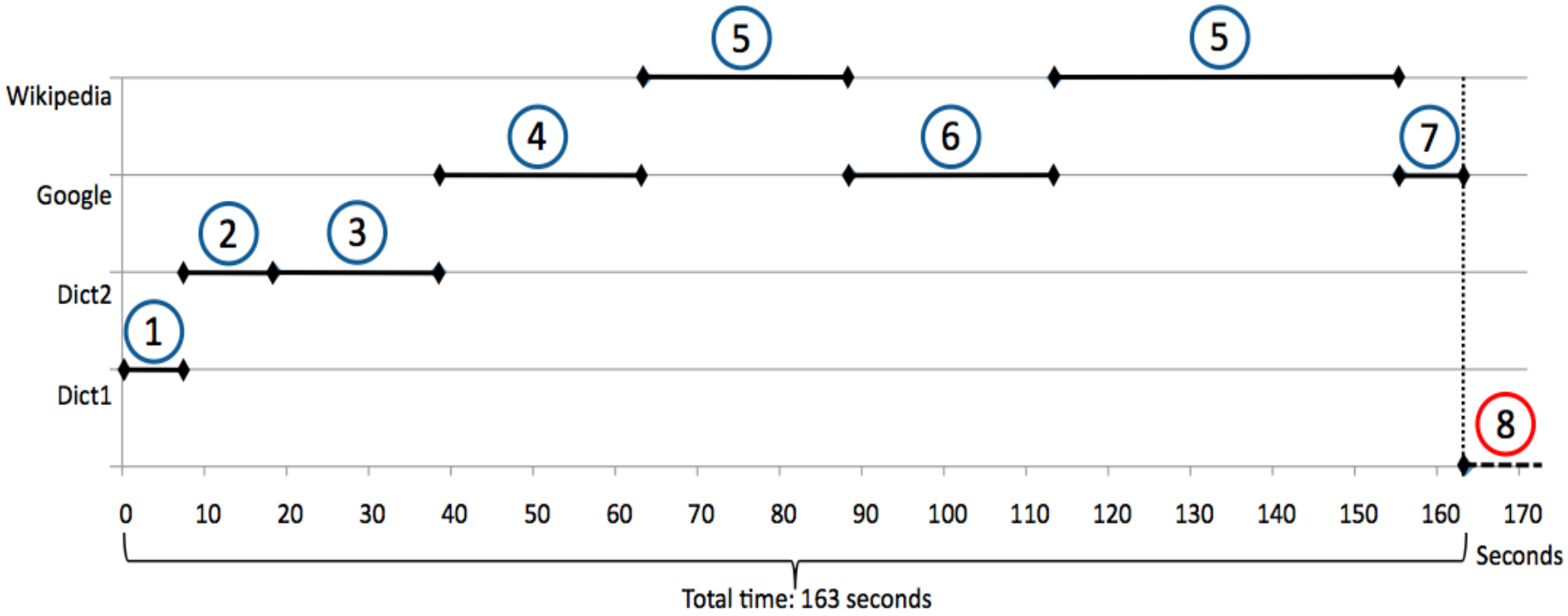
B1 Machine translate sentence

B2 Paraphrase machine translated sentence

Activity Analysis of Wikipedia Article Translation

Ari Hautasaari – Department of Social Informatics, Kyoto University

Information Search Activity (Chinese Translation for "Crowdfunding")



(1) Search "crowdfunding" in Dictionary 1 **(2)** Search "crowdfunding" in Dictionary 2 **(3)** Paraphrase to "crowdfund" in Dictionary 2 **(4)** Search Google for "crowdfunding" **(5)** Open and read Wikipedia article on "crowd funding**"** **(6)** Scan Google results on "crowdfunding" **(7)** Search "collective cooperation" in Google (8)Read source article *without finding the correct translation*

Ari Hautasaari – Department of Social Informatics, Kyoto University

Results

- On average, the translators were engaged in **single search actions** **27.23%** of the time, and in **information search activities** **72.77%** of the time, for one information item, such as one word.
- Non-expert translators tend to engage in the time-consuming information search activities only when the **first initially highly valued tool** returns an **unsatisfactory** result.

How To Support Non-Expert Translators?

Ari Hautasaari – Department of Social Informatics, Kyoto University

Experiment Setting

- 15 Chinese native speakers who are not professional translators or domain experts.
- **Task:** Create an **English – Chinese bilingual dictionary** based on one Wikipedia article.
- **Activity Analysis** based on Video Recording

Output

- 15 **domain specific bilingual dictionaries** created by non-experts.
- 5 dictionaries per article: Akan Volcanic Complex (**Object-based**), Crowdsourcing (**Conceptual**), Ying Huang (**Biographical**)

Bilingual Dictionary Creation in Wikipedia

Ari Hautasaari – Department of Social Informatics, Kyoto University

page | discussion | header | **dictionary** | setting

[Add record] [Delete record] [Add/Delete language] ** Updated*

English [↑][↓]	French
tin toy	jouet en fer-blanc
goldfish	poisson rouge
Sosuke	Sosuke
Orikaeshiten	Orikaeshiten
Miyazaki	Miyazaki
Hayao Miyazaki	Hayao Miyazaki
Ponyo	Ponyo
Ghibli	
pitch-black blackie	

Save Cancel | Upload | Download

Language Grid Extension for MediaWiki - Dictionary Creation Interface (Hautasaari et al. 2011)

Annotation scheme for **Actions** in bilingual dictionary creation by non-experts

- Based on the *goal* of each identified action
- Created iteratively in collaboration multiple annotators
  - Tested for accuracy by reliability annotation of a sample video with perfect agreement.

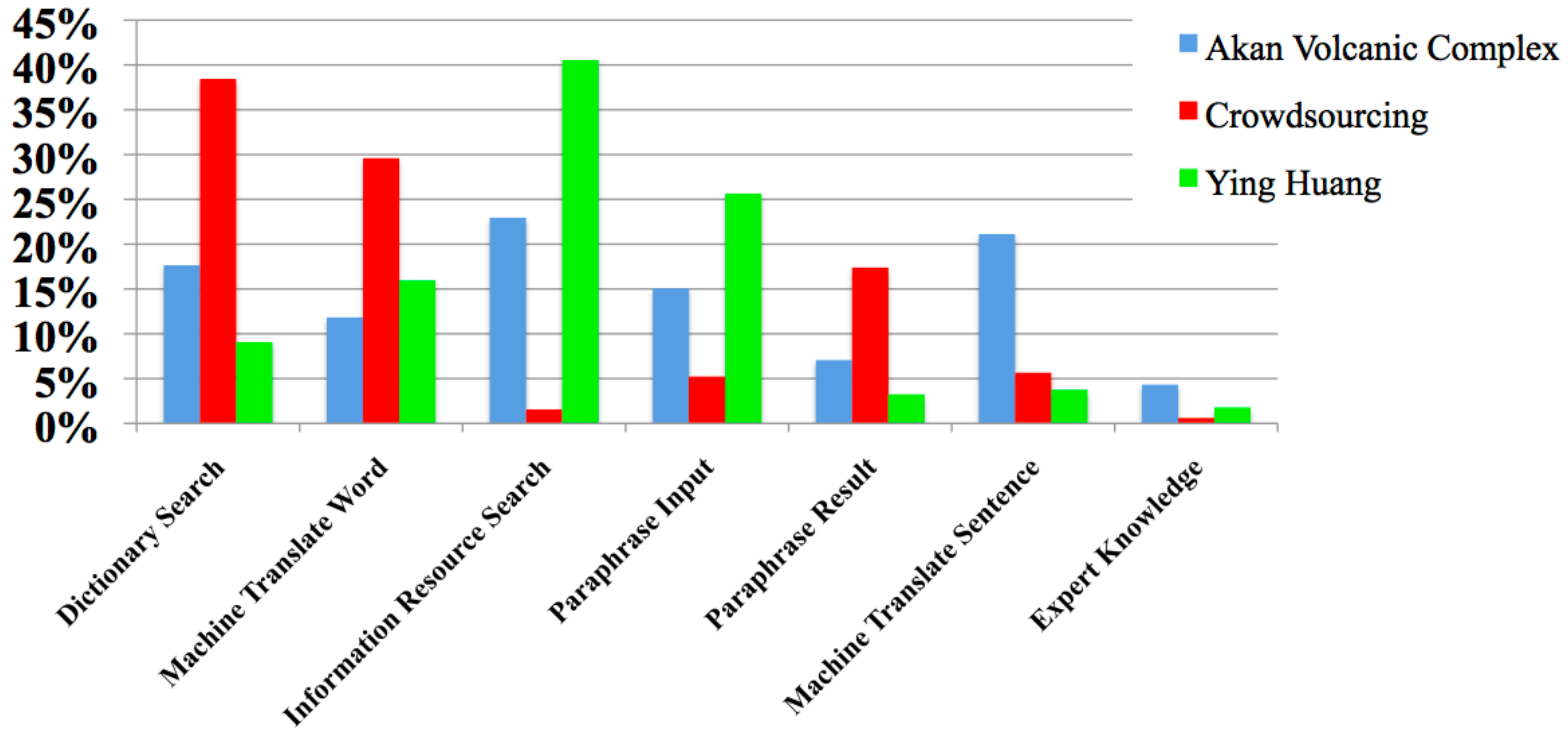
	Action ID	Action Description
	14	Expert knowledge / personal knowledge
	13	Machine translate sentence
Word search from information resources	12	Paraphrase search result
	11	Paraphrase search input
	10	Information resource search (Google, Wikipedia, etc.)
Word translation with machine translator	9	Paraphrase machine translation result
	8	Paraphrase machine translation input
	7	Machine translate word
Word search in online dictionaries	6	Paraphrase dictionary result
	5	Paraphrase dictionary input
	4	Dictionary search
Tool and article management	3	Input Chinese word to dictionary
	2	Input English word to dictionary
	1	Read source article
	-1	Open resources (text editor, web pages…)

Dictionary Creation Strategies – Activity Analysis

Ari Hautasaari – Department of Social Informatics, Kyoto University

- Overall distribution of time spent on **Actions** in all articles

- Significant difference in the distribution of actions between all three articles (Akan Volcanic Complex - Crowdsourcing ($X^2 = 21.2$, $p < .016$), Akan Volcanic Complex - Ying Huang ($X^2 = 29.2$, $p < .016$), Crowdsourcing - Ying Huang ($X^2 = 30.3$, $p < .016$))



Dictionary Creation Strategies – Activity Analysis

Ari Hautasaari – Department of Social Informatics, Kyoto University

- Effectiveness of each **Single Search action** per article type

	Object-based		Conceptual		Biographical	
Action	Success rate	Frequency	Success rate	Frequency	Success rate	Frequency
Dictionary Search	0.58	31.58%	0.35	44.44%	0.60	6.49%
Machine Translate Word	0.67	15.79%	0.73	33.33%	0.44	32.47%
Information Search (Google, Wikipedia, etc.)	0.25	10.53%	0	8.89%	0.76	44.16%
Machine Translate Sentence	0.60	13.16%	0	2.22%	0	2.60%

Ari Hautasaari – Department of Social Informatics, Kyoto University

- **Accuracy of User Created Wikipedia Dictionaries**
- **55.55% of incorrect entries are names and proper nouns**

	Article Type			
	Object-based	**Conceptual**	**Biographical**	**Total**
Unique English Entries	68	41	70	179
Unique Chinese Entries	59	41	66	166
Accuracy	0.771	0.753	0.723	0.749
Accuracy without Empty Entries	0.882	0.843	0.794	0.840

How Do The User Created Bilingual Dictionaries Help Non-Expert Translators?

Ari Hautasaari – Department of Social Informatics, Kyoto University

- Effects of User Created Bilingual Dictionaries in Article Translation (two-tailed t-test)
  - No significant effect on single word searches
    - Number of single searches between conditions ($p = 0.503$)
    - Number of single searches when Dictionary is available ($p = 0.906$)
  - No significant difference on number of Information Search Activities
    - Number of Information Search Activities between conditions ($p = 0.166$)
    - Number of Information Search Activities when Dictionary is available ($p = 0.238$)

No significant individual differences between participants

→ **Similar linguistic and translation skills**

Effects of User Created Dictionaries in Article Translation

Ari Hautasaari – Department of Social Informatics, Kyoto University

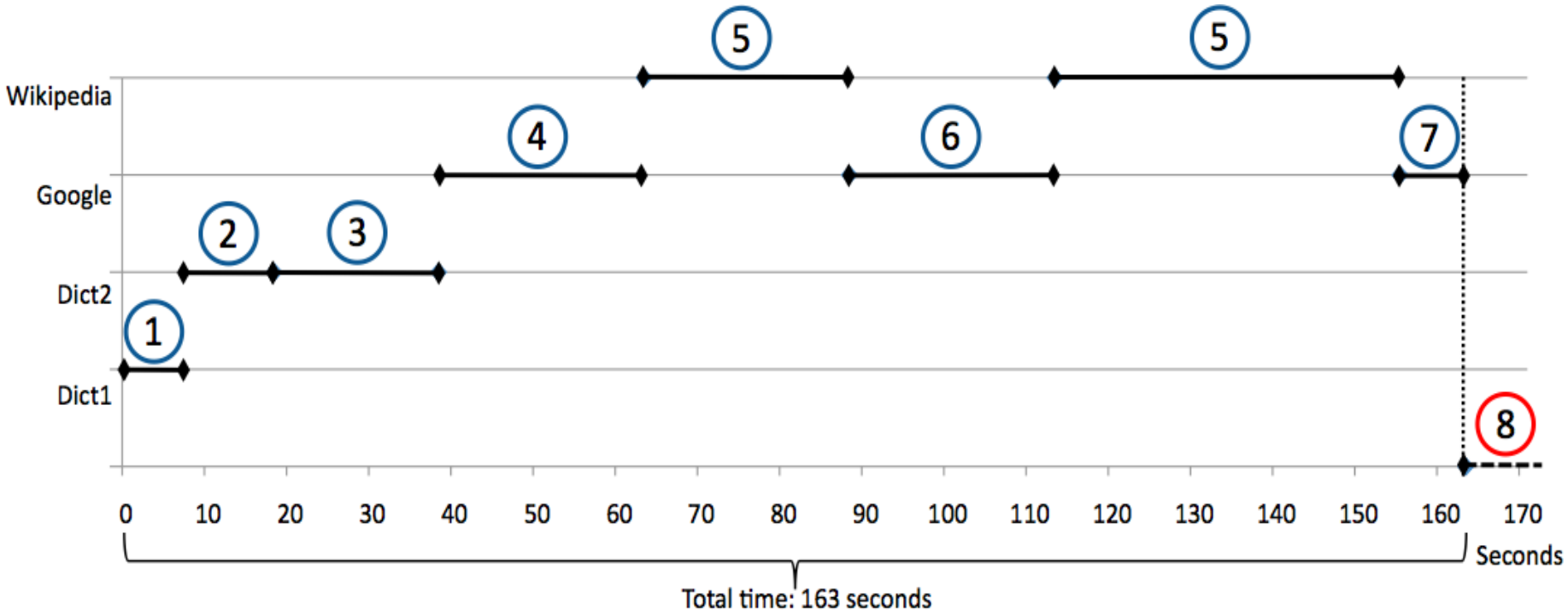
- Effects of Wikipedia Dictionaries to Article Translation
  - **Significant decrease in average time for Information Search Activities between conditions (mm:ss:ms)**

	Article Type			
	Object-based	**Conceptual**	**Biographical**	**Total**
Without Dictionary	1:22:33	1:07:37	1:08:19	1:12:49
With Dictionary	0:37:41	0:34:16	0:44:56	0:38:58
%-change	54.34%	49.31%	34.22%	46.00%
p-value	$p < .05$	$p < .05$	$p = .20$	$p < .05$

Effects of User Created Dictionaries in Article Translation

Ari Hautasaari – Department of Social Informatics, Kyoto University

Information Search Activity (Chinese Translation for "Crowdfunding")



(1) Search "crowdfunding" in Dictionary 1 **(2)** Search "crowdfunding" in Dictionary 2 **(3)** Paraphrase to "crowdfund" in Dictionary 2 **(4)** Search Google for "crowdfunding" **(5)** Open and read Wikipedia article on "crowd funding**"** **(6)** Scan Google results on "crowdfunding" **(7)** Search "collective cooperation" in Google (8)Read source article *without finding the correct translation*

Effects of User Created Dictionaries in Article Translation

Ari Hautasaari – Department of Social Informatics, Kyoto University

- Effects of Wikipedia Dictionaries to Article Translation
  - **Significant decrease in average time for Information Search Activities between conditions (mm:ss:ms)**

	Article Type			
	Object-based	**Conceptual**	**Biographical**	**Total**
Without Dictionary	1:22:33	1:07:37	1:08:19	1:12:49
With Dictionary	0:37:41	0:34:16	0:44:56	0:38:58
%-change	54.34%	49.31%	34.22%	46.00%
p-value	$p < .05$	$p < .05$	$p = .20$	$p < .05$

Effects of User Created Dictionaries in Article Translation

Ari Hautasaari – Department of Social Informatics, Kyoto University

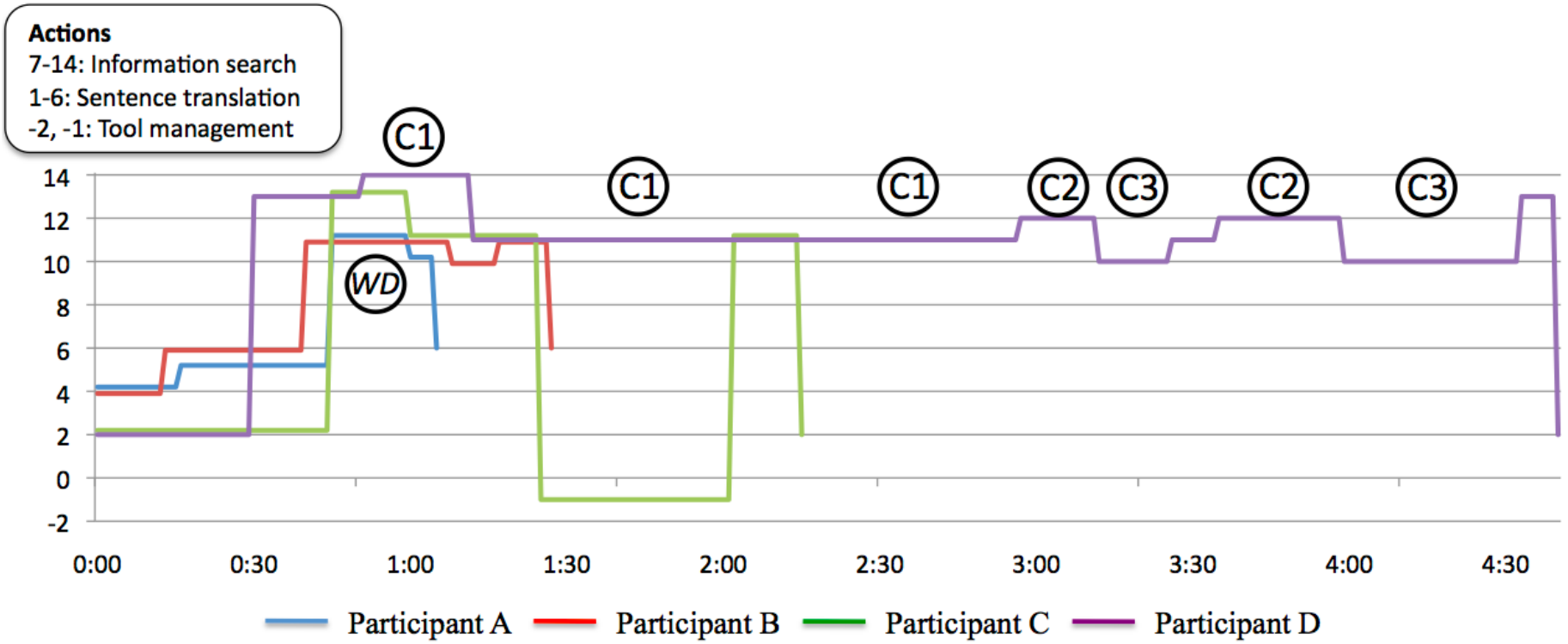
- Effects of Wikipedia Dictionaries to Article Translation
  - **Significant decrease in average translation time (mm:ss)**

	Article Type			
	Object-based	**Conceptual**	**Biographical**	**Total**
Without Dictionary	85:59	85:08	70:49	80:39
With Dictionary	63:53	61:53	57:44	61:10
%-change	34.59%	37.59%	22.66%	31.85%
p-value	$p < .05$	$p < .05$	$p = .20$	$p < .05$

Effects of User Created Dictionaries in Article Translation

Ari Hautasaari – Department of Social Informatics, Kyoto University

- Information Search activity with User Created Dictionary
- Behavior changes even though the dictionary entry is wrong



Effects of User Created Dictionaries in Article Translation

Ari Hautasaari – Department of Social Informatics, Kyoto University

- Non-expert translators adopt **3 strategies** when translating Wikipedia articles
- The most time-consuming task is **Information Search Activity**

- **Overall accuracy** of user created dictionaries was **.840,** where 55.55% of **incorrect** entries are proper names and nouns
- User created bilingual dictionaries:
  - **Significantly reduced the time spent on Information Search activities**
    - **(avg. 46.00%)**
  - **Significantly reduced the overall translation time**
    - **(avg. 31.85%)**

Tool Design Based on Activity Analysis

Activity Analysis of Wikipedia Article Translation

Ari Hautasaari – Department of Social Informatics, Kyoto University

Design Implications for Wikipedia Translation Support Tool Based on The Activity Analysis (Hautasaari, CSCW 2013)

1. **Design for Multiple Activities**
   - Support multiple activities (information search, translation, etc.) in the Wikipedia translation process with the same tool.
2. **Design for Continuous Action**
   - Decrease the time spent changing views between actions and activities.
3. **Design for Context Awareness**
   - Let the system/user define the context of words based on article/category domain to avoid word polysemy.
4. **Design for Knowledge Sharing**
   - Allow translators to share previously acquired knowledge through repositories for reuse in same domain translations.

Wikipedia Dictionary Design Based on Activity Analysis

Ari Hautasaari – Department of Social Informatics, Kyoto University

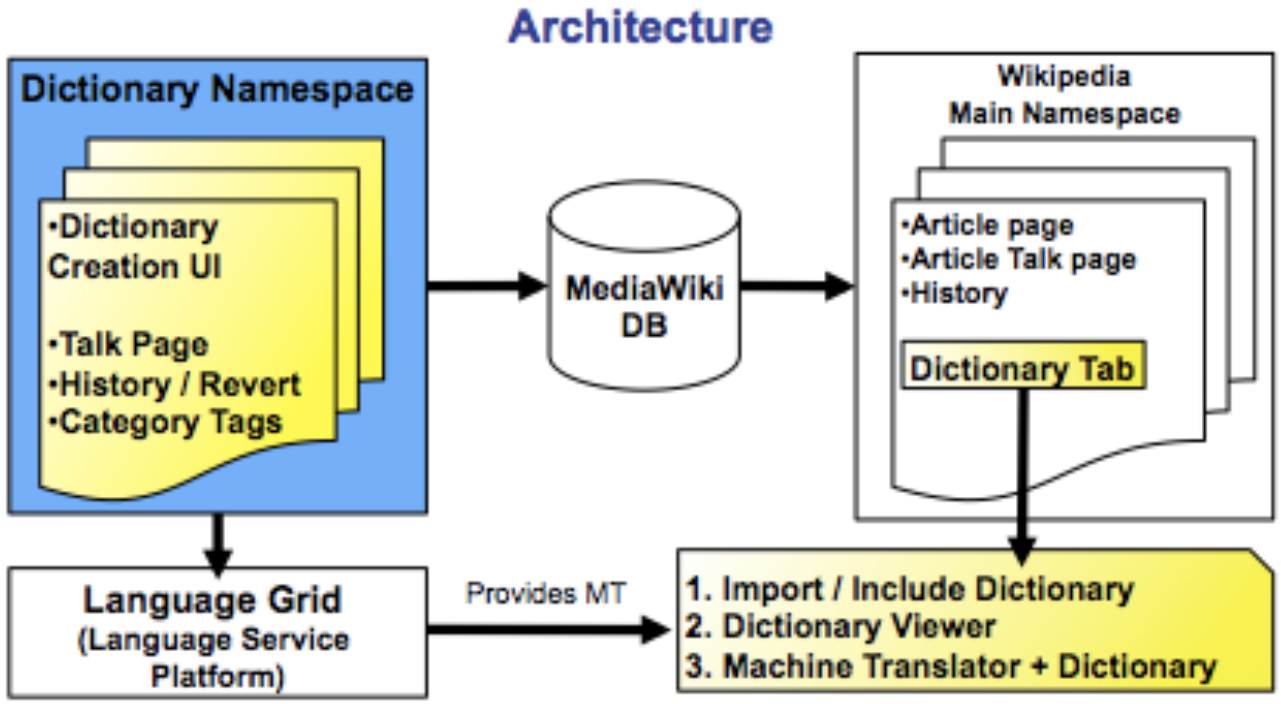
I. MediaWiki Extension that creates a Dictionary namespace to Wiki-sites

II. Domain specific dictionaries as individual pages (e.g. Dictionary:Football)

III. Discussion and history/revert function in dictionary editing

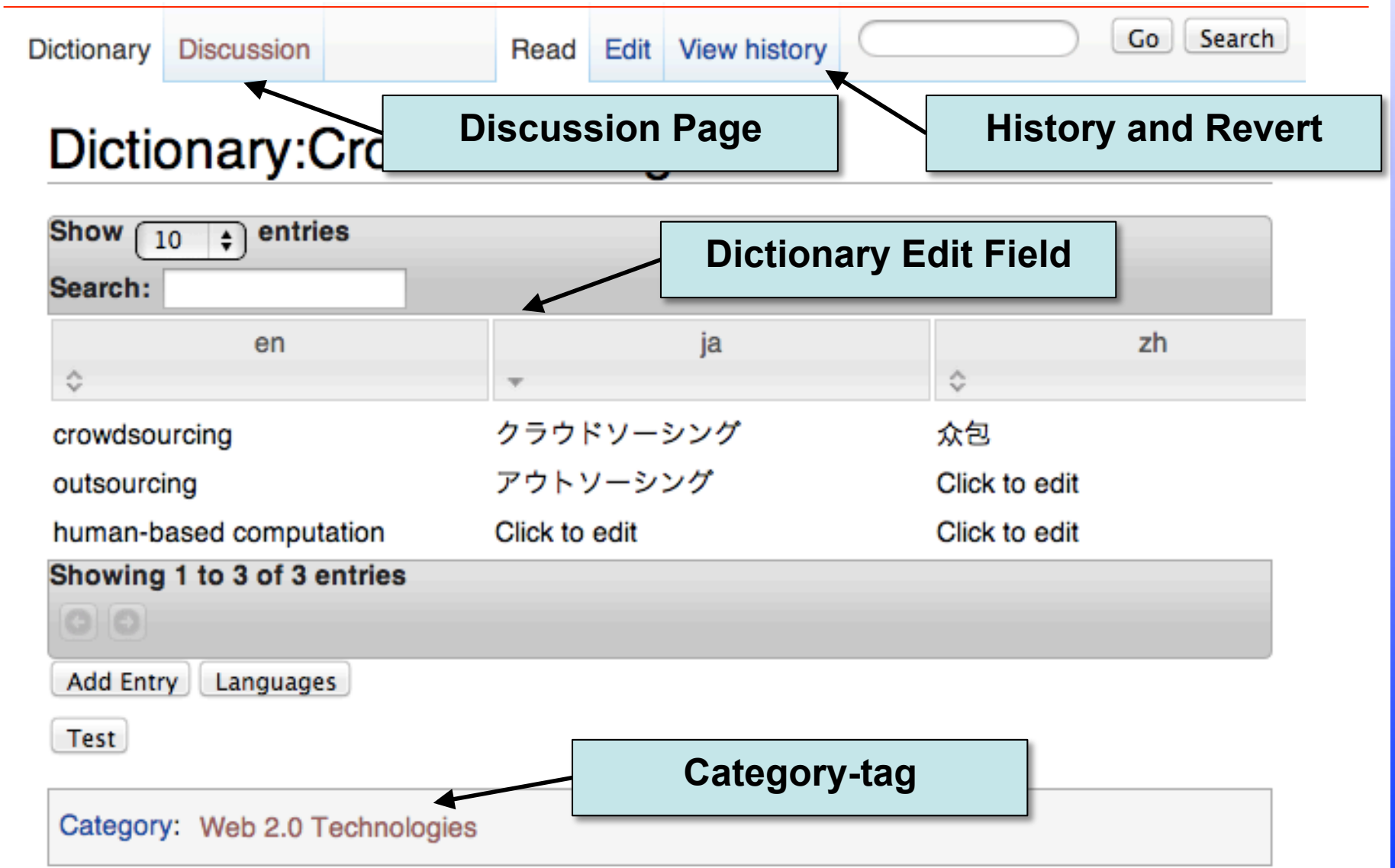
IV. Save dictionaries in Wiki-syntax (XML). Import to articles via namespace search

V. Combine with machine translators in the Dictionary tab to improve machine translation quality of sentences



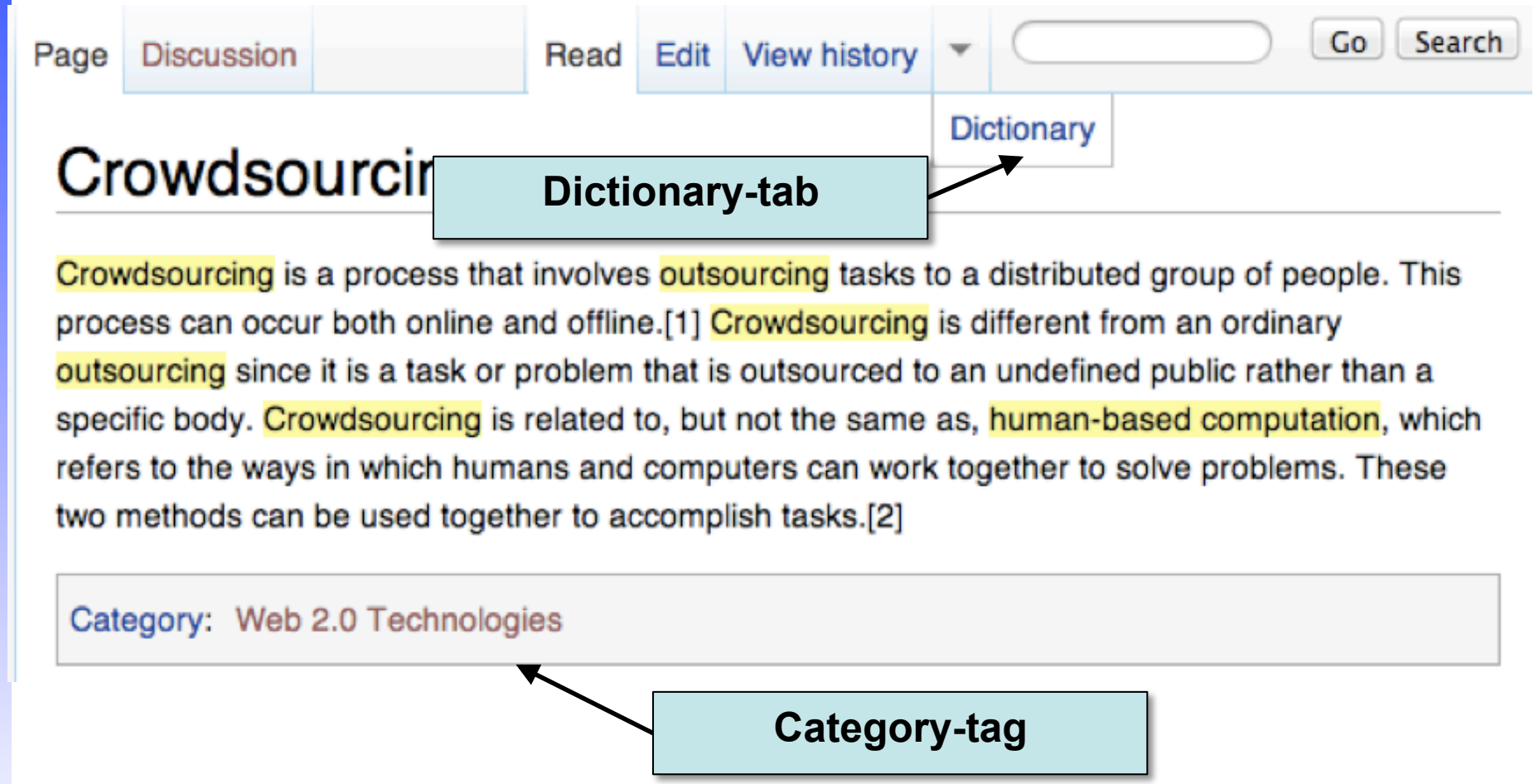
Dictionary Creation Interface

Ari Hautasaari – Department of Social Informatics, Kyoto University



Dictionary Interface in Wikipedia Articles

Ari Hautasaari – Department of Social Informatics, Kyoto University



Dictionary Interface in Wikipedia Articles

Ari Hautasaari – Department of Social Informatics, Kyoto University

Dictionaries for Crowdsourcing

Include New Dictionaries

Web_2.0_Technologies

- Crowdsourcing +

Others

- Abccd +
- Classical Music +

Automatically add user created dictionaries in the same domain (Category)

Included Dictionaries

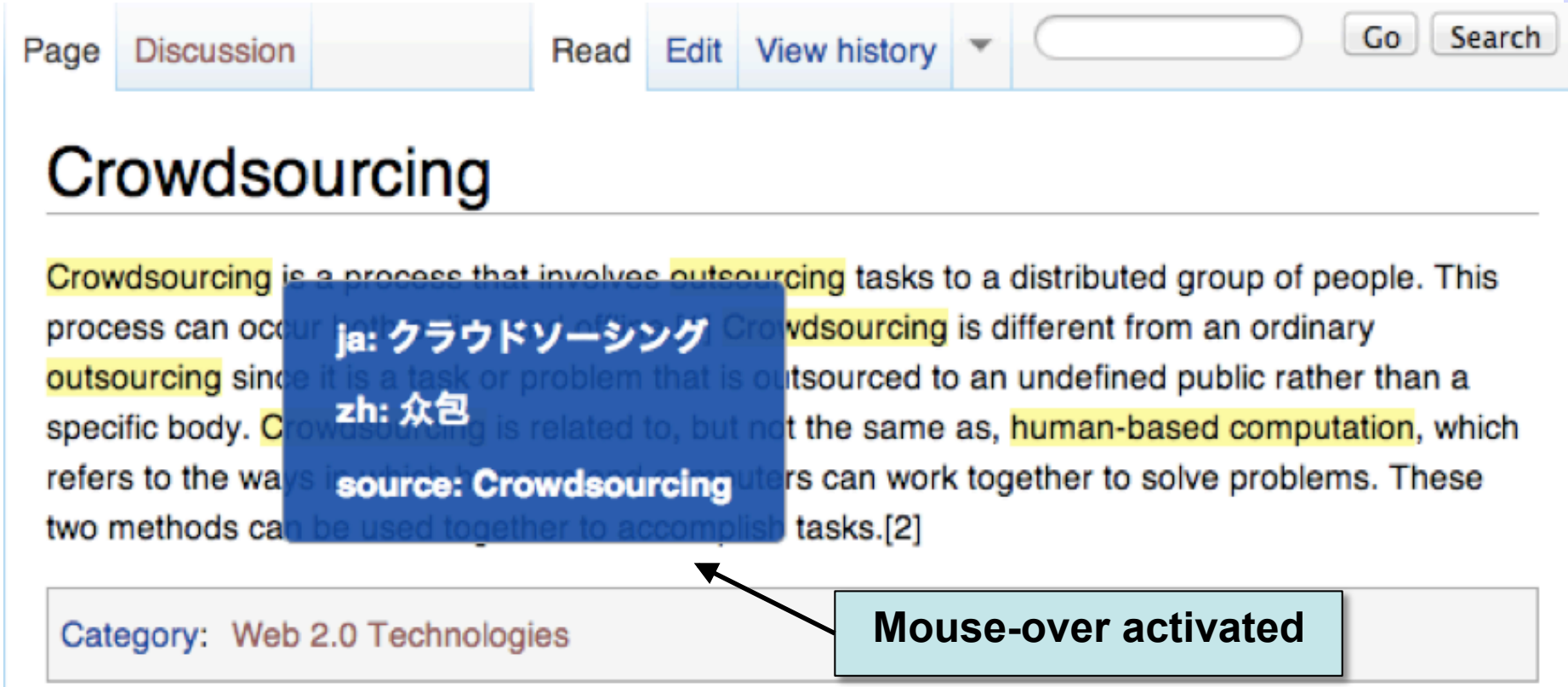
- Crowdsourcing -

Contents

en	ja	zh	Source
crowdsourcing	クラウドソーシング	众包	Dictionary:Crowdsourcing
outsourcing	アウトソーシング		Dictionary:Crowdsourcing
human-based computation			Dictionary:Crowdsourcing

Dictionary Interface in Wikipedia Articles

Ari Hautasaari – Department of Social Informatics, Kyoto University



Ari Hautasaari – Department of Social Informatics, Kyoto University

Lessons to take home:

- Activity Analysis is a very effective tool to analyze:
  - Emergent novel work practices
  - Effects of novel supporting tools in laboratory settings

- Activity Analysis informs the design of supporting tools for complex collaborative work (such as article translation) based on structured analysis of work practices.
  - As opposed to trial and error.

Thank you!