文部科学省科学研究費補助金基盤研究(A)

研究成果報告書（平成 21~23 年度）

（課題番号：21240014）

サービスコンピューティングに基づく集合知の研究

2012年3月

研究代表者 石田 亨（京都大学情報学研究科社会情報学専攻）

まえがき

インターネット上の多言語基盤をサービス指向の集合知で形成するという，言語グリッドのアイデアは2005年に1年間をかけた検討を経て生まれたものである.そのアイデアは,2006年1月のSAINTの招待講演で発表されている.

その後，2006年4月より，NICTで5年間の言語グリッドプロジェクトが始まり，基盤ソフトウェアの開発が行われた．そのソフトウェアを用いて，2007年12月に京都大学情報学研究科社会情報学専攻で運営が開始され，現在に至っている．その間に運営方式は，初期の単独組織から，複数の組織が連携する連邦制運営へと進化している．現時点では，バンコクの NECTEC，ジャカルタのインドネシア大学に運営組織が立ち上がり，京都大学の運営組織と相互に連携が行われている．

言語グリッド構築の動機は，2001年の9.11の直後に行われた異文化コラボレーション実験に遡る．機械翻訳を用いた日中韓馬の共同実験の際に，その実験にカスタマイズされた多言語環境を構築したのだが，その作業は容易ではなかった．言語グリッドの着想が言語処理研究の出口としてではなく，異文化コラボレーション環境の実現を容易にするためのものであったことが，その後のプロジェクトの性格を決定づけている．ソフトウェア開発を行うNICT，運営を行う京都大学に加え，プロジェクトの当初から異文化コラボレーション環境を必要とするNPO/NGOや大学研究室が言語グリッドアソシエーションを形成し,開発に参加した．

基盤研究「サービスコンピューティングに基づく集合知の研究」が実施された2009年～2011 年は，言語グリッドの初期開発が一段落し，運営が軌道に乗り始めた頃であった．言語グリッドは，開発，運営，利用が連携したプロジェクトであることは既に述べたが，本基盤研究は，その水先案内としての研究を担当している．大学の研究室で博士課程や修士課程の学生が様々に行う研究は，利用現場で生じる問題を先取りし，開発の効率を高める．一方，学生にとっては，望まれる研究を行っているという手ごたえを感じることができる．

以下の報告は 2 部に分かれる．第一部は言語グリッドプロジェクト全体の報告であり，第二部は本基盤研究の主要成果の論文からなる．なお，言語グリッドの成果は，Springerから “The Language Grid: Service-Oriented Collective Intelligence for Language Resource Interoperability” と題する書籍として出版した．本基盤研究の成果が該当する章を記載したので併せて参照いただきたい．最後に，本基盤研究の拠り所となった言語グリッドを開発・運営いただいたNICTと，京都大学 情報学研究科 社会情報学専攻に感謝する．

2012年3月
石田 亨 (研究代表者)

研究組織

研究代表者：	石田亨	(京都大学情報学研究科社会情報学専攻)
研究分担者：	松原繁夫	(京都大学情報学研究科社会情報学専攻)
研究分担者：	服部 宏充	(京都大学情報学研究科社会情報学専攻)

研究経費

平成 21 年度	15,990 千円
平成 22 年度	15,080 千円
平成 23 年度	15,990 千円
合計	47,060 千円

研究発表

(1) 著書・編書

1. Toru Ishida Ed. *The Language Grid: Service-Oriented Collective Intelligence for Language Resource Interoperability*. Springer, 2011. ISBN 978-3-642-21177-5.

（本研究課題に関係する章は以下の通り）

Chapter 5　Service Supervision for Runtime Service Management
Masahiro Tanaka, Toru Ishida, and Yohei Murakami

Chapter 7　Cascading Translation Services
Rie Tanaka, Yohei Murakami, and Toru Ishida

Chapter 12　Conversational Grounding in Machine Translation Mediated Communication
Naomi Yamashita and Toru Ishida

Chapter 13　Humans in the Loop of Localization Processes
Donghui Lin

Chapter 14　Collaborative Translation Protocols
Daisuke Morita and Toru Ishida

Chapter 15　Multi-Language Discussion Platform for Wikipedia Translation
Ari Hautasaari, Toshiyuki Takasaki, Takao Nakaguchi, Jun Koyama, Yohei Murakami, and Toru Ishida

Chapter 16　Pipelining Software and Services for Language Processing
Arif Bramantoro, Ulrich Schäfer, and Toru Ishida

(2) ジャーナル

1. 石田 亨, 村上陽平, 稲葉利江子, 林 冬惠, 田仲正弘. 言語グリッド：サービス指向の多言語基盤, 電子情報通信学会論文誌 D, Vol.J95-D, No.1, pp.2-10, 2012.（招待論文）
2. 石田憲幸, 高崎俊之, 石松昌展, 石田 亨. Wikipedia 翻訳のための多言語議論の支援. 電子情報通信学会論文誌 D, Vol.J95-D, No.1, pp.39-46, 2012.

3. 石田　亨, 村上　陽平. サービス指向集合知のための制度設計. 電子情報通信学会論文誌 D Vol.J93-D, No.6, pp. 675-682, 2010. （招待論文）
4. Tomoko Koda, Toru Ishida, Matthias Rehm and Elisabeth André. Avatar Culture: Cross-Cultural Evaluations of Avatar Facial Expressions. *AI & Society*, Vol.24, No.3, Springer, pp. 237-250, 2009.
5. 稲葉利江子, 山下直美, 石田　亨, 葛岡英明. 機械翻訳を用いた 3 言語間コミュニケーションの相互理解の分析. 電子情報通信学会論文誌, Vol. J92-D, No.6, pp. 747-757, 2009.
6. 森田大翼, 石田　亨. 共同翻訳のためのプロトコルの開発. 電子情報通信学会論文誌, Vol. J92-D, No.6, pp. 739-746, 2009.
7. 境　智史, 後藤雅樹, 村上陽平, 森本智史, 石田　亨. 言語グリッドプレイグラウンド: 軽量の構成部品を用いた異文化コラボレーション環境. ヒューマンインタフェース学会論文誌, Vol. 11, No. 1. pp. 115-123, 2009.
8. 田仲正弘, 石田　亨. 複合 Web サービスの実行可能性予測. 情報処理学会論文誌, Vol.50, No. 2. pp. 701-708, 2009.

(3) 国際会議およびシンポジウム・ワークショップ

1. Ari Hautasaari. Analysis of Discussion Contributions in Translated Wikipedia Articles An Intercultural Collaboration Experiment. *3rd international conference on intercultural collaboration (ICIC-12),* ACM, 2012.
2. Donghui Lin, Toru Ishida, Yohei Murakami, and Masahiro Tanaka. Improving Service Processes with the Crowds. *9th International Conference on Service Oriented Computing (ICSOC-2011)*, industry track, Paphos, Cyprus, December 6th 2011.
3. Noriyuki Ishida, Toshiyuki Takasaki, Masanobu Ishimatsu and Toru Ishida. Supporting Multilingual Discussion for Wikipedia Translation. *International Conference on Culture and Computing (Culture and Computing 2011)*, poster session, pp.129-130, Kyoto, Japan, October 21th 2011.
4. Julien Bourdon and Toru Ishida. A Graph Based Model for Understanding Localisation Patterns in Multilingual Websites. *International Conference on Culture and Computing (Culture and Computing-11)*, poster session, Kyoto, Japan, October 22nd 2011.
5. Ari Hautasaari and Toru Ishida. Discussion About Translation in Wikipedia. *International Conference on Culture and Computing (Culture and Computing-11)*, poster session, Kyoto, Japan, October 21th 2011.
6. Linsi Xia, Naomi Yamashita and Toru Ishida. Analysis on Multilingual Discussion for Wikipedia Translation. *International Conference on Culture and Computing (Culture and Computing-11)*, Kyoto, Japan, October 21th 2011.
7. Jun Matsuno and Toru Ishida. Constraint Optimization Approach to Context Based Word

Selection. *International Joint Conference on Artificial Intelligence (IJCAI-11)*, pp. 1846-1851, Bercelona, Spain, July 20th 2011.

8. Arif Bramantoro and Toru Ishida. Cultural Language Service: A Discovery, Composition and Organization. *IEEE International Conference on Services Computing (SCC-11)*, pp.402-409, Washington DC, USA, July 8th 2011.
9. Shinsuke Goto, Yohei Murakami and Toru Ishida. Reputation-Based Selection of Language Services. *IEEE International Conference on Services Computing (SCC-11)*, pp.330-337, Washington DC, USA, July 6th 2011.
10. Ari Hautasaari, Nadia Bouz-Asal, Rieko Inaba, Toru Ishida. Intercultural Collaboration with the Language Grid Toolbox. *The 2011 ACM Conference on Computer Supported Cooperative Work (CSCW-2011) Videos,*pp.579-580, Hangzhou, China, March 23rd 2011.
11. Nadia Bouz-Asal, Rieko Inaba, Toru Ishida. Analyzing patterns in composing teaching materials from the Web. *The 2011 ACM Conference on Computer Supported Cooperative Work (CSCW-2011) Interactive papers,* pp.605-608, Hangzhou, China, March 21st 2011.
12. Donghui Lin, Masahiro Tanaka, Yohei Murakami, Toru Ishida. Language Grid Toolbox for Customized Multilingual Communities. *The 2011 ACM Conference on Computer Supported Cooperative Work (CSCW-2011) Demonstrations,* pp.747-748, Hangzhou, China, March 21st 2011.
13. Ari Hautasaari. Machine Translation Effects on Group Interaction: An Intercultural Collaboration Experiment. *International Conference on Intercultural Collaboration (ICIC-10), ACM, pp. 69 - 78.* August 19th, 2010.
14. Masahiro Tanaka, Yohei Murakami, Donghui Lin and Toru Ishida. Service Supervision for Service-oriented Collective Intelligence. *IEEE International Conference on Services Computing (SCC-10)*, pp.154-161, July 7th, 2010.
15. Yohei Murakami, Naoki Miyata and Toru Ishida. Market-Based QoS Control for Voluntary Services. *IEEE International Conference on Services Computing (SCC-10)*, pp. 370-377, July 7th, 2010.
16. Toru Ishida. The Language Grid for Intercultural Collaboration. *Web Science Conference (WebSci-10)*, April 27th, 2010.
17. Arif Bramantoro, Ulrich Schäfer and Toru Ishida. Towards an Integrated Architecture for Composite Language Services and Components. *International Conference on Language Resources and Evaluation (LREC-10)*, pp.3506-3511, May 21st, 2010.
18. Yohei Murakami, Donghui Lin, Masahiro Tanaka, Takao Nakaguchi and Toru Ishida. Language Service Management with the Language Grid. *International Conference on Language Resources and Evaluation (LREC-10)*, May 21st, 2010.
19. Donghui Lin, Yoshiaki Murakami, Toru Ishida, Yohei Murakami and Masahiro Tanaka.

Composing Human and Machine Translation Services: Language Grid for Improving Localization Proces. *International Conference on Language Resources and Evaluation (LREC-10)*, May 19th, 2010.

20. Toru Ishida and Yohei Murakami. Federated Operation for Service-Oriented Language Resource Sharing. *FLaReNet Forum*, Position Paper, Barcelona, Catalonia, Spain, February 12th, 2010.
21. Mika Yasuoka, Toru Ishida, Yohei Murakami, Donghui Lin, Masahiro Tanaka and Rieko Inaba. Supporting Local Jargon in Multilingual Collaboration. *International Conference on Computer Supported Cooperative Work (CSCW-10)*, demo session, pp.553-554, February 8th, 2010.
22. Toru Ishida, Rieko Inaba, Yohei Murakami, Tomohiro Shigenobu, Donghui Lin and Masahiro Tanaka. The Language Grid: Creating Customized Multilingual Environments. International Conference on Global Interoperability for Language Resources *(ICGL-10)*, January 19th, 2010.
23. Daisuke Morita and Toru Ishida. Designing Protocols for Collaborative Translation. *International Conference on Principles of Practice in Multi-Agent Systems (PRIMA-09), Lecture Notes in Artificial Intelligence*, 5925, Springer-Verlag, pp. 17-32, Nagoya, Japan, December 14th, 2009.
24. Julien Bourdon, Laurent Vercouter and Toru Ishida. A Multiagent Model for Provider-Centered Trust in Composite Web Services. *International Conference on Principles of Practice in Multi-Agent Systems (PRIMA-09), Lecture Notes in Artificial Intelligence*, 5925, Springer-Verlag, pp. 216-228, Nagoya, Japan, December 14th, 2009.
25. Donghui Lin, Yoshiaki Murakami, Toru Ishida, Yohei Murakami and Masahiro Tanaka. Lessons Learned from Composing Web Services and Human Activities. International Joint Conference on Service Oriented Computing (ICSOC-09), *Industry Session*, Stockholm, Sweden, November 25th, 2009.
26. Masahiro Tanaka, Toru Ishida, Yohei Murakami and Donghui Lin. Service Supervision Patterns: Reusable Adaption of Composite Services. *In Proceedings of International Conference on Cloud Computing (CLOUDCOMP-09)*, Springer, Munich, Germany, October 21st, 2009.
27. Arif Bramantoro and Toru Ishida. User-Centered QoS in Combining Web Services for Interactive Domain. *In Proceedings of International Conference on Semantics, Knowledge and Grid (SKG-09)*, IEEE, pp.41-48, Zhuhai, China, October 12th -14th, 2009.
28. Satoshi Morimoto, Satoshi Sakai, Masaki Gotou, Heeryon Cho, Toru Ishida and Yohei Murakami. Building Blocks: Layered Components Approach for Accumulating High-Demand Web Services. *In Proceedings of IEEE/ACM/WIC International Conference on Web Intelligence (WI-09),* short paper, IEEE Computer Society, pp.430-433, Milano, Italy, September 15th -18th, 2009.

29. Rie Tanaka, Yohei Murakami and Toru Ishida. Context-Based Approach for Pivot Translation Services. *In Proceedings of International Joint Conference on Artificial Intelligence (IJCAI-09),* AAAI Press, pp.1555-1561, Pasadena, California, USA, July 16th, 2009.
30. Rie Tanaka, Toru Ishida and Yohei Murakami. Towards Coordination of Multiple Machine Translation Services. *JSAI 2008 Conference and Workshops*, Revised Selected Papers, *Lecture Notes in Artificial Intelligence*, 5447, Springer-Verlag, pp. 73-86, Asahikawa, Japan, June 11th -13th, 2009.
31. Heeryon Cho, Naomi Yamashita and Toru Ishida. Towards Culturally-Situated Agent Which Can Detect Cultural Differences. *Pacific Rim International Conference on Multi-Agents (PRIMA-07), Lecture Notes in Artificial Intelligence*, 5044, Springer-Verlag, pp. 458-463, Bangkok, Thailand, 2009.
32. Naomi Yamashita, Rieko Inaba, Hideaki Kuzuoka and Toru Ishida. Difficulties in Establishing Common Ground in Multiparty Groups using Machine Translation. *In Proceedings of International Conference on Human Factors in Computing Systems (CHI-09)*, ACM, pp. 679-688, Boston, USA, April 7th, 2009.
33. Yumiko Mori, Rieko Inaba, Toshiyuki Takasaki and Toru Ishida. Patterns in Pictogram Communication. *In Proceedings of International Workshop on Intercultural Collaboration (IWIC-09)*, *Poster Session,* ACM, pp. 277-280, Palo Alto, California, USA, February 21st, 2009.
34. Satoshi Sakai, Masaki Gotou, Satoshi Morimoto, Daisuke Morita, Masahiro Tanaka, Toru Ishida and Yohei Murakami. Language Grid Playground: Light Weight Building Blocks for Intercultural Collaboration. *In Proceedings of International Workshop on Intercultural Collaboration (IWIC-09)*, *Poster Session,* ACM, pp. 297-300, Palo Alto, California, USA, February 21st, 2009.
35. Heeryon Cho, Toru Ishida, Naomi Yamashita, Tomoko Koda and Toshiyuki Takasaki. Human Detection of Cultural Differences in Pictogram Interpretations. *In Proceedings of International Workshop on Intercultural Collaboration (IWIC-09)*, ACM, pp. 165-174, Palo Alto, California, USA, February 21st, 2009.
36. Daisuke Morita and Toru Ishida. Collaborative Translation by Monolinguals with Machine Translators. *In Proceedings of International Conference on Intelligent User Interfaces (IUI-09)*, *Poster Session*, ACM, pp. 361-366, Sanibel Island, Florida, USA, February 8th-11th, 2009.

(4) 解説

1. Toru Ishida. Intercultural Collaboration Using Machine Translation. *IEEE Internet Computing*, pp. 26-28, 2010.
2. 石田 亨. コミュニティと機械翻訳の出会い. 人工知能学会誌, Vol. 24, No. 1, pp. 88-94, 2009 年 1 月 1 日.

(5) 新聞

1. 「異文化つなぐ「言語グリッド」試み本格化」, 日本経済新聞, 2011 年 3 月 21 日.
2. 「ネット多言語システム　タイ研究機関と連携」, 京都新聞, 2011 年 2 月 15 日.
3. 「言語グリッドを連携運営」, 日刊工業新聞, 2011 年 2 月 15 日.
4. 「京大の翻訳サービス　タイの研究所と提携」, 産経新聞, 2011 年 2 月 15 日.
5. 「留学生も快適に・・・「誤訳」少ない翻訳サービス」, 産経新聞, 2011 年 2 月 10 日.
6. 「"京大用語"正しく翻訳」, 京都新聞, 2011 年 2 月 10 日.
7. 「京大留学生に必須語翻訳」, 読売新聞, 2011 年 2 月 10 日.
8. 「ネット介して 8 言語を翻訳，商店やホテル向け」, 日経新聞, 2010 年 3 月 16 日.
9. 「ネットで瞬時に多言語翻訳」, 京都新聞, 2010 年 3 月 13 日.
10. 「多言語使って外国人と交流」, 朝日新聞, 2010 年 3 月 13 日.
11. 「４か国語即翻訳サイト 京大 留学生の生活支援目指す」, 読売新聞, 2010 年 2 月 20 日, 朝刊 京都 35 面.
12. 「京大 自動翻訳 留学生向け」, 京都新聞, 2010 年 2 月 16 日, 朝刊 23 面.
13. 「NICT 多言語コラボレーション支援ツールを OSS 公開 4 つの汎用的な多言語モジュールも同時開示」, 電波タイムズ, 2010 年 1 月 25 日, 1 面.
14. 「NICT 多言語コラボ支援 汎用ツールを OSS で」, 日本情報産業新聞, 2010 年 1 月 25 日, 朝刊 2 面.
15. 「文化とコンピューティング国際会議」, 日本経済新聞, 2010 年 1 月 25 日, 夕刊 9 面.
16. 「多言語交流 ソフトで支援」, 京都新聞, 2010 年 1 月 21 日, 朝刊 25 面.
17. 「情報通信研究機構など支援ソフト　翻訳辞書ソフト取り込み容易に」, 日経産業新聞［日経テレコン 21］, 2010 年 1 月 19 日, 朝刊 11 面.
18. 「情報通信研究機構と京都大学が言語グリッドツール OSS として開発公開 多言語コラボレーションを支援」, 電経新聞，2010 年 1 月 18 日, 朝刊 4 面.
19. 「情報通信研究機構など多言語機能容易に 支援ツール OSS で公開」, 日刊工業新聞, 2010 年 1 月 15 日, 朝刊 11 面.
20. 「IT で京文化発信を」, 京都新聞, 2010 年 1 月 14 日, 朝刊 25 面.

(6) 雑誌

1. 稲葉利江子, 村上陽平, 田仲正弘, 林冬惠, 石田亨. 言語グリッドを用いたスマート翻訳―京大翻訳!―, AAMT, Vol.49, 2011.
2. 石田 亨. 留学生が担う研究活動. 日本語学, 2009年5月臨時増刊号, pp. 207-214, 2009年5月15日.

(7) TV

1. 「多言語交流支援システムが完成」, 京プラス, KBS 京都, 2010 年 3 月 12 日（金）.
2. 「多言語交流支援システム」, 京 bizW , KBS 京都, 2010 年 3 月 12 日（金）.
3. 「自動翻訳システムの実験」, 京いちにち, NHK 京都, 2010 年 3 月 12 日（金）.

第一部　言語グリッドの概要

1　サービス指向の集合知形成

インターネットは世界の人々を繋いだと言われるが，言語の壁は依然として存在している．インターネット上には多数の言語資源（データ及びソフトウェア）が存在しているが，専門家でなければ異文化コラボレーションの現場で利用することは難しい．複雑な契約や知的財産，データ構造やインタフェースの多様性が，言語資源の利用を困難にしている．

本研究は，言語資源をサービス化して共有する多言語基盤を実現することを目的とする．開発されたシステムは「言語グリッド（The Language Grid）」[Ishida 11] と呼ばれる．利用者は，言語グリッドにアクセスすることによって，大学や研究機関，企業が提供する言語サービスを利用し，さらにそれらのサービスを自由に組み合わせて用いることができる．また，利用者がその目的に合わせて，新たな言語サービスを作成し登録することも可能である．言語グリッド実現までには特に下記の二つの課題が挙げられた．

サービス指向の多言語基盤の構築: 言語サービスを蓄積し，共有するためには，標準のインタフェースを持つ原子サービスに基づいてサービスを連携する基盤ソフトウェアが必要である．さらに，利用者がそれらの言語サービスを用いて異文化活動のためのアプリケーションシステムを簡単に開発できなければならない．

ユーザ参加型デザインの実践：提供される言語サービスが多ければ多いほど，利用者はそのサービスによる利益を享受できる．つまり，サービス指向の集合知を形成するには，利用者とコミュニティを積極的に参加させることが必要である[1]．

クラウドコンピューティングなどのように，サービスを世界規模で集積し実行する計算環境が整いつつある．しかしながらサービス指向のアプローチの課題は，大規模な計算環境のみにあるのではない．スケールアップを可能とする計算環境を前提として，どのようにサービスを集積し，利用し，組み合わせて新たなサービスを生み出していくのかという制度設計も重要な課題である[Papazoglou 03]．ここで，Web サービスを要素として集合知を形成する枠組みを「サービスグリッド」と呼ぶ．

筆者らは実際にサービスグリッドのための基盤ソフトウェアを開発し，「言語グリッド（The Language Grid）」を運営してきた[Ishida 06]．本研究では，言語グリッドの運営経験から得られた多くの知見に基づき，大学や研究機関などの非営利組織を中心とする公共的なサービスグリッドの制度設計を試みた[2]．

2　言語資源から言語サービスへ

2.1　設計思想

言語グリッドは，集合知のアプローチを取っている．即ち，専門家や様々な利用現場のユーザが開発した言語資源を共有し利用できる環境として設計されている（図 1）．言語グ

[1] 集合知の成長は利用者の自発的な努力によるものとされている[Weiss 05]．

[2] サービスグリッドという用語は，従来から，サービス提供者の課す制約の範囲で，サービス利用者のコミュニティの要求を満たすようサービス合成が行われる枠組みの総称として用いられている[Furmento 02, Krauter 02]．

リッドの特徴は，言語資源をサービスの形で共有することである．そこには，サービスグリッド運用者，サービス提供者，サービス利用者の 3 種のステークホルダーが存在する．サービスグリッド運用者は，言語グリッドを管理し，言語サービスの実行を制御する．サービス提供者は，機械翻訳や形態素解析，辞書などの言語資源をサービスとして言語グリッドに登録する．サービス利用者は登録されたサービスを異文化コラボレーション活動に利用する．

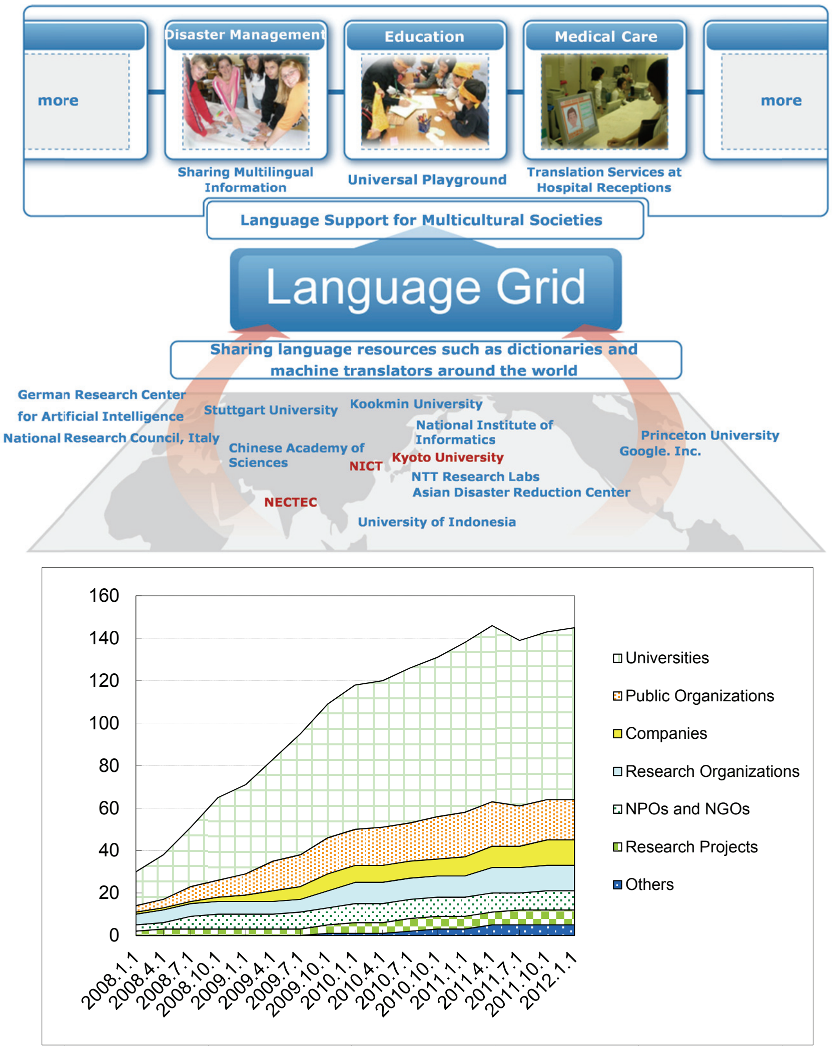


図 1　言語グリッドとその参加組織数の経緯

言語グリッドは，このように異なる組織から提供される言語サービスを結合するプラットフォームである．これまでも言語処理プログラムを結合しようとする試みとして DFKI の Heart of Gold[Callmeier 04] や IBM の UIMA[Ferrucci 04] が存在したが，主に研究開発者のためのプラットフォームで，共有データに対して，多様な言語処理プログラムをパイプライン的に適用することができる．UIMA 準拠の U-Compare [Kano 10] は統合自然言語処理

システムで，自動組み合わせ比較，統計評価，ワークフロー作成実行，結果の視覚化などの汎用基盤機能を有している．それに加え，様々な言語資源群をプログラミング作業なしで利用できるよう提供している．一方，言語グリッドは応用指向のプラットフォームで，サービス指向アーキテクチャに基づいて知財を管理することに焦点を当てている．このように目的が直交するため，DFKI の Heart of Goal と言語グリッドをシステム的に連結する共同研究を行った[Bramantoro 08]．今後，UIMA にもその成果を展開する予定である．

2.2　サービス階層

図 2 に示すように，言語グリッドは以下の 4 層から構成される[Murakami 08]．P2P サービスグリッドは，コアノードとサービスノードという 2 種類のノードを接続することを目的としている．コアノードはサービスの登録情報を管理し，サービスのアクセス制御を行い，サービスを連携させる．一方，サービスノードには，サービス実体とそのラッパーが配備される．

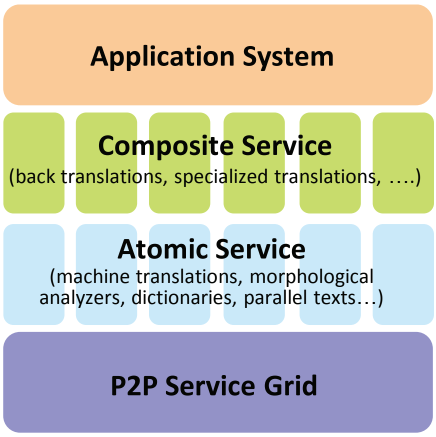


図 2　言語グリッドの階層

原子サービスは，個々の言語資源に対応した Web サービスである．例えば，機械翻訳や形態素解析，辞書，用例対訳が典型的な言語資源である．これらの資源は標準化されたサービスインタフェースに基づいてラッピングされる．既に，様々な言語データや言語処理ソフトウェアのサービスインタフェースを階層的に標準化するためのオントロジー体系が提案されている[Hayashi 08]．言語グリッド上で提供される言語サービスのインタフェースは，このオントロジー体系に基づいて規定されている．

複合サービスは，ワークフローによって原子サービスを合成したものである[Khalaf 03]．ワークフローは WS-BPEL によって記述され，BPEL 実行エンジンによって解釈，実行される[Andrews 03]．言語ドメインでは，折り返し翻訳や専門翻訳といった多様な複合サービスが必要となる．例えば，専門翻訳は，機械翻訳サービスや形態素解析サービス，および専門用語辞書サービスを合成して実現される．

言語グリッド Playground は京都大学の学生チームによって開発された応用システムで，

言語グリッド上の様々な言語サービスに，Web ブラウザを通じてアクセスすることができる（図 3）．Playground には，原子サービスの利用のための Basic サービス，原子サービスを組み合わせた複合サービスを利用するための Advanced サービス，異文化コラボレーション活動への応用に特化した Customized サービスがある．

図 3　言語グリッド Playground

3　サービスグリッドの制度設計

サービスグリッドのステークホルダー（利害関係者）について以下にまとめる．単純化のために，主要なステークホルダーは以下の 3 者とする．

(a) 「サービス提供者」はサービスグリッドに対して各種のサービスを提供する．

(b) 「サービス利用者」はサービスグリッドに提供されたサービスを呼び出して利用する．

(c) 「サービスグリッド運営者」はサービス提供者からサービスの提供を受け，そうしたサービスをサービス利用者に供する．

なお，サービス提供者とサービス利用者を「サービスグリッド利用者」と総称する．実際，サービスグリッド利用者は，サービス提供者とサービス利用者の両方の立場を取ることができる．サービスグリッド運営者の果たす役割は，サービスグリッド利用者の間（典型的にはサービス提供者とサービス利用者の間）に立って，サービスの提供と利用を促進することにある．以下では，サービスグリッド運営者とサービスグリッド利用者の契約という観点から制度設計を進める．

本研究扱うサービスは，「原子サービス」(atomic service)と「複合サービス」(composite service) に分かれる．原子サービスはサービスグリッド利用者からの資源へのアクセスを可能とする Web サービスをいう．ここで「資源」とは，サービスグリッドによって共有されるデータ，ソフトウェアや人的資源を言う．一方，複合サービスは，単数あるいは複数の原子サービスを呼び出す手続き（以下，「ワークフロー」）により実現される Web サービス

をいう．

ところでサービスや資源の知的財産権に関しては，運営者が統一的なライセンスを示し，それに合意した利用者がサービスを登録することが考えられる．しかし，統一的なライセンスはサービスグリッドの運営を単純化しその拡大を促進する一方で，サービス提供者にインセンティブを失わせる可能性がある．そこで以下では，多様なサービス提供者の立場を認め，運営者が統一的なライセンスを課すことを制度設計の前提とはしないこととする．

なお，本研究で議論するサービスグリッドの運営は，大学や研究機関などの非営利組織が中心となり，公共の場で行うことを想定している．企業内のサービスグリッドのように，サービス提供者とサービス利用者のインセンティブを完全にあるいは部分的に制御できる状況は前提としない．

3.1　サービスの提供

3.1.1　サービス利用目的の分類

サービス提供者の立場を考えると，自らの知的財産を守るために，サービス利用者の利用目的に関心を持つのは当然である．実際，研究機関や公的機関のホームページ上には，提供するサービスの利用を「非営利あるいは研究目的に限る」と明示していることも多い．そこで，こうしたサービス提供者の関心を反映するために，サービスの利用目的を以下の 3 種に分類し，その利用範囲を選択することを可能とする．

(a) 「非営利目的での利用」とは，(i) 公的機関や非営利組織の本来業務のための利用または, (ii) 公的機関や非営利組織以外の企業･団体の CSR（corporate social responsibility）活動のための利用をいう．

(b) 「研究目的での利用」とは，各種研究のための利用で，営利的収益に直接的に寄与しないものをいう．

(c) 「営利目的での利用」とは，非営利目的又は研究目的での利用以外の利用で，直接的又は間接的に営利的収益に寄与するものをいう．

公的機関や非営利組織の本来業務以外の業務を非営利目的での利用から除外するのは，活動資金確保のための活動でのサービス利用を認めないためである．一方，企業の CSR 活動を非営利目的での利用に含めるのは，こうした活動が公的機関や非営利組織の本来業務と連携して行われることが多いためである．

上記の分類は，組織による利用に限らず，個人による利用にも適用できる．しかし，個人利用が私的な利用のみを意味する場合には，個人利用を非営利目的での利用として扱うこともできる．

3.1.2　サービスの登録

サービス提供者は，自らのサービスをサービスグリッドに登録するとき，提供する資源の著作権及びその他の知的財産権の所在に関わる情報（第三者から使用許諾を受けている

のであればその旨を含む）を明示する必要がある．またサービス提供者は，登録した資源をサービス提供者が保有しているか，第三者に提供可能なものとして管理していることを保証する必要がある．これはサービス利用者が，誤ってサービス提供者や第三者の知的財産権を侵害することを防ぐためである．

では，サービスの登録や維持は誰によって行われるべきだろうか．集合知の形成がサービス提供者によって自律的に行われるという前提に立てば，提供する資源の維持，資源を原子サービスとするラッピング作業，提供するサービスの維持，提供するサービスとサービスグリッドとの接続の維持は，サービス提供者が行うものとせざるを得ない．一方，サービスの品質と安全性を重視する立場からは，サービスの登録や維持は，運営者によってあるいは運営者の承諾を得て行われるべきである．従って，サービスの登録や維持を誰が行うべきかについては，サービス提供者の自律的活動とサービスグリッドの品質や安全性とのトレードオフを検討して決める必要がある．

同様に，サービスの登録解除についても，サービス提供者に任せるのか，サービス利用者の利便性を重視し登録解除に制約を設けるのかを検討する必要がある．サービスグリッドの品質と安全性を重視する立場に立てば，少なくとも緊急時には運営者によってサービスの登録解除が行える必要がある．

3.1.3　サービス利用の制御

サービス提供者の立場からは，提供するサービスの利用条件を定める自由度があることが望ましい．例えば，以下のような利用条件が考えられる．

(a)　サービス利用者の制限

(b)　サービスの利用目的の制限

(c)　サービスを利用する応用システムの制限

(d)　サービスへのアクセス回数やダウンロードされるデータ量の制限

サービス利用者は，サービスグリッドに登録されたサービスを，サービス提供者が指定する利用条件の範囲内で利用できる．このため，サービスの利用時には，利用目的が非営利目的，研究目的又は営利目的のいずれであるかを指定する必要がある．例えば，サービス提供者がサービスを別途自治体などに販売している場合には，サービスグリッドを通じた非営利利用を認めたくないと考えるかもしれない．

一般にサービス利用条件のきめ細かな設定を可能とすることは，サービス提供者の満足度を増す一方で，そうした制限を順守することをサービス利用者に求めることを意味する．その結果，サービス利用者が利用条件に違反しないことを保証する技術的手段の提供が運営者に求められることになる．さらに複合サービスの利用に際しては，構成要素である全ての原子サービスの利用条件が満足されなければならない．これを自動的に保証しようとすると，サービスグリッドの機能は高度で複雑なものとなる．従って，サービス提供者の権利行使の自由度と，サービス利用者の利便性や運営者の負担との間のトレードオフを検

討する必要がある．

3.2　サービスの利用

3.2.1　応用システムを介したサービスの利用

サービスグリッドの利用が個人利用でない場合には，サービス利用者は何らかの応用システムを通じて，サービスをさらに広い範囲の利用者に提供することが多い．ここで「応用システム」とは，図 4 に示すように，サービス利用者が自ら運営するシステムで，サービスグリッドの ID やパスワードを知らなくても，当該応用システムの利用者が間接的にサービスグリッドを利用することができるものをいう．このような場合，サービス利用者は応用システム利用者に，応用システムの実現に用いられるサービスの利用条件を遵守させる責任が生じる．

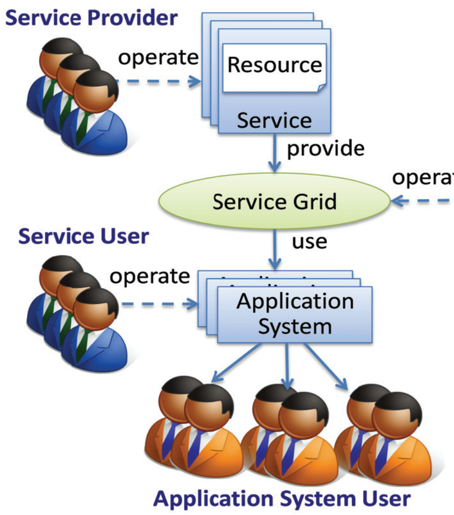


図 4　応用システムを介したサービスの利用

3.2.2　応用システムの運営方式

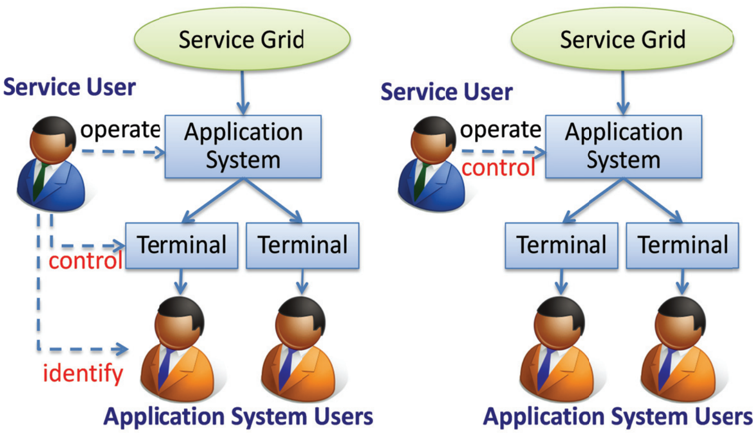
サービス利用者が運営する応用システムには様々なものが考えられる．Web を介して不特定多数の応用システム利用者にサービスを提供するものや，受け付け窓口などの特定の端末でサービスを提供するものなどがある．本研究では，応用システムがサービスの利用をどのように制御できるかに着目し，応用システムの運営を，クライアント制御下とサーバー制御下の運営に分類する．

「クライアント制御下」とは，応用システム利用者がサービス利用者の制御下にある場合をいう．即ち，応用システム利用者の端末機器がサービス利用者の制御下にある場合か，応用システム利用者をサービス利用者が特定できる場合をいう．いずれの場合も，サービス利用者が各端末機器の，あるいは各応用システム利用者の利用状況を常時把握でき，かつ必要に応じて端末機器あるいは応用システム利用者を特定して，その利用を随時停止で

きる権限を保持していることが求められる．

「サーバー制御下」とは，応用システム利用者がサービス利用者の制御下にはないが，応用システムを稼働させるためのサーバーがサービス利用者の制御下にある場合をいう．この場合には，サービス利用者が応用システムのサーバーの利用状況を把握でき，かつ必要に応じて応用システムのサーバーを随時停止する権限を保持していることが求められる．

応用システムの運営方式を図 5 に示す．例えば Web を介してサービスを提供する応用システムは，応用システム利用者が各々自宅から認証なしで利用できるとすれば，クライアント制御下で運営されているとは言えない．しかし，その Web サーバーをサービス利用者が管理していれば，サーバー制御下で運営されていると言える．一方，受け付け窓口の端末でサービスを提供する応用システムは，その端末がサービス利用者によって管理されていれば，クライアント制御下での運営に分類される．



(1) Client Control　　(2) Server Control

図 5　応用システムの運営方式

3.2.3　サービス提供者へのリターン

サービス提供者がサービスを提供するインセンティブはどこから来るのだろう．サービス提供者が有償でサービスを提供する場合には，サービス利用者と別途契約して，有償でサービスを利用させることができる．このとき，運営者は契約内容に関与する必要はない．

サービス提供者が無償でサービスを提供する場合には，サービスグリッド運営者に求められるものは，サービス提供者にサービスの利用統計情報を提供することである．この利用統計情報は，どのサービス利用者がどのようなサービスをどの程度利用しているかを示すものである．こうした情報は，サービス提供者とサービス利用者とのインタラクションを刺激する．但し，利用統計情報には，通信内容や通信当事者に関する個人情報は含むべきではない．サービス提供者が利用統計情報以外の情報の取得を望む場合には，別途サービス利用者と情報の提供について契約を締結する．このとき，サービスグリッド運営者はこうした契約に関与する必要はない．

こうした分類を行うのは，サービス利用者による応用システムの開発を許容するとともに，サービス提供者がサービスの提供範囲を適切に選択できるようにするためである．例えば，別途自治体にサービスを販売しているサービス提供者は，病院窓口での患者へのサービスの提供（クライアント制御下の運営）に異存がない場合でも，自治体の Web 上で市民へサービスを提供すること（サーバー制御下の運営）には難色を示すことがある．このような場合，サービス提供者は応用システムの運営方式をサービス利用条件に指定することによって，サービスの提供範囲を制限する．一方，サービス利用者は，それぞれの運営方式で利用が許可されたサービスのみを用いて応用システムを構築する．

4　基盤ソフトウェアとツール

4.1　システムアーキテクチャ

図 6 に P2P サービスグリッドのシステム構成を示す．サービス提供者は，Web サービスのインタフェース記述である WSDL ファイルとサービスの著作権情報，ライセンス情報，アクセス制約をサービスマネージャ (Service Manager) に登録する．サービスマネージャは，WSDL ファイルを取得すると，インタフェース情報とエンドポイントの URL を抽出し，同じインタフェースの仮想エンドポイントをサービススーパバイザ（Service Supervisor）上に生成する．仮想エンドポイントの目的は，サービスへの直接のアクセスを禁止し，指定されたアクセス制約に基づいて，サービスへのアクセスを制御することである．

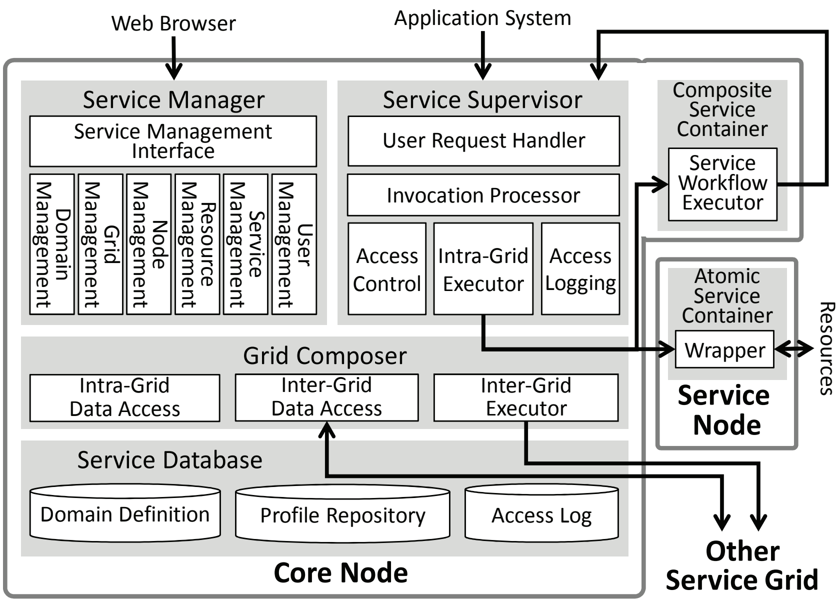


図 6　P2P サービスグリッドのシステム構成

サービスを利用するときには，応用システムから仮想エンドポイントに SOAP リクエストを送りサービスを呼び出す．サービススーパバイザは，そのリクエストをユーザリクエ

ストハンドラで受け取ると，サービス登録時に設定されたアクセス制約を満たしているかどうか検証する．満たしていれば，サービススーパバイザは実際のエンドポイントをプロファイルレポジトリから取得しサービスにアクセスする．サービスからのレスポンスはアクセスログに蓄積され，アクセス制約が守られていることの検証や，サービス利用のモニタリングに利用される．

4.2　サービススーパビジョン

複合サービス内に定義されたサブタスクを抽象サービスと呼び，その抽象サービスを実際に実行する Web サービスを具象サービスと呼ぶ．サービス合成問題は，いずれに注目するかによって，以下の 2 種類に分けられる．

(a) 垂直型合成: 最善の抽象サービスの組み合わせを求める
(b) 水平型合成: 機能的に等価な Web サービスの集合から，最善の具象サービスの組み合わせを求める

我々は水平型サービス合成に取り組み，初めて制約最適化問題として定式化した[Hassine 06]．言語サービスでは具象サービスの組み合わせの数が大きくなることに注目し，利用制約を満たし，かつ QoS を最大化する具象サービスの組み合わせを求める手法を示した．

また，多様な組織から異なるポリシーの元で提供されるサービスを連携させるため，サービス実行時の振る舞いを制御するサービススーパビジョンと呼ぶ機構を開発した[M.Tanaka 09]．サービススーパビジョンは，例えば，文脈に基づくピボット翻訳に利用できる．ピボット翻訳は，軸になる言語を介した 2 つの機械翻訳機の連携によって実現される．ピボット翻訳では，2 つの機械翻訳機の訳語選択が一貫しないことから，意味のドリフト[3]が起こることがある．訳語選択[R.Tanaka 09, Matsuno 11]の文脈を，サービススーパビジョンを用いて引き継ぐことによって，この問題を解決できる．

4.3　言語グリッド ToolBox

国際的な NPO は海外に拠点を持ち，各拠点でボランティアスタッフが活動しているが，相互に連携して拠点間のアクティビティを計画することは，母語が異なるために容易ではない[Mori 07]．例えば，NPO パンゲア[4]は，世界の子供たちのつながりを作ることを目的として活動している．日本，韓国，オーストリア，ケニア，マレーシア，ベトナムに拠点を持ち，ICT を利用して非同期・同期アクティビティを行い，子どもたちの相互理解を育てようとしている．各拠点のボランティアスタッフのコミュニケーション手段として，多言語のコミュニティサイト（図 7）を開発し，活動報告を多言語掲示板により共有している．この多言語コミュニティサイトは，言語グリッドを用いて実装されている．ボランティアスタッフは，母語で報告を書き込み，他拠点の書き込みを母語で閲覧できる．NPO が活動内で利用する外来語や造語，固有名詞などを独自の辞書に登録し，機械翻訳と連携し利用す

[3] 機械翻訳機から機械翻訳機へと訳文が引き継がれ，伝言ゲームのように意味が変化していく．
[4] http://www.pangaea.org/

ることで翻訳品質を向上させている．さらに，コミュニティ内で，翻訳結果を修正し合うことにより，自然な翻訳文を共有することができるようになっている．

NPO において，多言語コミュニティサイトが日常的に利用されていることは，言語グリッドの研究開発に大きなフィードバックを与えた．実際に，このコミュニティサイトを参考に，多言語コミュニケーションを支援するツール群である言語グリッド Toolbox が開発され，現在，多くのグループが利用している（図 8）．

図 7　多言語コミュニティサイト（日本語画面）

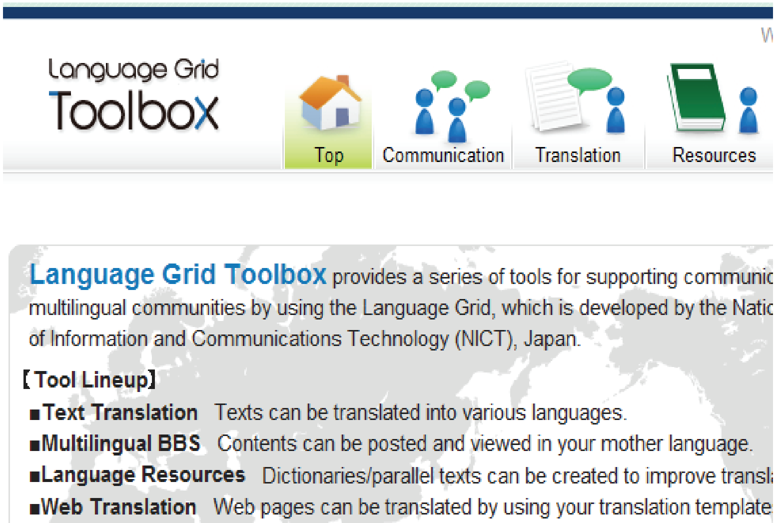


図 8　言語グリッド Toolbox

言語グリッド Toolbox は，コミュニティにおける異文化コラボレーションを支援するモジ

ュール群であり，多言語 BBS，辞書作成などの機能を持つ．また，オープンソースソフトウェアとして提供されており，各コミュニティが必要に応じて拡張できる．

現在，NPO パンゲアは，自ら開発したツールのメンテナンスを中止し，言語グリッド Toolbox を利用して多言語コミュニティサイトを再構築している．このような利用者と開発側のアイデアの循環を通じて，異文化コラボレーションツールの参加型デザインが実践されている．

5　言語グリッドの利用

5.1　ローカルコミュニティでの利用

在日外国人の増加に伴い，医療の現場においても，十分に日本語を話すことができない外国人患者との対話が大きな問題となっている．医療現場の場合，病状，薬，保険制度などが，医療従事者と患者の双方で正しく伝わらなければならない．京都では，医療通訳ボランティアが同行する支援が行なわれているが，その需要は増大している．

そこで，用例対訳を利用し，医療従事者と患者間の対面でのコミュニケーションを支援する多言語医療受付支援システム M^3（図 9）が，和歌山大学と多文化共生センターきょうとにより開発された[宮部 09]．医療現場，特に医療受付時に高頻度で利用される用例が必要となるため，医療用例収集システム TackPad が開発され，医療通訳ボランティアによる用例対訳の収集が行われている．

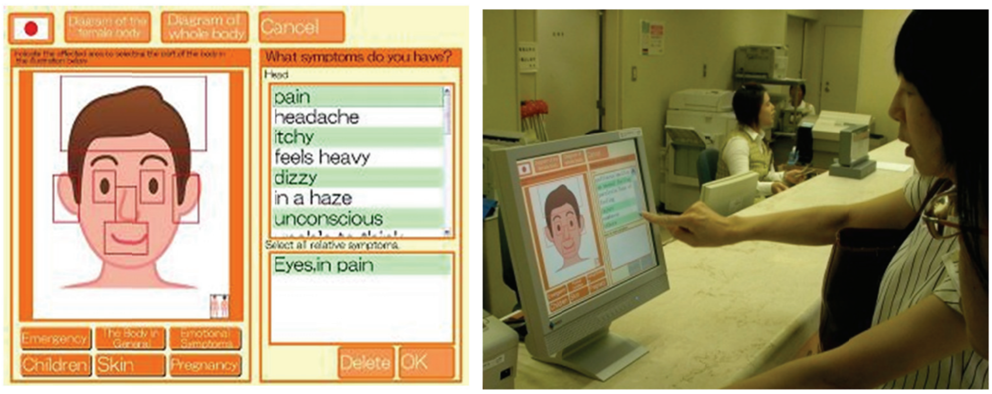


図 9　多言語医療受付支援システム M^3

現在，M^3 は，京都市立病院，京都大学医学部附属病院，洛和会音羽病院，東京大学医学部附属病院に導入され，多言語受付の支援が行われている．また，病院に行く前の医療支援を目的とした Web 版 M^3 やモバイル版 M^3 の公開も行われている．

5.2　グローバルコミュニティにおける利用

Wikipedia は，誰でも記事を作成・編集できるため，約 270 もの言語により情報が共有さ

れている．これらの記事はそれぞれの文化を背景に執筆されているため，異文化の相互理解のための知識の宝庫と言える．

しかしながら，その内訳を調べると，英語では 354 万本の記事があるのに対し，日本語では 73 万本，タイ語では 6 万本など言語によって記事の数に大きな偏りがある．知識の翻訳を加速するためには，翻訳に関する議論が可能な多言語掲示板が必要である．

そこで Wikimedia 財団と共同で，言語グリッドを応用した多言語掲示板を MediaWiki 上に開発した[5]．この多言語掲示板を用いれば，世界中の Wikipedia ボランティアは，記事の翻訳のために，多言語での質問応答を行うことができる．

実現方法としては，まず，MediaWiki 上に，言語グリッドへのアクセス手段を提供する言語グリッドエクステンション（図 10）を開発した．次に，これを利用し，Wikimedia 財団が開発した単言語の掲示板『Liquid Thread』を拡張した多言語掲示板『Multilingual Liquid Thread』を開発した．Multilingual Liquid Thread は，記事ごとに多言語用語集を作成できるため，記事ごとに機械翻訳をカスタマイズし，翻訳精度を向上させることができる．今後，Wikimedia 財団のサーバーにセットアップされ，テストを開始する予定である．

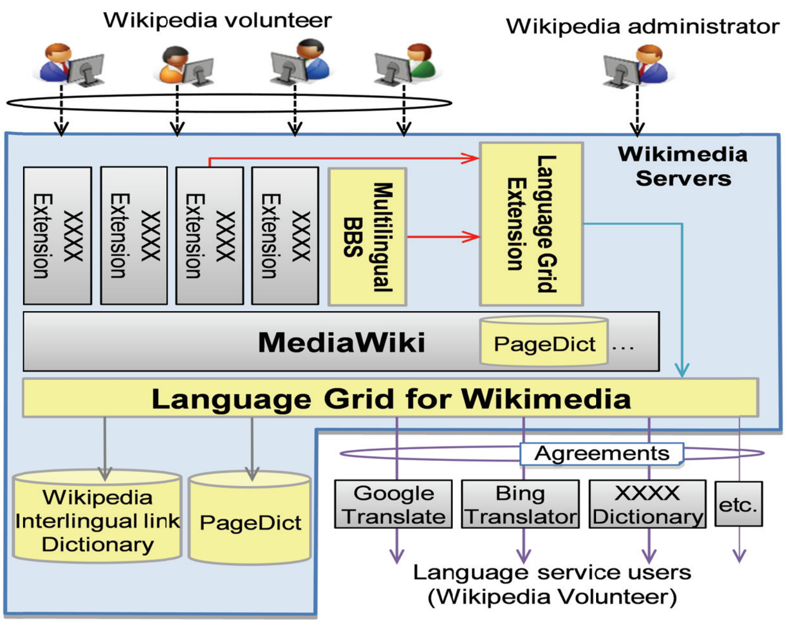


図 10　言語グリッドエクステンション

6　運営

6.1　言語グリッドの運営

筆者らが考案した言語グリッドの運営モデルは，世界各地の研究機関や NPO などの利用グループの意向を反映したものである[Ishida 08]．運営モデルの策定は言語グリッドの基盤ソフトウェアの開発と並行して行われたが，運営モデルの合意には半年以上を要した．運

[5] MediaWiki は Wikipedia など，Wikimedia 財団が提供するサービスのプラットフォームである．

営モデルを実現するために，基盤ソフトウェアが開発されたと言っても過言ではない． 言語グリッドは，2007 年 12 月に京都大学によって運営が開始された．その後，17 カ国 145 組織が覚書に署名している[6]．参加組織は，例えば，中国科学院や CNR，DFKI，NII といった研究機関や，シュツットガルト大学，プリンストン大学，清華大学，そして多くの日本の大学，NPO/NGO や公的機関などである．NTT や東芝，沖電気，Google といった企業も参加し無償で機械翻訳サービスなどを提供している．

2011 年 2 月には，タイの NECTEC が言語グリッドオペレーションセンターをバンコクに立ちあげ，京都大学のオペレーションセンターと連邦制運営を開始した[石田 10]．その結果，言語グリッド（京都，バンコク）に登録された言語サービスは，現在，130 を超えた．多様な原子・複合サービスが，Translation，Bilingual Dictionary，Parallel Text，Morphological Analysis，Text-to-Speech など 20 種のサービスタイプに分類され共有されている．

ところで，「言語資源から言語サービスへ」という言語グリッドの方向性が，欧州，米国の言語資源研究者の間で共有され始めている．米国では，自然言語処理，情報検索，機械翻訳，音声，セマンティックウェブなどの分野で，これまで個別に作成されてきた言語資源を，分野を超えて再利用するプロジェクト SILT (Sustainability Interoperability for Language Technology)が進められてきた[Ide 09]．SILT の次期プロジェクトは，言語グリッドの基盤ソフトウェアを利用する計画になっている．

また，欧州では，効率的に新規の言語技術や言語資源を開発できるように，今後の技術課題の優先度付けやロードマップを検討するプロジェクト FLaReNet （Fostering Language Resources Network）が進められてきた[Calzolari 10]．この FLaReNet は言語グリッドを参考に，言語資源から言語サービスへの移行を提唱し，MetaNet という新しいプロジェクトを生みだしている．言語サービスを世界規模で共有するために，欧米とアジアの協力が今後ますます必要となると思われる．

6.2　サービスグリッドの運営

大学や研究機関などの非営利組織を中心とするサービスグリッドが世界的な広がりを見せるためには，複数の運営者の連携が求められる．これを「連邦制の運営」と呼ぶ．連邦制の運営が必要となる理由は，運営者が管理できるサービスグリッド利用者の数に限りがあるからだけではない．運営者がコミュニケーションを行えるサービスグリッド利用者の範囲に地理的あるいは専門的な観点からの局所性があるからである．

連邦制の運営には２つの方式が考えられる．第一は集権的な方式で，運営者を構成員とする連邦組織を別途構成し，合意に基づいてサービスグリッド間の連携の仕組みを決定していく．この方式は，合意により連携の在り方を柔軟に決定できるが，連邦組織の維持には多大な労力を要する．第二は分権的な方式で，サービスグリッド利用者が，同一の覚書

[6]図 1 に示したように，参加組織の数は順調に伸びている．連邦制の開始に伴い既存ユーザと覚書の再締結を進めた結果，2011 年 4 月に，一時的に参加組織数が減少している．

を用いて別のサービスグリッドの運営者となることを許す．この方式は，運営者が P2P 型のネットワークを構成することを促すものである．連携の仕組みは予め共通に用いる覚書により定められているが，連邦組織のネットワーク形成は柔軟で，その維持も容易である．以下では，大学や研究機関などの非営利組織に向くと思われる，分権的な連邦制の運営方式を詳しく述べる．

「連携運営者」とは，同一の覚書を用いて別途自らサービスグリッドを運営しているサービスグリッド利用者をいう．また「連携利用者」とは，同一の覚書を用いて連携運営者が運営するサービスグリッドの利用許諾を受けているものをいう．このとき連携利用者が，図 11 に示すように，連携運営者がサービスグリッド利用者として参加しているサービスグリッドを利用できるというのが連邦制のアイデアである．但し，その場合にも，サービス提供者が連携利用者に利用許諾をするか否かの選択をする権限は継承される．

一般に 2 つのサービスグリッドが対等の関係で連携するには，双方の運営者が各々相手方のサービスグリッド利用者となり覚書を締結すればよい．こうした双方向の連携は，同種のサービスグリッドが地理的な制約を超えてネットワークを形成していくのに適している．

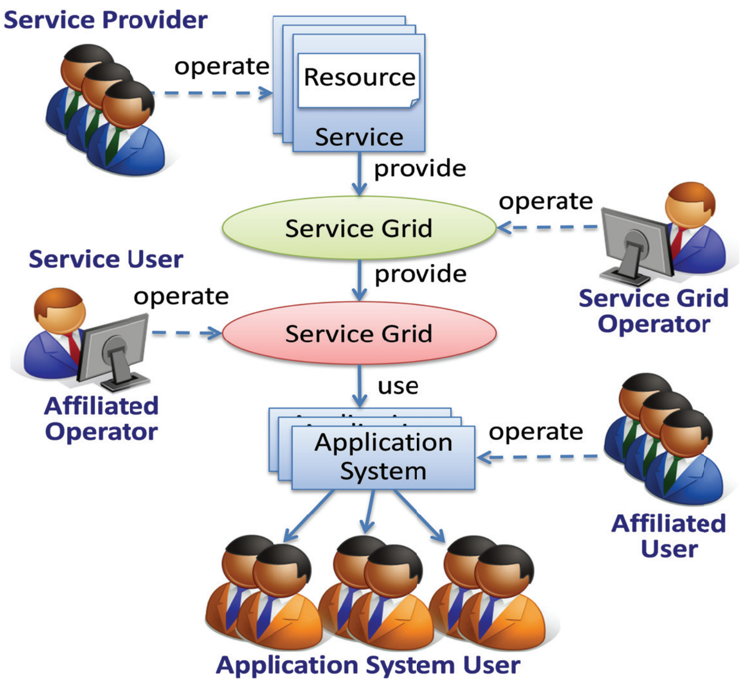


図 11　連邦制によるサービスグリッドの運用

しかしながら，一方向の連携が意味を持つこともある．例えば，一方が基盤的なサービスを提供するグリッドで，他方が応用的なサービスを提供するグリッドの場合には，後者が前者のサービスグリッドの利用者となればよい．このような一方向の連携は，異種のサービスグリッド間で機能的な補完をする場合に適している．

異なるサービスグリッドが同一の覚書を用いることが困難な場合もある．特に問題となるのは準拠法である．国際的な連携では，ニューヨーク州法など特定の法令を準拠法と定めることも考えられるが，運営者はそれぞれが所在する地の法令を準拠法とすることを望むかもしれない．そのような場合には，運営者ごとに準拠法を除いて同一の覚書を作成することになる．このような場合には，サービス提供者は，連携利用者が異なる準拠法の下でサービスを利用することを理解しておく必要がある．

7　むすび

言語グリッドは，利用者の目的に合わせた多言語環境を構築するためのサービス指向の多言語基盤である．各大学や研究機関，企業等が提供している言語サービスを利用者が自由に組み合わせることを可能にする．各地域の学校の多言語支援，商店街のコミュニティの支援等の活動に利用されている[Ishida 07, Fussell 09]．例えば，世界中の子ども達が描いた災害安全マップをインターネット上で共有し，防災協働学習を支援するシステム CoSMOS（Collaborative Safety Maps on Open System）などが開発されている[Ikeda 10]．言語グリッドを活用して多言語チャットシステムも実装されている[Nakatsuka 10]．このチャットシステムには，機械翻訳サービスで活用できる領域固有の対訳用例を収集する機能が組み込まれている．

言語グリッドを用いた新しい研究も生まれている．例えば，機械翻訳を介したコミュニケーションというインタラクションスタイルの分析が行われている[Yamashita 06, Yamashita 09]．また，研究者とフィールドワーカーとのコラボレーションは，創作絵文字とその解釈の文化差に関する研究を生み出している [Takasaki 07, Cho 08]．ユビキタス分野では，スマートクラスルームの機能をサービスとして再構築し，言語サービスと結合した多言語のオープンスマートクラスルームが開発された[Suo 09]．人文，社会科学系の論点からも，言語グリッドを利用した多文化共生支援の可能性と問題点が論じられている[喜多 08]．特に，翻訳リペアの営みを共生日本語の実践と比較し，その類似点と相違点が論じられている．また，工学的アプローチのフィールド情報学と人文学系のいうアクションリサーチとの比較が行われている．

本研究は 2001 年の 9.11 を契機として京都大学で始めた異文化コラボレーション実験が出発点となっている．それから 10 年が過ぎたが，インターネット上に公共のサービス指向の多言語基盤が必要だという認識は変わっていない．それどころか，今後，益々その必要性は高まり，欧米アジアの協力が必要になると感じている．

本研究では，Web サービスを要素として集合知を形成する枠組みをサービスグリッドと呼び，大学や研究機関などの非営利組織を中心とする公共的なサービスグリッドの制度設計を試みた．本論文の提案は，筆者らの 2 年間に及ぶサービスグリッドの運営経験に基づいている．こうした経験の共有が，制度設計の知見の蓄積を促し，サービス指向の集合知の発展に寄与することを願っている．

参考文献

[Andrews 03] T. Andrews, F. Curbera, H. Dolakia, J. Goland, J. Klein, F. Leymann, K. Liu, D. Roller, D. Smith, S. Thatte, I. Trickovic, and S. Weeravarana, “Business process execution language for Web services,” Specification, 2003.

[Bramantoro 08] A. Bramantoro, T. Tanaka, Y. Murakami, U. Schäfer, and T. Ishida, “A hybrid integrated architecture for language service composition,” *IEEE International Conference on Web Services (ICWS-08)*, pp. 345-352, 2008.

[Callmeier 04] U. Callmeier, A. Eisele, U. Schäfer, and M. Siegel, “The deep thought core architecture framework,” *LREC 2004*, pp.1205-1208, 2004.

[Calzolari 10] N. Calzolari, and C. Soria, “Planning the future of language resources: the role of the FLaReNet,” *International Conference on Network Computational Linguistics and Intelligent Text Processing (CICLing-10)*, LNCS 6008, pp.1-11, 2010.

[Cho 08] H. Cho, T. Ishida, T. Takasaki, and S. Oyama, “Assisting pictogram selection with semantic interpretation,” *European Semantic Web Conference (ESWC-08)*, LNCS 5021, pp. 65–79, 2008.

[Ferrucci 04] D. Ferrucci, and A. Lally, “UIMA: an architectural approach to unstructured information processing in the corporate research environment,” *Natural Language Engineering*, Vol. 10, pp. 327-348, 2004.

[Furmento 02] N. Furmento, W. Lee, A. Mayer, S. Newhouse, and J. Darlington, “ICENI: an open grid service architecture implemented with Jini,” *International Conference on High Performance Networking and Computing*, pp.1-10, 2002.

[Fussell 09] S. Fussell, P. Hinds, and T. Ishida (Eds), *The Second International Workshop on Intercultural Collaboration*, ACM Press, 2009.

[Hassine 06] A. Ben Hassine, S. Matsubara, and T. Ishida, “Constraint-based approach for Web service composition,” International Semantic Web Conference (ISWC-06), LNCS 4273, pp. 130-143, 2006.

[Hayashi 08] Y. Hayashi, T. Declerck, P. Buitelaar, and M. Monachini, “Ontologies for a global language infrastructure,” *Proc. of ICGL2008*, pp.105-112, 2008.

[Ide 09] N. Ide, J. Pustejovsky, N. Calzolari, and C. Soria, “The SILT and FlaReNet international collaboration for interoperability,” *Third Linguistic Annotation Workshop*, pp.178-181, 2009.

[Ikeda 10] Y. Ikeda, Y. Yoshioka, and Y. Kitamura, “Intercultural collaboration support system using disaster safety map and machine translation,” *Culture and Computing*, Lecture Notes in Computer Science 6259, Springer, 100-112, 2010.

[Ishida 06] T. Ishida, “Language Grid: An Infrastructure for Intercultural Collaboration,” *IEEE/IPSJ Symposium on Applications and the Internet (SAINT-06)*, pp. 96-100, keynote address, 2006.

[Ishida 07] T. Ishida, S. Fussell, and P. Vossen (Eds.), *The First International Workshop on Intercultural Collaboration, Lecture Notes in Computer Science*, vol.568, Springer-Verlag, 2007.

[Ishida 08] T. Ishida, A. Nadamoto, Y. Murakami, R. Inaba, T. Shigenobu, S. Matsubara, H. Hattori, Y. Kubota, T. Nakaguchi, and E. Tsunokawa,"A Non-Profit Operation Model for the Language Grid," *International Conference on Global Interoperability for Language Resources*, pp. 114-121, 2008.

[Ishida 11] T. Ishida (Ed.), *The Language Grid: Service-Oriented Collective Intelligence for Language Resource Interoperability*, Springer, 2011.

[Kano 10] Y. Kano, M. Miwa, K. Cohen, L. Hunter, S. Ananiadou, and J. Tsujii, "U-Compare: a modular NLP workflow construction and evaluation system," *IBM Journal of Research and Development*, Vol. 55, No. 3, pp. 11:1-11:10, 2010.

[Khalaf 03] R. Khalaf, N. Mukhi, and S. Weerawarana, "Service-oriented composition in BPEL4WS," *World Wide Web Conference,* 2003.

[Krauter 02] K.Krauter, R. Buyya and M.Maheswaran, "A taxonomy and survey of grid resource management systems for distributed computing," *Software-Practice & Experience*, Vol.32, No.2, pp.135-64, 2002.

[Matsuno 11] J. Matsuno, and T. Ishida, "Constraint optimization approach to context based word selection," *International Joint Conference on Artificial Intelligence (IJCAI-11)*, 2011.

[Mori 07] Y. Mori, "Atoms of bonding: communication components bridging children worldwide," *Intercultural Collaboration*, LNCS 4568, pp. 335-343 (2007).

[Murakami 08] Y. Murakami, and T. Ishida, "A layered language service architecture for intercultural collaboration," *International Conference on Creating, Connecting and Collaborating through Computing (C5-08)*, 2008.

[Nakatsuka 10] M. Nakatsuka, S. Yasunaga, and K. Kuwabara, "Extending a multilingual chat application: towards collaborative language resource building," *9th IEEE Int. Conf. on Cognitive Informatics (ICCI '10)*, pp. 137-142, 2010.

[Papazoglou 03] M.P. Papazoglou, "Service-Oriented Computing: Concepts, Characteristics and Directions," *International Conference on Web Information Systems Engineering*, p.3, 2003

[Suo 09] Y. Suo, N. Miyata, H. Morikawa, T. Ishida, and Y. Shi, "Open smart classroom: extensible and scalable learning system in smart space using Web service technology," *IEEE Transactions on Knowledge and Data Engineering*, Vol.21, No.6, pp. 814-828 , 2009.

[Takasaki 07] T. Takasaki, and Y. Mori, "Design and development of a pictogram communication system for children around the world," *Intercultural Collaboration*, LNCS 4568, pp. 193-206, 2007.

[M.Tanaka 09] M. Tanaka, T. Ishida, Y. Murakami, and S. Morimoto, "Service supervision: coordinating Web services in open environment," *IEEE International Conference on Web Services (ICWS-09)*, pp. 238-245, 2009.

[R.Tanaka 09] R. Tanaka, Y. Murakami, and T. Ishida, "Context-based approach for pivot translation services," *International Joint Conference on Artificial Intelligence (IJCAI-09)*, pp.1555-1561, 2009.

[Weiss 05] A. Weiss, "The power of collective intelligence," *Networker*, Vol. 9, No.3, pp. 16-23, 2005.

[Yamashita 06] N. Yamashita, and T. Ishida, "Effects of machine translation on collaborative work," *International Conference on Computer Supported Cooperative Work (CSCW-06)*, pp. 515-523, 2006.

[Yamashita 09] N. Yamashita, R. Inaba, H. Kuzuoka, and T. Ishida, "Difficulties in establishing common ground in multiparty group using machine translation," *ACM Conference on Human Factors in Computing Systems (CHI-09)*, pp.679-688, 2009.

[喜多 08] 喜多千草, "情報通信基盤による多言語環境支援の可能性について —『言語グリッド』構築の実践とその思想," 多言語多文化—実践と研究, no.1, pp.77-100, 2008.

[宮部 09] 宮部真衣, 吉野 孝, 重野亜久里, "外国人患者のための用例対訳を用いた多言語医療受付支援システムの構築," 電子情報通信学会論文誌 D, vol.J92-D, no.6, pp.708-718, 2009.

[石田 10] 石田 亨, 村上 陽平, "サービス指向集合知のための制度設計," 電子情報通信学会論文誌 D, vol.J93-D, no.6, pp.675-682, 招待論文, 2010.

第 2 部　主要論文

本基盤研究では，言語グリッドプロジェクトの開発，運営，利用の水先案内として，以下に示す研究が先駆的に行われ，その一部は言語グリッドの改良にも反映されている．

[サービスグリッドアーキテクチャ]

多言語環境を, 言語サービスを連携させて構成するアイデアは 2006 年に発表しているが, 解説を 2010 年に IEEE Internet Computing で発表している．同様の試みの先駆的なものとしては，DFKI の Heart of Gold がある．そこで，言語グリッドと Heart of Gold の相違点を検証し，接続を可能とする研究を DFKI と共同で行い，LREC 2010 で発表している．また，実際に多言語環境を構築するビルディングブロックを構成し，ICIC（異文化コラボレーション国際会議）の前身であるワークショップに発表している．

1. Toru Ishida. Intercultural Collaboration Using Machine Translation. *IEEE Internet Computing*, pp. 26-28, 2010.
2. Arif Bramantoro, Ulrich Schäfer and Toru Ishida. Towards an Integrated Architecture for Composite Language Services and Components. *International Conference on Language Resources and Evaluation (LREC-10)*, March 21st, 2010.
3. Satoshi Sakai, Masaki Gotou, Satoshi Morimoto, Daisuke Morita, Masahiro Tanaka, Toru Ishida and Yohei Murakami. Language Grid Playground: Light Weight Building Blocks for Intercultural Collaboration. *International Workshop on Intercultural Collaboration (IWIC-09), Poster Session,* ACM, pp. 297-300, Palo Alto, California, USA, February 21st, 2009.

[機械翻訳連携]

複数の翻訳機をカスケード状につなぐものである．機械翻訳が，主に英語と他の言語との間で開発されているために，アジア言語と欧州言語の翻訳を実現するには，機械翻訳連携が必要となる．この時の問題点は，インタラクション分析の手法を用いて解明され CHI 2009 で報告されている．その結果を用いた問題点の解決は，IJCAI 2009, IJCAI 2011 で報告されている．

4. Naomi Yamashita, Rieko Inaba, Hideaki Kuzuoka and Toru Ishida. Difficulties in Establishing Common Ground in Multiparty Groups using Machine Translation. *In Proceedings of International Conference on Human Factors in Computing Systems (CHI-09)*, ACM, pp. 679-688, Boston, USA, April 6th, 2009.
5. Rie Tanaka, Yohei Murakami and Toru Ishida. Context-Based Approach for Pivot Translation Services. *In Proceedings of International Joint Conference on Artificial Intelligence (IJCAI-09),* AAAI Press, pp.1555-1561, Pasadena, California, USA, July 16th, 2009.

6. Jun Matsuno and Toru Ishida. Constraint Optimization Approach to Context Based Word Selection. *International Joint Conference on Artificial Intelligence (IJCAI-11)*, pp. 1846-1851, Bercelona, Spain, July 20th 2011.

[ユーザ中心 QoS]

ユーザによって評価の変わるサービス品質の問題を捉えようとする試みである．英語の不得意なユーザにとっては，英語のサービスより母語のサービスの方が価値は高い．しかし，一方で，英語しか話せない外国人が一人でも会話に参加すると，会話の言語が英語に切り替わるのは，研究室においても日常的に経験することである．この問題は SKG 2000 に招待論文として発表している．また，実行時でのサービス切り替えを可能とする Service Supervision と名付けた仕組みは，SCC 2010 で発表している．

7. Arif Bramantoro and Toru Ishida. User-Centered QoS in Combining Web Services for Interactive Domain. *In Proceedings of International Conference on Semantics, Knowledge and Grid (SKG-09)*, IEEE, pp.41-48, Zhuhai, China, October 12th -14th, 2009.
8. Yohei Murakami, Naoki Miyata and Toru Ishida. Market-Based QoS Control for Voluntary Services. *IEEE International Conference on Services Computing (SCC-10)*, pp. 370-377, July 7th, 2010.
9. Masahiro Tanaka, Yohei Murakami, Donghui Lin and Toru Ishida. Service Supervision for Service-oriented Collective Intelligence. *IEEE International Conference on Services Computing (SCC-10)*, pp.154-161, July 7th, 2010.
10. Shinsuke Goto, Yohei Murakami and Toru Ishida. Reputation-Based Selection of Language Services. *IEEE International Conference on Services Computing (SCC-11)*, pp.330-337, Washington DC, USA, July 6th 2011.

[共同翻訳]

異言語のユーザの協力による翻訳を研究対象としている．言葉が通じないために機械翻訳を活用するのだが，翻訳精度が悪いため，適切なプロトコルを用いなければ最終的によい翻訳は得られない．基本的なアイデアは IUI 2009 で発表している．また，このアイデアは，多くのボランティアにより進められている Wikipedia 翻訳にも適用可能である．実際に行われている Wikipedia 翻訳の観察結果は，Culture and Computing 2011 で発表している．また，別途，Wikimedia 財団と協力したプロトタイプ開発が行っているが，本研究成果はその検討にも生かされている．

11. Daisuke Morita and Toru Ishida. Collaborative Translation by Monolinguals with Machine Translators. *In Proceedings of International Conference on Intelligent User Interfaces (IUI-09), Poster Session*, ACM, pp. 361-366, Sanibel Island, Florida, USA, February 8th-11th, 2009.
12. Linsi Xia, Naomi Yamashita and Toru Ishida. Analysis on Multilingual Discussion for

Wikipedia Translation. *International Conference on Culture and Computing (Culture and Computing-11)*, Kyoto, Japan, October 20-22nd 2011.

Intercultural Collaboration Using Machine Translation

Toru Ishida
Kyoto University

Almost every country on Earth is engaged in some form of economic globalization, which has led to an increased need to work simultaneously in multiple cultures and a related rise in multilingual collaboration. In local communities, we can already see this trend emerging in the rising number of foreign students attending schools. Regional communities have had to solve the communication problems among teaching staffs, foreign students, and their parents, typically by focusing on relieving culture shock and its related stress with the aid of bilingual assistants. When turning our eyes to global communities, problems such as the environment, energy, population, and food require something more — mutual understanding. In both local and global cases, the ability to share information is the basis of consensus, thus language can be a barrier to intercultural collaboration.

Because there's no simple way to solve this problem, we must combine several different approaches. Teaching English to both foreign and local students is one solution in schools, but learning other languages and respecting other cultures are almost equally important. Because nobody can master all the world's languages, machine translation is a practical interim solution. Although we can't expect perfect translations, such systems can be useful when customized to suit the communities involved. To customize machine translations, however, we need to combine domain-specific and community-specific dictionaries, parallel texts with machine translators. Furthermore, to analyze input sentences to be translated, we need morphological analyzers; training machine translators with parallel texts requires dependency parsers. In the future, users might also want to use speech recognition/synthesis and gesture recognition. Even for supporting local schools, which include students from different countries, we need worldwide collaboration to generate all the necessary language services (data and software). Fortunately, Web service technologies enable us to create a workflow that assists in their creation. At Kyoto University and NICT, we've been working on the Language Grid,[1] which is an example of a service-oriented language infrastructure on the Internet.

Customized Language Environment Everywhere

Let's look at what could happen in the very near future in a typical Japanese school, where the number of Brazilian, Chinese, and Korean students is rapidly increasing. Suppose the teacher says "you have cleanup duty today (あなたは今日掃除当番です)" in Japanese, meaning "it is your turn to clean the classroom today." Now imagine that some of the foreign students don't understand what she said — to figure it out, they might go to a language-barrier-free room, sit in front of a computer connected to the Internet, and watch the instructor there type the following words in Japanese on the screen: "you have cleanup duty today." The resulting translation appears as "今天是你负责打扫卫生" in Chinese, "오늘은 네가 청소 당번이야" in Korean, and "Hoje é seu plantão de limpeza" in Portuguese. "Aha!" say the kids with excited faces. One of them types in Korean, "I got it," and the translation appears in Japanese on the screen.

Is machine translation that simple to use? Several portal sites already offer some basic services, so let's challenge them with my example from the previous paragraph. Go to your favorite Web-based translation site and enter, "you have cleanup duty today" in Japanese and translate it into Korean. But let's say you're a Japanese teacher who doesn't understand Korean, so you aren't sure if the translation is correct; to test it, you might use back translation, clicking on the tabs to translate the Korean translation back into Japanese again, which yields, "you should clean the classroom today." It seems a little rude, but it might be acceptable if accompanied with a smile. Let's try translating the Chinese translation in the same way. When we back translate it into Japanese, we might get the very strange sentence, "today, you remove something to do your duty." It seems the Japanese word "cleanup duty" isn't registered in this machine translator's dictionary.

Basically, machine translators are half-products. The obvious first step is to combine a domain-specific and community-specific multilingual dictionary with machine translators. Machine-translation-mediated communication might work better in high-context multicultural communities, such as an NPO/NGO working for particular international issues. Computer scientists can help overcome language barriers by creating machine translators that generalize various language phenomena; multicultural communities can then customize and use those translators to fit their own context by composing various language services worldwide.

Issues with Machine-Translation-Mediated Communication

Even if we can create a customized language environment, we still have a problem in that most available machine translators are for English and some other language. When we need to translate Asian phrases into European languages, we must first translate them into English, then the other European language. If we use back translation to check the translation's quality, we must perform translation four times: Asian to English, English to European, and back to English and then to the original Asian language. Good translation depends on luck — for example, when we translate the Japanese word "タコ," which means octopus, into German, the back translation returns "イカ," which means squid, two totally different sushi ingredients.

The main reason for mistranslation is the lack of consistency among forward/backward translations. Different machine translators are likely to have been developed by different companies or research institutions, so they independently select words in each translation. The same problem appears in machine-translation-mediated conversation: when we reply to what a friend said, he or she might receive our words as totally different from what we actually, literally said. *Echoing*, an important tool for the ratification process in lexical entrainment (the process of agreeing on a perspective on a referent) is disrupted, and it makes it difficult to create a common ground for conversation.[2]

Even if translation quality increases, we can't solve all communication problems through translation, so we must deepen our knowledge of different cultures to reach an assured mutual understanding. For example, we can translate the Japanese term "cleanup duty" into Portuguese, but it can still puzzle students because there's no such concept in Brazil. As is well known, deep linkage of one language to another is the first step in understanding, thus we need a system that associates machine translation results with various interpretations of concepts to help us better understand different cultures. I predict that Wikipedia in particular will become a great resource for intercultural collaboration when combined with machine translators because a large portion of Wikipedia articles will be provided in different languages and linked together.

References

1. T. Ishida, "Language Grid: An Infrastructure for Intercultural Collaboration," *IEEE/IPSJ Symp. Applications and the Internet* (SAINT 06), IEEE Press, 2006, pp. 96–100.
2. N. Yamashita et al., "Difficulties in Establishing Common Ground in Multiparty Groups using Machine Translation," *Proc. Int'l Conf. Human Factors in Computing Systems* (CHI 09), ACM Press, 2009, pp. 679–688.

Toru Ishida is a professor in Kyoto University's Department of Social Informatics and a leader of the NICT Language Grid project. He has been named an IEEE fellow for his contribution to autonomous agents and multiagent systems. His research interests include social information systems, digital cities, and intercultural collaboration. Ishida has a PhD in computer science from Kyoto University. Contact him at ishida@i.kyoto-u.ac.jp.

Towards an Integrated Architecture for Composite Language Services and Multiple Linguistic Processing Components

Arif Bramantoro[1], Ulrich Schäfer[2], Toru Ishida[1]
[1]Department of Social Informatics, Kyoto University, Japan
Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501, Japan
[2]Language Technology Lab, German Research Center for Artificial Intelligence, Germany
Campus D 3 1, Stuhlsatzenhausweg 3, D-66123 Saarbrücken, Germany
E-mail: arif@ai.soc.i.kyoto-u.ac.jp, ulrich.schaefer@dfki.de, ishida@i.kyoto-u.ac.jp

Abstract

Web services are increasingly being used in the natural language processing community as a way to increase the interoperability amongst language resources. This paper extends our previous work on integrating two different platforms, i.e. Heart of Gold and Language Grid. The Language Grid is an infrastructure built on top of the Internet to provide distributed language services. Heart of Gold is known as middleware architecture for integrating deep and shallow natural language processing components. The new feature of the integrated architecture is the combination of composite language services in the Language Grid and the multiple linguistic processing components in Heart of Gold to provide a better quality of language resources available on the Web. Thus, language resources with different characteristics can be combined based on the concept of service oriented computing with different treatment for each combination. Having Heart of Gold fully integrated in the Language Grid environment would contribute to the heterogeneity of language services.

1. Introduction

One of the wide implementations of Web Services is language service (Shimohata, et al., 2001). The number of language service available on the Web is inevitably increasing. Computer scientists have been trying to develop more and more infrastructures to improve the quality and accuracy of the services. To utilize the language service more robustly, we need to integrate multiple infrastructures. Two of the famous ongoing developments of language infrastructures are the Language Grid (Ishida, 2006) and HoG (Heart of Gold; Schäfer, 2006).

The Language Grid is a framework of collective intelligence built on service oriented architecture which enables access to various language services and language resources in the world based on a single powerful protocol, HTTP. For the Language Grid, the more language resources it has the better it is for the availability of composite services. Composite language service means the ability to create a new service by combining existing services.

Heart of Gold (HoG) is also a framework that bridges user application and external natural language processing (NLP) components regardless the depth of the linguistic analysis. This framework provides integration between deep and shallow NLP annotations. Deep NLP applies as much linguistic knowledge as possible to analyze natural language sentences (Pollard & Sag, 1994). On the other hand, shallow NLP neglects the use of the whole range of linguistic details, but concentrates on specific aspects.

Only few shallow tools such as ChaSen and TreeTagger are provided by the Language Grid so far. There are various natural language processing (NLP) functions in HoG which are not provided by the Language Grid, especially the efficient deep analyzers for various languages. Moreover, hybrid and composite workflows can be defined that consist of combinations of the language components, the main goals being increased robustness and computation of formal semantics representations of natural language utterances.

This paper proposes an enhancement of the integrated architecture of the Language Grid and HoG that extends our previous work presented at the 2008 International Conference on Web Services (Bramantoro et al., 2008). Previously, the integrated architecture only provides HoG as an atomic service unable to be combined with other services in the Language Grid. Now, we utilize the composite language services in the Language Grid together with the multiple linguistic processing components in HoG.

The main contributions of this paper are (*i*) interoperability among various language services by creating new possible composition between multiple linguistic processing components of HoG and composite language services of the Language Grid; (*ii*) a new functionality of language services available on the Web by enabling the substitution of language components in HoG with additional in the Language Grid and vice versa within integrated composition.

2. Integrated Architecture

We identify three general problems concerning the integration.

- HoG is a framework based on components, while the

- Language Grid is a service-oriented framework. We need to survey which architecture is suitable and reliable to accommodate these frameworks.
- The standard interfaces of these two frameworks are not the same. HoG provides XML annotations as output, while in the Language Grid standard interface there is no such type for output parameter.
- Both frameworks provide a processing strategy for language resources but in different ways. The Language Grid provides service workflows for composite language services, while HoG uses a compilable description language for composing multiple components.

To combine the two frameworks, a number of experiments were designed to combine HoG and the Language Grid. We found out that the best possible one for combining HoG and the Language Grid is by wrapping HoG as a Web service that can be accessed through the Language Grid. We proposed that the Language Grid can utilize HoG by adding it to the language resources layer, a layer where atomic services are wrapped and registered. Although it is not common in the Language Grid to have a composite service in this layer, the standard wrapping technique of the Language Grid requires doing so. Consequently, we have to treat HoG differently in this layer since it contains multiple NLP components that behave as composite services.

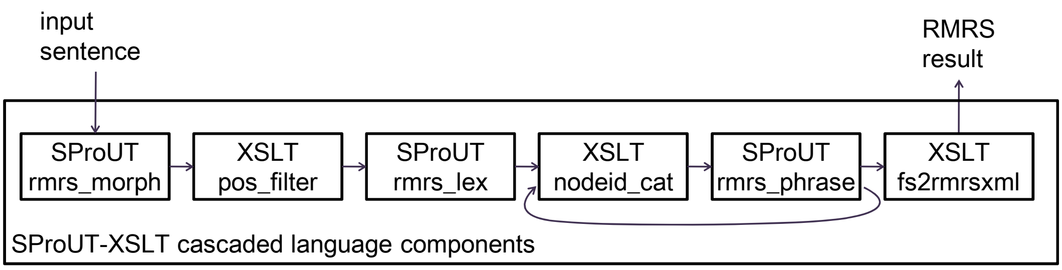
We create a new Web service that can connect to HoG and implement the Language Grid standard interface. From HoG's point of view, this Web service acts as an application, whilst from the Language Grid's point of view, this Web service is considered as a wrapped language resource. The wrapped Web service connects to the Module Communication Manager via XML RPC. Therefore, the HoG server can be located at any nodes in the Language Grid.

3. Processing Flow and Workflow

To get a higher quality of language processing we need to integrate more than one processing tool. HoG allows the user to execute more than one language component. In fact, this multiple component processing is the original characteristic of HoG since the default strategy is to execute the shallowest component first, then other components with increasing depth up to the requested depth. Unless a user defines smallest depth value, there is more than one language component executed.

There are three ways to configure the sequence of the components in HoG, (1) varying the depth value, (2) varying input and output, (3) using the SDL extension. In this paper, we focus on using SDL extension for running multiple components in a HoG service integrated in the Language Grid. It is impractical to implement the concept of depth value in service oriented computing. Moreover, Web services should be autonomous so that it is difficult to vary the input and output of language services during the composition.

SDL (System Description Language; Krieger, 2003), is a specific language initially used for building NLP systems and may be used in HoG to define sub-architectures of composite components. SDL uses a declarative specification language to define a flow of information (input and output) between linguistic processing components. The declarative specification consists of operators, symbolic module names, assignment of these symbolic module names to Java class names and constructor arguments. The basic operators currently available in HoG are + (*sequence*), | (*parallelism*), and * (*unrestricted iteration*). For example, multiple linguistic components consist of three SProUT grammar components and three XSLT transformation components described in Figure 1 together with its definition in SDL syntax.




```
chunkiermrs = ( sprout_rmrs_morph + xslt_pos_filter + sprout_rmrs_lex +
                (* xslt_nodeid_cat + sprout_rmrs_phrase ) + slt_fs2rmrsxml)

sprout_rmrs_morph = SproutModulesTextDom("rmrs-morph.cfg")
xslt_pos_filter = XsltModulesDomDom("posfilter.xsl", "aid", "Chunkie")
sprout_rmrs_lex = SproutModulesDomDom("rmrs-lex.cfg")
xslt_nodeid_cat = XsltModulesDomDom("nodeinfo.xsl", "aid", "Chunkie")
sprout_rmrs_phrase = SproutModulesDomDom("rmrs-phrase.cfg")
xslt_fs2rmrsxml = XsltModulesDomDom("fs2rmrsxml.xsl")
```

Figure 1: Composing NLP components in Heart of Gold with SDL

Composite services in the Language Grid are formulized in constraint satisfaction problem specification (Bramantoro & Ishida, 2009). Constraint satisfaction problem adopted from artificial intelligence theory is characterized with triplet entities (X, D, C) as follows:

- $X=\{X_1,\ldots,X_n\}$ is a set of abstract Web services, with $X_i.IN$ is a set of required input types, $X_i.OUT$ is a set of required output types, $X_i.QOS$ is a set of required QoS types. These requirements are defined as abstract service specifications..
- $D=\{D_1,\ldots,D_n\}$ where D_i a set of concrete Web services X_i that can perform the task of the corresponding abstract Web services.

  $D_i=\{s_{i1},\ldots,s_{ik}\}$ where s_{ij} is a concrete Web service of the corresponding X_i with $s_{ij}.IN$ is a set of provided input types, and $s_{ij}.OUT$ is a set of provided output types, $s_{ij}.QOS$ is a set of provided QoS types. In semantic matching of web service (Paolucci et al., 2002), every element of the input set in concrete service specification should be also an element of the input set in abstract service specification and every element of the output set in abstract service specification should be also an element of the output set in concrete service specification. We argue that in QoS based matching every element of the QoS set in abstract service specification should be also an element of the output set in concrete service specification. Therefore, we define semantically matched service specification as follows.

  $D_i=\{s_{ij} \mid s_{ij}.IN \subseteq X_i.IN \wedge X_i.OUT \subseteq s_{ij}.OUT \wedge X_i.QOS \subseteq s_{ij}.QOS\}$
- $C=\{C_1,\ldots,C_p\}$ is a set of constraints which consists of workflow control, QoS-related, provider-defined and user-defined constraints.

In the Web service composition, there are four possible controls of workflow, i.e. *sequence*, *split*, *choice* and *loop* that can be specified in a constraint satisfaction problem. For example, in order to increase the quality of translation, we can compose a translation service with the community dictionary service in the Language Grid as described in Figure 2.

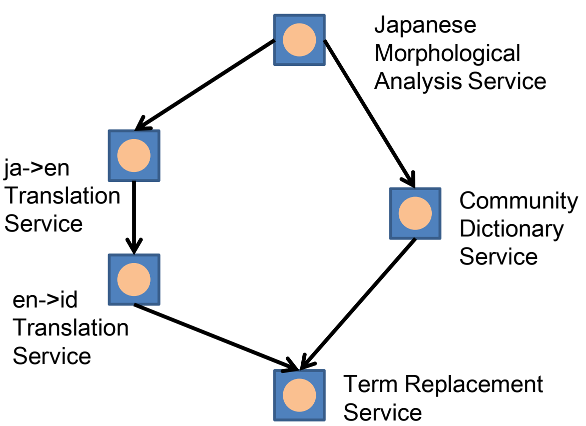


Figure 2: A workflow of specialized translation service between Japanese and Indonesian

The formulization for this workflow is as follows:

- $X=\{X_1, X_2, X_3, X_4, X_5\}$, where:
  - X_1: Morphological analyzer service;
  - X_2: ja-en translation service;
  - X_3: en-id translation service;
  - X_4: Community dictionary service;
  - X_5: Term replacement service;
- $D=\{D_1, D_2, D_3, D_4, D_5\}$, where (for the sake of simplicity, we omit the input and output parameters of D_i)
  - D_1: {mecab at NTT, ICTCLAS, KLT at Kookmin University, treetagger at IMS Stuttgart};
  - D_2: {JServer at Kyoto-U, JServer at NICT, WEB-Transer at Kyoto-U, WEB-Transer at NICT};
  - D_3 : {ToggleText at Kyoto-U, ToggleText at NICT};
  - D_4: {Science Dictionary, Natural Disasters Dictionary, Tourism Dictionary at NICT, Academic Terms Dictionary at NII};
  - D_5: {TermRepl service};
- C including (due to page limitation, only example constraints are shown)
  - C_1: For multi hop translation, $X_2.OUT=X_3.IN$;
  - C_2: For composite service which involves X_2 and X_4 (translation service and multilingual dictionary), serverLocation(X_2)=serverLocation(X_4);
  - C_3: For morphological analysis used together with community dictionary services, partialAnalyzedResult($X_1.OUT$) $\in X_4.IN$.

4. Combination of Two Flows

There are two urgent combinations between the multiple linguistic processing components of HoG service and composite language services in the Language Grid. These combinations involve the processing flow of HoG service and the workflow of the Language Grid.

Firstly, we need to incorporate composite components of HoG into the Language Grid's workflow. For example, there is a specialized Japanese-English translation service in the Language Grid that includes a Japanese morphological analyzer, an English morphological analyzer and some community dictionary services. The concrete Web service for English morphological analyzer available in the Language Gird is TreeTagger.

Multiple linguistic processing components (TreeTagger and RMRS) in HoG provide not only morphological analysis but also named entity recognition. This new functionality in the Language Grid's workflow enables users to dynamically select the right community dictionary service during workflow execution. Therefore, we can substitute the English morphological analyzer service in the workflow with the ones from HoG. To realize this combination, we have to instrument a new Web service in the workflow, i.e. an XML decoding service to detach the XML code in the HoG service output.

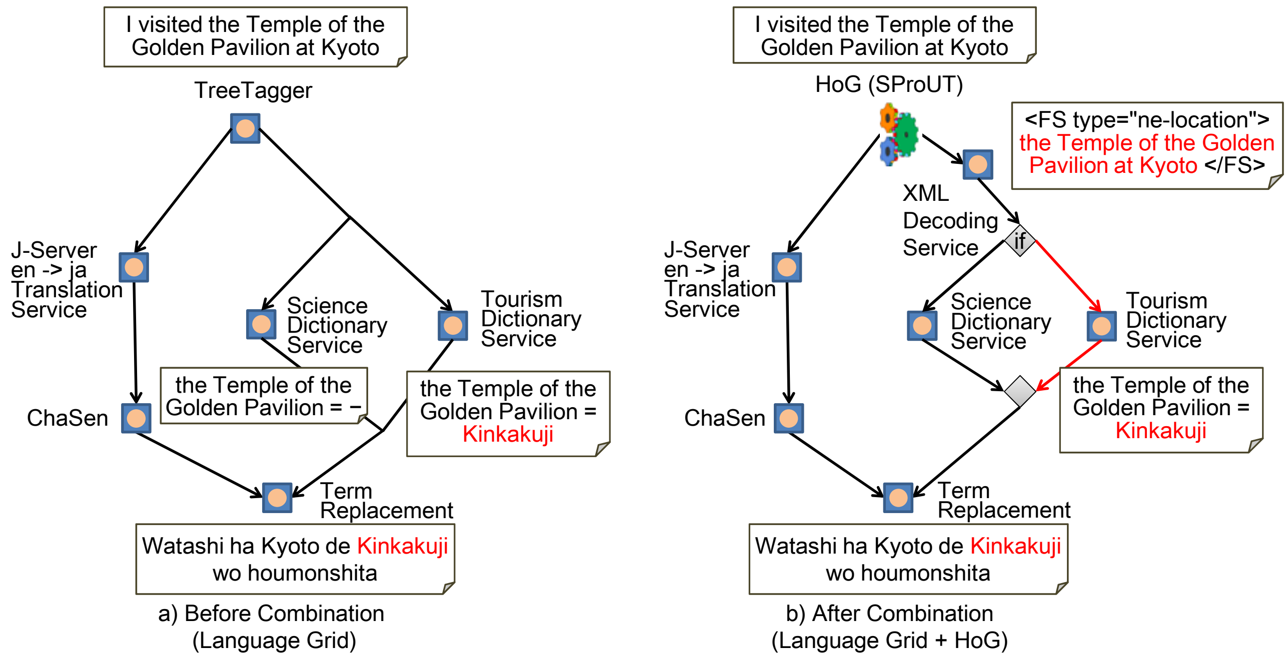


Figure 3: HoG composite components in the Language Grid's workflow

Figure 3 shows the scenario of combining HoG service in the Language Grid's workflow. In this scenario, a location term in the sentence could be detected and tagged by named entity recognition component (SProUT). When the location term is tagged by SProUT, the workflow execution engine automatically chooses Tourism Dictionary Service instead of Science Dictionary Service. The final result is the same as the existing workflow before combination, but the workflow execution by using HoG service should be more efficient since it runs one dictionary service in one time, not all dictionaries in parallel.

The scenario of using HoG service in the Language Grid workflow is also applicable to other dictionary services in the Language Grid. This could be realized by using the current tag set in the named entity recognition component related to the dictionary service or training a new tag set according to dictionary service entries. The integration will deliver efficiency since most of the community dictionary services are not free. Currently, there are more than 15 dictionary services available in the language grid. It should be costly to run all community dictionary services in each workflow without utilizing HoG service.

Secondly, we need to incorporate language service(s) of the Language Grid inside the processing flow of HoG. To do this, it is necessary to realize a mechanism of Service as a Software (SaaS) by wrapping language service(s) in the Language Grid as a HoG component that has additional parameters of XML output and, therefore, needs a special tool to convert the service output into XML format.

This integration is useful when we want to try the NLP components of HoG in different languages. For example, ChunkieRMRS in HoG is only available in German and English. Hence, deep NLP for Japanese could also be realized by utilizing Japanese-English translation service from the Language Grid (it is important to note that composite language service such as multi-hop translation service can be also wrapped as a language component) as described in Figure 4.

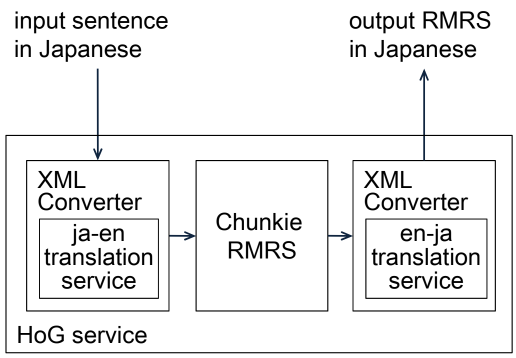


Figure 4: Language service inside HoG's processing flow

To realize the combinations, we propose a service and its architecture to integrate the processing flow and workflow. This service consists of processing flow analyzer, workflow analyzer and SDL writer. Three repositories are utilized by this service, i.e. language component information, language service information and extended workflow repository represented in constraint satisfaction problem.

An alternative workflow is automatically created and stored in the workflow repository together with its generated SDL description of incorporated HoG's components. When a user requests a particular task to be performed by composite language services, the processing flow & workflow integrator service analyzes an alternative workflow, enriches it with deeper composite language components provided by the HoG service, and calls SDL Writer to generate a new SDL description based on a new workflow combination to be delivered to the user. In addition, this integrator service can run offline so that the processing time of a user request is not affected since the new workflow has already been stored in the repository before runtime. The overall service architecture is illustrated in Figure 5.

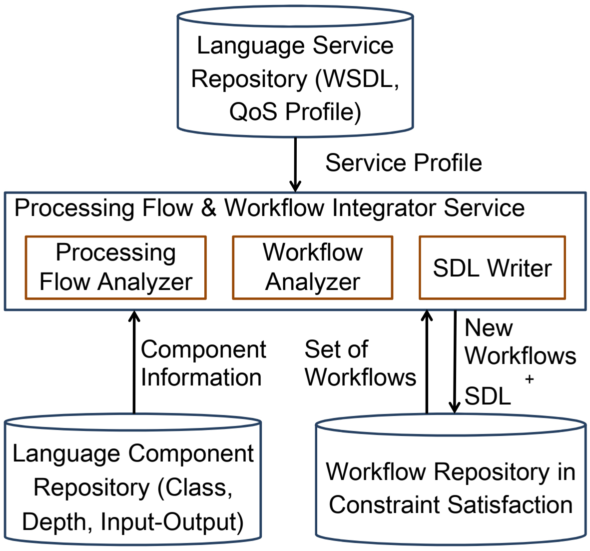


Figure 5: Integrator service architecture for composite language services and components

5. Related Work

We realize that there have been some breakthroughs in NLP researches that try to transform language software components into more loosely coupled components by using standard internet technology so called Web services. However, it is hard to find a good reference that provides a real solution for a complex integration task between a huge web service framework (the Language Grid) and a dynamic, highly customizable software system such as HoG.

Today's era is service oriented computing that creates everything as a service. There are many considerations to be examined before transforming software into a service. We can accommodate all language resources as a service but converting individual resources takes a lot of efforts as in the Language Grid. It is much easier to convert an existing platform that contains multiple language resources. Then, one would still be able to intervene inside the platform to choreograph individual resources.

A hybrid approach proposed by Jang et al. (2004) provides a workflow architecture based on Web services and object-oriented techniques. The authors argue that this architecture supports workflow systems with multiple process languages and standardized resource management. An interesting idea of this paper is the ability to support different web service-supporting process definition languages, such as BPML, XPDL, BPEL, and WSCI. This idea has been inspiring us to have different description languages in a single architecture. However, this paper only provides a few explanations on the implemented prototype.

A similar effort has been proposed in W3C to deal with different types of web services. Kavantzas et al. (2005) propose WS-CDL (Web Service Choreography Description Language) that is mainly used to integrate several web services from different providers, implementing different Web service technologies, such WS-BPEL and .Net C#. More specifically, WS-CDL supports the interoperability and interactions between web services in various programming languages and platforms within one business function by optimizing messaging between web services. This situation is different from what we face in the language domain. The Language Grid uses constraint satisfaction for its composite services. The HoG service is integrated into the Language Grid at a language resource layer (considered as atomic service), but contains composite components within its processing flow in SDL. Problems faced during the integration are not related to messaging between web services but mostly lie in transforming existing multiple linguistic processing components into machine-readable composite web services.

There is another candidate recommendation by W3C to define a new language, XProc (XML Processing Language; Walsh et al., 2009), to compose XML processes and deal with operations to be performed on XML documents. One of the advantages of this language is that it supports HTTP requests. By using this feature, this specification might be useful to integrate language services defined in WSDL and SOAP (both use XML over HTTP) and language components with XML output and called by XML-RPC. A specific pipeline can be created to process composite language services and multiple linguistic processing components at the same time. The concept of XProc is suitable to integrate two XML-based architectures, but currently there is no guarantee that XProc can fully support language services, especially for language services which are not merely an XML document.

Another open platform for natural language processing, Unstructured Information Management Architecture (UIMA) developed by IBM researchers (Ferrucci & Lally, 2004), enables association of each element of an unstructured document with semantic results of analysis. This paradigm can be adapted to the Language Grid. Any word in the source text translated by the Language Grid can be initially assigned a semantic value from UIMA. To give a simple example, the word "car" in a text document can be associated with multiple analysis engines, e.g. a morphological analysis and a translation engine. The result would be the word “car” with associated semantic values “noun:en” and “kuruma: en → ja”. These associations could be further processed by more advanced language-aware applications. Having two frameworks, HoG and UIMA, in the Language Grid could be another research topic, taking into account considerations on HoG and UIMA integration discussed in Schäfer (2008).

6. Conclusion

In this paper, we showed that language resources with different characteristic can be combined based on the concept of service oriented computing with different combinations. Multiple linguistic processing components in HoG can be combined with the existing workflow of composite services in the Language Grid environment. On the other hand, the composite language services in the Language Grid can be utilized in the processing flow of HoG components.

The next step that can be done on the basis of this prototype is to build more applications for visualizing computed annotation results. Currently, the return value of HoG service is an XML document, which is complicated for layman to understand and use. By providing client applications that process and visualize the XML result, the users of the Language Grid, not only linguists, could hopefully benefit better from natural language processing results returned by HoG.

7. Acknowledgements

This research was partially supported by Strategic Information and Communications R&D Promotion Programme from Ministry of Internal Affairs and Communications, and also from Global COE Program on Informatics Education and Research Center for Knowledge-Circulating Society.

The work described in this paper was partially supported by the German Federal Ministry of Education and Research under contract 01IW08003 (project TAKE: Technologies for Advanced Knowledge Extraction).

8. References

Bramantoro, A. & Ishida, T. (2009). User-Centered QoS in Combining Web Services for Interactive Domain, In *Proceedings of the International Conference on Semantics*, Knowledge and Grid, Zhuhai, China, October 2009, pp. 41-48.

Bramantoro, A., Tanaka, M., Murakami, Y., Schäfer, U., & Ishida, T. (2008). A Hybrid Integrated Architecture for Language Service Composition, In *Proceedings of the IEEE International Conference on Web Services*, Beijing, China, September 2008, pp. 345-352.

Ferrucci, D. & Lally, A. (2004). Building an Example Application with the Unstructured Information Management Architecture, *IBM Systems Journal*, 43(3), pp. 455–475.

Ishida, T. (2006). Language Grid: An Infrastructure for Intercultural Collaboration, In *Proceedings of the IEEE/IPSJ Symposium on Applications and the Internet*, Arizona, USA, January 2006, pp. 96-100.

Jang, J., Choi, Y., Zhao, J.L. (2004). An Extensible Workflow Architecture through Web Services, *International Journal of Web Services Research*, 1(2), pp. 1-15.

Kavantzas, N., Burdett, D., Ritzinger, G., Fletcher, T., Lafon, Y., & Barreto, C. (2005). Web Service Choreography Description Language (WS-CDL) Version 1.0, W3C Candidate Recommendation, World Wide Web Consortium. Retrieved November 9, 2009, from http://www.w3.org/TR/ws-cdl-10.

Krieger, H.-U. (2003). SDL—A Description Language for Building NLP Systems, In *Proceedings of the HLT-NAACL Workshop on the Software Engineering and Architecture of Language Technology Systems*, Edmonton, Canada, May 2003, pp. 84–91.

Paolucci, M., Kawamura, T., Payne, T.R., Sycara, K. (2002). Semantic Matching of Web Services Capabilities, In *Proceedings of the International Semantic Web Conference*, Sardinia, Italy, pp. 333-347.

Pollard, C. J. & Sag, I. A. (1994). *Head-Driven Phrase Structure Grammar*, University of Chicago Press.

Schäfer, U. (2006). Middleware for Creating and Combining Multi-dimensional NLP Markup. In *Proceedings of the EACL-2006 Workshop on Multi-Dimensional Markup in Natural Language Processing*. Trento, Italy, April 2006, pp. 81–84.

Schäfer, U. (2008). Shallow, Deep and Hybrid Processing with UIMA and Heart of Gold, In *Proceedings of the LREC-2008 Workshop Towards Enhanced Interoperability for Large HLT Systems: UIMA for NLP*, Marrakesh, Morocco, May 2008, pp. 43-50.

Shimohata, S., Kitamura, M., Sukehiro, T., & Murata, T. (2001). Collaborative Translation Environment on the Web, In *Proceedings of the Machine Translation Summit VIII*, Santiago de Compostela, Spain, September 2001, pp. 331-334.

Walsh, N., Milowski, A., & Ritzinger, S. T. (2009). XProc: An XML Pipeline Language, W3C Candidate Recommendation, World Wide Web Consortium. Retrieved December 7, 2009, from http://www.w3.org/TR/xproc/.

Language Grid Playground: Light Weight Building Blocks for Intercultural Collaboration

Satoshi Sakai[1], Masaki Gotou[1], Yohei Murakami[2], Satoshi Morimoto[1], Daisuke Morita[1], Masahiro Tanaka[1], Toru Ishida[1]

[1]Department of Social Informatics, Kyoto University
Yoshida-Honmachi, Sakyo-ku, 606-8501, Japan
[2]Language Grid Project, National Institute of Information and Communications Technology
3-5 Hikaridai, Seikacho Soraku-gun, Kyoto, 619-0289, Japan
{s-sakai, m-goto, morimoto, morita, mtanaka}@ai.soc.i.kyoto-u.ac.jp,
yohei@nict.go.jp, ishida@i.kyoto-u.ac.jp

ABSTRACT

Various types of multilingual collaboration tasks must be performed in the fields of education, medical care, and so on. Members in such fields need support customized for each field. Therefore, multilingual collaboration tools should allow customization to suit the tasks and circumstances. The tools provided by portal sites such as Google and Excite are not flexible enough to solve the problems in various fields because they fail to support customization. Therefore, we have developed the Language Grid Playground: an environment in which it is easy to make customized multilingual tools. The basic idea is to organize language services in a layered architecture and develop light weight building blocks that form collaboration tools by combining services. Our system, which is composed of components designed in this way, makes it easy to create tools customized for various intercultural collaboration fields. As a practical example, we develop a customized tool for the field of education in just 6 man-weeks. It confirms the efficiency of our approach for developing tools.

ACM Classification Keywords

D2.13. Reusable Software: Reusable libraries

General Terms

Design

Keywords

Web Service, Service Oriented Programming, Service Oriented Architecture, Intercultural Collaboration

INTRODUCTION

In recent years, the opportunities for international exchange and the number of multicultural communities have increased. There are various fields such as medical front-desks for foreign patients in hospitals and guidance for foreign students or parents in the field of education. In the multicultural communities, tools customized for tasks in each field are needed. They include a tool that can translate domain specific terms correctly and a tool with which users can make multilingual handouts. However, tools available on portal sites provide only general tools such as translation and dictionary tools. In other words, these tools cannot be customized for particular fields and consequently cannot solve collaboration problems in these fields. Our solution is the Language Grid Playground, which is an environment that makes it easy to develop multilingual tools customized for various scenarios; it rests on two basic approaches.

Organizing layered architecture of language services: The Language Grid project [1] creates various web services by wrapping language resources from all around the world. Moreover, in order to accumulate useful components which can compose language tools, we also wrap language services, each of which is composed of language resources, as web services and then share them. We organize these components by classifying the language services into four layers. This makes components more reusable, in other words, easier to search and easier to modify.

Developing building blocks with service-oriented programming: We provide several multilingual tools and accumulate useful components. We develop these components as programs which can be deployed as web services and publish them. By using these building blocks, people can easily develop multilingual tools that suit the tasks in the field of interest.

This paper introduces, as background, the Language Grid project. We then explain our approaches: the four layered architecture of language services and the service-oriented programming. We then introduce the Language Grid Playground. Finally, we describe the result of an experiment in which we create a customized tool by using the building blocks available on our system.

BACKGROUND

The Language Grid is an infrastructure for enabling users to create new language tools by combining web services that represent wrapped language resources published on the Internet. The Language Grid Association is organized as a

IWIC'09, February 20–21, 2009, Palo Alto, California, USA.
ACM 978-1-60558-502-4/09/02.

user group to discuss issues about the Language Grid from various perspectives and accumulate knowledge to better utilize it [4].

The Language Grid has two main structures. One, called the horizontal Language Grid, involves the combination of existing bilingual dictionaries or machine translation systems. It combines language resources and a language processing system for standard languages. The other one is called the vertical Language Grid. It involves concerns specific scenes of intercultural collaboration activities, which require new specialized language services. It enables the use of specific community dictionaries and parallel texts used in the field of intercultural collaboration.

LAYERED ARCHITECTURE OF LANGUAGE SERVICES

In the Language Grid project, many language resources are wrapped and presented as web services. However, there is a big gap between the functions that the language resources provide and the functions end users need. Therefore, the Language Grid offers composite web services, which are created by combining language resources, and lie in the language service. However, if these composite web services are constructed ad hoc, they will include many difficult-to-reuse services. In order to accumulate highly reusable language services, Murakami and Ishida classified them [3]. This architecture makes components more reusable. It means that users can easily select services which they need in each layer and replace a sub-service with an appropriate one chosen from a lot of interchangeable alternatives. Following their approach we reformed the layer structure as shown in Figure 1 and classified language services into four layers.

Resource Adaptation layer: The goal of the resource adaptation layer is to resolve the problems unique to each language resource, for example, No-sentence-break translation service which deletes all breaks.

Combination layer: This layer combines the adapted language resources. This layer offers abstract workflows that are domain independent, for example, multi-hop translation and translation with user dictionary.

Application layer: The goal of this layer is to create composite web services in order to solve the problems of specific domains. An example is a service that supports multilingual communication in hospitals. It retrieves medical question-and-answer pairs from adjacency pair services and translates them using medical parallel text services.

User Adaptation layer: The purpose of this layer is to provide language services customized for the intended end users by combining language resources. For example, the pictogram translation service created by combining pictogram dictionary [1] of NPO Pangaea and machine translation.

[1]Pictogram dictionary services take keywords as argument and returns binary data of the pictograms annotated by the keywords.

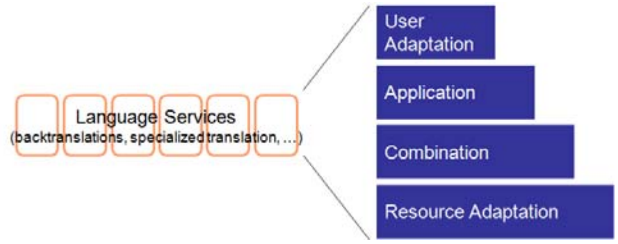


Figure 1. Four-layered language service architecture

SERVICE ORIENTED PROGRAMMING

There are so many organizations which need support their multilingual activities that it is impossible to make custom-made support tools for each of them from scratch because of the high cost. Therefore, customizing existing tools specified to specific organizations and making them usable in other communities are very important.

We can incorporate service oriented programming [5], a paradigm to integrate services on the Internet, to achieve this goal. In this paradigm, we can create a component to represent a service and components can be combined to create a new complex component and, finally, an application. In this approach, it becomes possible to break down an application into several components that are hierarchically organized. Moreover, components developed with this approach are reusable because they have appropriate grain sizes. In detail, their sizes are large enough because unskilled people can easily compose them, and are also small enough because each of them represents a single step in users' work. In addition, creating components by breaking down the processing of tasks into several services allows the structure of the components to be greatly simplified, i.e. they become light weight.

However, executing all components composing a tool in one environment gives a heavy load to the environment. Therefore, components need to be executed in distributed environments. Since each component created by service oriented programming provides a service, transforming these components into web services enables the decentralization. It is, however, difficult to describe workflows that are equal to complex components described in workflow description languages such as BPEL4WS [2]. Therefore, the first step in construction is to create highly reusable components as services using a simple scripting language such as PHP, which is much easier to code than BPEL4WS. The next step is to transform the highly reusable components into web services. These steps minimize the cost of describing workflows. Moreover, in order to minimize the cost of modifying processes implemented in these components into web service, they are developed as programs that can be transformed into web services. The components created in this way are regarded as building blocks. They enable the construction of systems easily.

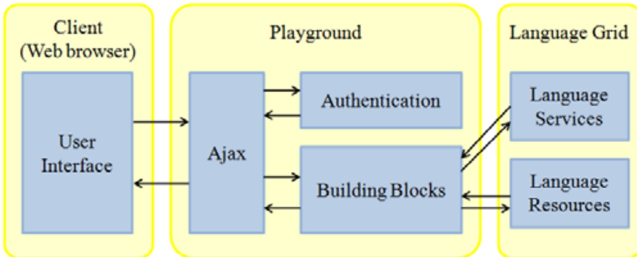


Figure 2. System Architecture

LANGUAGE GRID PLAYGROUND

We constructed the Language Grid Playground to supporting multicultural communities according to two approaches: layered architecture of language services and service oriented architecture.

System Architecture

The Language Grid Playground provides GUIs that are easily accessed via web browsers so end users can use language resources more easily. Figure 2 illustrates the system architecture of the system. Client side (web browser) is a GUI described in HTML and JavaScript. Playground side is described in PHP and Java and consists of three parts. Ajax part executes queries from the client side. At first, it authenticates users. If the authentication is successful, it sends a query to the building blocks part. Building blocks part gets the query and calls some language resources and language services.

Building Blocks

The Language Grid Playground provides tools which are classified into three categories. First category is called BASIC. Tools in this category provide GUIs that make it easy to invoke the language resources. Second category is called ADVANCED; these tools provide complex tools that call multiple language resources. The last category, CUSTOMIZED, has tools that provide functions customized to support intercultural collaboration activities in a certain community.

For constructing BASIC category tools, we created building blocks that enable the use of language resources. In ADVANCED category, we construct tools by combining building blocks used in tools in BASIC category and other new building blocks. One example is the composite translation services. The building blocks created for the composite translation services are *edit user dictionary*, *search of user dictionary, translation with user dictionary*, and *translation with public dictionary*. The composite translation services achieves multi-hop translation by using these building blocks. End user can raise the fluency and adequacy of translation with registering terms in the user dictionary and choosing public dictionary. The composite translation services also provides auto completion by using

Table 1. Building block list

Category	Building Blocks
BASIC	-search of public dictionary -search of parallel text -execution of translator -cross search of public dictionaries -cross search of parallel texts -cross execution of translators -adaptation to EDR -adaptation to Pangaea pictogram
ADVANCED	-edit user dictionary -search of user dictionary -translation with user dictionaries -translation with public dictionaries -translation with user and public dictionaries -back translation with user dictionaries -back translation with public dictionaries -back translation with user and public dictionaries

cross search of parallel texts block. When the end user inputs a part of a sentence, the building block searches for the text in the selected parallel text resources and the system displays the result below the input area. Table 1 shows a list of building blocks.

EXPERIMENT

We constructed custom pages by combining the building blocks. The pages were created to support Fujimi Junior High School. This school has 14 foreign students but few teachers can speak the students' mother tongue. Our solution was to construct a tool in which users can chat in their mother language. The page is shown in Figure3. In this system, students and teachers can chat by accessing the same GUI. Users can input text in their mother tongue, translate the sentence, check the back translation and post it to the log area at the top of the page. In addition, users can register the terms used in their school in the user dictionary, which makes the translations more correct.

In order to construct this system, we used several building blocks. The back translation functions are realized by using several *translation with user dictionary* blocks. This service provides parallel text auto completion by using BASIC building blocks. Moreover, the *edit user dictionary* block enables users in the school to create their own dictionary.

Usually, constructing such a system requires a lot of programming. However, combining light weight building blocks makes it easy to implement language processing parts of the system. In fact, time required for constructing this system was 6 man-weeks. Therefore it is proved that constructing by using light weight building blocks in our system is extremely useful. Besides this system, we created a glossary viewer system and a multilingual handout system for the same school.

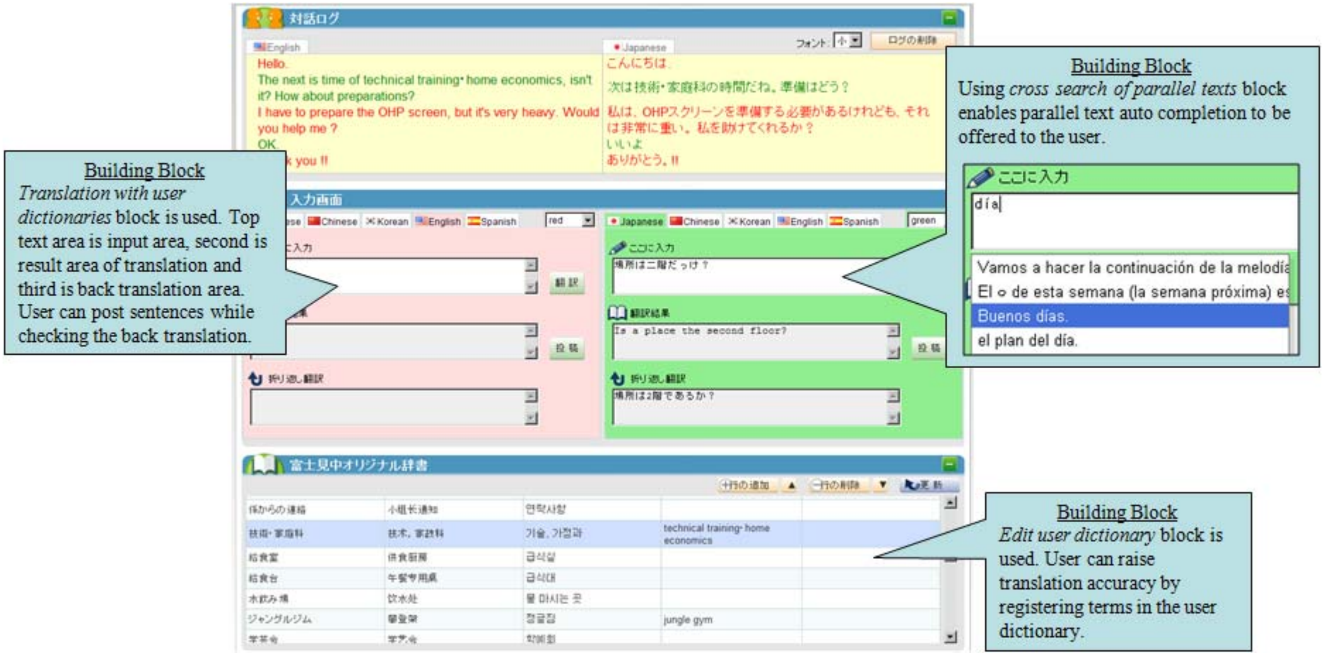


Figure 3. Custom page for Fujimi Junior High School

CONCLUSION

In order to realize effective multilingual communication we need support tools that can be easily customized to realize the tasks demanded by each multicultural community. It is, however, impossible to customize the tools provided by portal sites. To solve this problem, we constructed the Language Grid Playground as an environment in which tools can be easily customized for various tasks. The main contributions of this research are as follows

Achieving a layered architecture of language services Since existing language resources in the Language Grid provide only simple services such as translation, there is a big gap between these services and the tools needed by the end users. Therefore, we have created many language services in order to provide language tools for the actual fields. Moreover we make these language services easier to find and modify by classifying them into four layers.

Developing Building Blocks We created light weight building blocks using the service oriented programming approach. Moreover, we published the building blocks and a tutorial on how to use them in the Language Grid Playground web site. The site allows end users to construct customized tools for intercultural collaboration.

We have constructed a customized page by combining several building blocks. The time required for constructing this system was just 6 man-weeks. This proves that the light weight building blocks offered by the Language Grid Playground are extremely useful for constructing multilingual tools.

ACKNOWLEDGMENTS

This research was partially supported by Kyoto University Global COE Program on "Informatics Education and the Research Center for Knowledge-Circulating Society" and "Strategic Information and Communications R&D Promotion Programme" by the Japanese Ministry of Internal Affairs and Communications. And this report would not have been possible without the collaboration of the members of NICT Language Grid Project and Kyoto University Language Grid Operation Center.

REFERENCES

1. Ishida, T. Language Grid: An Infrastructure for Intercultural Collaboration. In *IEEE/IPSJ Symposium on Applications and the Internet (SAINT-06)*, pp.96-100, 2006.
2. Khalaf, R., Mukhi, N., and Weerawarana, S. "Service-Oriented Composition in BPEL4WS", *Proceedings of the World Wide Web Conference*, 2003.
3. Murakami, Y., and Ishida, T. A Layered Language Service Architecture for Intercultural Collaboration *The Sixth International Conference on Creating, Connecting and Collaborating through Computing (C5 2008)*, 2008.
4. Sakai, S. et al. Language Grid Association: Action Research on Supporting Multicultural Society, *International Conference on Informatics Education and Research for Knowledge-Circulating Society (ICKS'08)*, 2008.
5. Sillitti, A., Vernazza, T., and Succi, G. Service Oriented Programming: a New Paradigm of Software Reuse, *Seventh International Conference on Software Reuse (ICSR-7 2002), 2002.*

Difficulties in Establishing Common Ground in Multiparty Groups using Machine Translation

Naomi Yamashita[1], Rieko Inaba[2], Hideaki Kuzuoka[3], Toru Ishida[2,4]

[1] NTT Communication Science Labs.
Kyoto, Japan
naomi@cslab.kecl.ntt.co.jp

[2] National Institute of Information and Communications Technology
Kyoto, Japan
rieko.inaba@nict.go.jp

ABSTRACT

When people communicate in their native languages using machine translation, they face various problems in constructing common ground. This study investigates the difficulties of constructing common ground when multiparty groups (consisting of more than two language communities) communicate using machine translation. We compose triads whose members come from three different language communities—China, Korea, and Japan—and compare their referential communication under two conditions: in their shared second language (English) and in their native languages using machine translation. Consequently, our study suggests the importance of not only grounding between speaker and addressee but also grounding between addressees in constructing effective machine-translation-mediated communication. Furthermore, to successfully build common ground between addressees, it seems important for them to be able to monitor what is going on between a speaker and other addressees.

Author Keywords

Machine translation, Referential communication, Grounding, Computer-mediated communication.

ACM Classification Keywords

H.5.3 [**Group and Organization Interfaces**]: Computer-supported cooperative work, Synchronous interaction.

INTRODUCTION

Although communication technology has increased collaboration across international borders, language remains the biggest barrier to intercultural collaboration. In fact, most people have difficulty thinking and communicating in their non-native languages [20, 1].

For such people, machine translation appears to be an attractive technology, since it allows them to speak (write) and listen (read) in their native language. Indeed, an increasing number of multilingual organizations and Internet communities are proposing machine translation for communication support [8, 13]. One project that provides various language supports for such organizations is the "Language Grid Project [13]", which also served as a basis of this study.

Although machine translation liberates people from language barriers, it also poses hurdles to establishing mutual understanding. As one might expect, translation errors are the main source of inaccuracies that complicate mutual understanding [18]. In addition to translation errors, people have trouble constructing mutual understanding because they are not aware how each message is translated into other languages [19]. Furthermore, pairs have trouble grounding references because echoing and shortening of referring expressions are disrupted by asymmetries and inconsistencies in machine translation [22].

Although some novel solutions have been proposed [19, 13], machine translation still imposes excessive burdens on establishing mutual understanding. As a preliminary investigation, we interviewed members of an NPO [17] that has been using a machine-translation-embedded chat system to manage its overseas offices for almost two years. From these interviews, we found that they were facing particular difficulties when conducting multiparty group meetings. All of the interviewees mentioned that it was virtually impossible to conduct a group meeting when the total number of languages within the group was larger than two. For example, it seemed that members were easily left behind in the conversations of such meetings.

This study, inspired by these interviews, aims to clarify the reasons why machine-translation-mediated conversation is so difficult when the number of group members is larger than two. Research has demonstrated the difficulties of grounding references between pairs using machine translation [22]. Building on this previous work by expanding the experiment on referential communication from pairs to triads, we consider ways of supporting machine-translation-mediated collaboration for group work.

[3] University of Tsukuba, kuzuoka@iit.tsukuba.ac.jp
[4] Kyoto University, ishida@i.kyoto-u.ac.jp

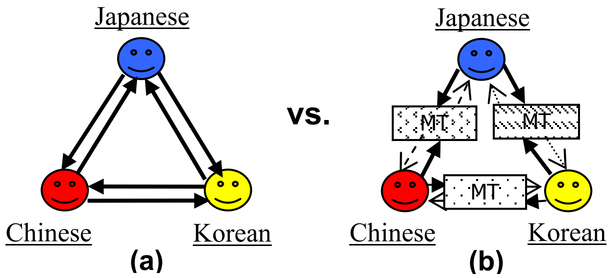


Figure 1 Three members communicating: (a) in their shared second language (English) or (b) in their native languages using machine translation software.

In the remainder of this paper, we first draw on prior research and predict how machine translation might affect referential communication within triads. Next, we describe a study that compares referential communication within triads in English (their shared second language) (Figure 1(a)) and referential communication within triads in their native languages using a machine-translation-embedded chat system (Figure 1(b)). We conclude with a discussion and issues raised by our study.

DIFFICULTIES IN ESTABLISHING COMMON GROUND IN MACHINE-TRANSLATION-MEDIATED COMMUNICATION

Common Ground

Regular Communication

Establishing *common ground* [4, 7, 6]—mutual knowledge, beliefs, assumptions, etc.—is important because communication is more efficient when participants share a greater amount of common ground [4, 9]. According to Clark and Marshall [6], people construct their common ground based on information they share by belonging to the same community, a shared physical setting (i.e., *physical co-presence*) or shared conversational content (i.e., *linguistic co-presence*). In each case, to successfully establish common ground, people not only must share the same information but also be aware that they are sharing this information with others [4, 15].

Grounding [4], then, refers to a process by which "common ground is updated in an orderly way, by each participant trying to establish that the others have understood their utterances well enough for the current purpose." During the grounding process, people become aware of what others do and do not know [5]. Such information helps them to formulate appropriate utterances, which leads to effective communication [5, 12].

In sum, for communicators to efficiently ground their utterances (particularly when members do not share the same physical space), the following three conditions must hold:

(1) they must share the same conversational content with others [4, 15]; (2) they must be aware that they are sharing the conversational content with others [4, 15]; and (3) they must be able to distinguish between information they do and do not share with others [5, 12].

Machine-Translation-Mediated Communication

It is important to satisfy the above three conditions in constructing common ground [4], but these conditions are not satisfied in machine-translation-mediated communication: As for condition (1), members cannot share the same conversational content because machine translation often mistranslates some parts of their utterances. As for condition (2), members cannot be aware whether they have the same conversational content, since they have no idea whether machine translation translated each utterance correctly into every language. Finally, as for condition (3), members cannot assess which parts of the utterance others do or do not understand because they have no idea where translation errors exist in other languages.

To improve machine-translation-mediated communication, researchers have proposed a novel solution called *back translation* [19]. Back translation offers speakers the awareness of how their utterances are translated into other languages by retranslating the translated utterances back to the speaker's language. Studies have demonstrated that the technique improves translation quality in machine-translation-mediated communication [19].

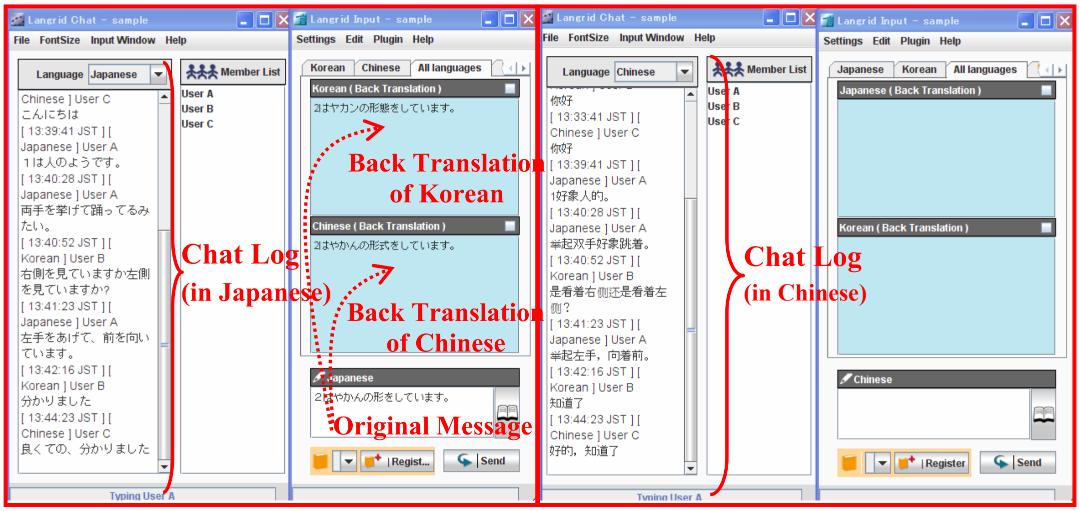
Despite this breakthrough, some problems remain unresolved in multiparty machine-translation-mediated communication. Even with the use of back translation, an addressee in a three-way machine-translation-mediated communication cannot monitor how the speaker's utterance is translated to the other addressee. For example, speaker A's message is translated into B's and C's languages simultaneously and back translations from both languages are shown to A. However, B (C) cannot monitor the translation between A and C (B). Consequently, conditions (2) and (3) do not hold between the two addressees: As for condition (2), the two addressees (B and C) cannot be aware whether they share the same information (i.e.. A's utterance); as for condition (3), addressee B (C) cannot be aware what addressee C (B) did and did not understand of A's utterance.

Since conditions (2) and (3), which are important in establishing common ground, do not hold in three-way machine-translation-mediated communication, it would clearly be difficult to build common ground, even with the use of back translation.

Referential Communication

Regular Communication

One type of communication that has been extensively studied to examine people's grounding process is "referential communication [7, 10, 14]." In referential communication, speakers and addressees work together to build common ground on a referent by adopting the same perspective [7]. Once speakers and addressees have enough evidence to believe that they are talking about the same thing, mapping is grounded between the referent and the perspective [3].



Japanese Interface Chinese Interface

Figure 2 Langrid Chat Interface (Japanese Director and Chinese Matcher)

The most basic task for examining referential communication is called the "referential communication task." Research applying this task typically studies how pairs arrange an identical set of figures into matching orders [7, 10, 14]. In each trial, one partner (the Director) is given a set of figures in a predetermined order. The other partner (the Matcher) is given the same figures in a random order. The Director must explain to the Matcher how to arrange the figures in the predetermined order. Typically, this matching task is repeated for several trials, each using the same figures but in different orders.

The process of agreeing on a perspective on a referent is known as *lexical entrainment* [3, 11]. Studies have shown that people make references based on historical factors such as recency, frequency of past references, and partner-specific conceptualization of the referent [2]. Studies have also shown that once communicators have entrained on a particular referring expression for a referent, they tend to abbreviate this expression in subsequent trials [2, 14].

Machine-Translation-Mediated Communication

Research on machine-translation-mediated communication has also studied referential communication between members of pairs. Yamashita [22] compared referential communication within pairs in English (their shared second language) and that within pairs in their native languages using machine translation software. Their results showed that lexical entrainment was disrupted in machine-translation-mediated communication because echoing was disrupted by asymmetries in machine translations. In addition, the process of shortening referring expressions was also disrupted because the translations did not produce the same terms consistently throughout the conversation.

Back translation can be used to alleviate the asymmetry issues because it offers speakers the awareness whether their utterances are symmetrically translated; when back translation does not yield the original expression, it implies that they cannot share the expression with others. While back translation may help communication within pairs, it is still unclear whether it improves communication within triads. Indeed, the NPO we interviewed had been using a machine-translation-embedded chat system with a back translation function, and they managed to conduct communication within language pairs; however, they said this was not possible within language triads.

As mentioned, we assume that problems peculiar to multiparty group communication arise when participants try to build common ground using machine translation; establishing common ground among multiple addressees would be difficult because addressees cannot monitor how the speaker's utterance is translated to the other addressees. To examine how this issue actually leads to real problems in the grounding process, we conducted an experiment using a machine-translation-embedded chat system with a back-translation function.

CURRENT STUDY

The present study builds on Yamashita's research [22] by expanding the experiment of referential communication from pairs to triads. We attempt to reveal how machine translation complicates referential communication within triads by comparing such communication in English (members' shared second language) and that in their native languages through machine translation software (Figure 1).

In the present task, three participants from three different language communities—China, Korea, and Japan—work together in a referential communication task in English or in their native languages. In the task, they must arrange an identical set of tangram figures into matching orders. In each trial, one participant (Director) is given a set of figures in a predetermined order, and the other two participants (Matchers) are given the same figures in different random orders. Using a multilingual chat system embedded with a back-translation function, the Director must explain to the Matchers how to arrange the figures in the predetermined order. Rotating the role of Director for each trial, this matching task is repeated for six trials (i.e., two cycles) using the same figures but in different orders.

Multilingual Chat System: Langrid Chat

For the experiment, we used a machine-translation-embedded chat system called "Langrid Chat [16]" (Figure 2). Langrid Chat translates each message into other languages while providing awareness information on the typing of other users. The machine-translation software embedded in Langrid Chat is a commercially available product that is rated as one of the very best translation programs on the market, in terms of translation quality. Langrid Chat is also equipped with a back-translation function: when a user types a sentence into the typing area, the system automatically translates the sentence into other languages, retranslates them back to the original language, and shows them to the user (Figure 2 (left)). Back translation is provided in real time so that users can edit their messages before sending them to others.

The chat interface allows each user to select his/her browsing and typing language from Chinese, English, Korean, and Japanese. For example, a Japanese participant who selects Japanese for his browsing and typing language can read and write in Japanese. Similarly, when a triad selects English as their browsing and typing language, they can both read and write in English.[1]

Hypotheses

We use quantitative and qualitative data analyses to examine three hypotheses:

In three-way machine-translation-mediated communication, machine translation translates each message into two other languages. Since translation from language A to B and translation from language A to C are carried out independently of each other, the original utterance in language A is often translated differently in language B than in C. In such conversations, two Matchers will not be able to share the same Director's utterance (i.e. condition (1) does not hold). Furthermore, they will not be aware whether they share the same Director's utterance (i.e., condition (2) does not hold). Under such conditions, we assume that participants will have trouble in identifying referents, leading them to low efficiency in their mutual acceptance process:

H1 (Efficiency of Mutual Acceptance Process): Participants will more efficiently identify a referent when using English rather than machine translation.

In the second cycle, each participant becomes the Director once again. When comparing referring expressions of the same participant between the first and second cycles, we expect that referring expressions will be shorter in the second cycle when using English because people often abbreviate referring expressions over time [2, 14]. However, we expect that abbreviation of referring expressions is at times very difficult when using machine translation for the following reason: Even when a Director A's referring expression is translated correctly to both Matchers (B and C), this does not ensure that the same referring expression will be correctly translated between B and C (i.e., condition (2) does not hold between the three participants); when B (or C) becomes the next Director, he or she might realize that the referring expression does not work between B and C, and thus change the referring expression to something else or add some details so that C (or B) understands it. Such changes in referring expression may complicate their mutual acceptance process, making it difficult to abbreviate their referring expressions:

H2 (Abbreviation of Referring Expressions over Trials): Participants will abbreviate their referring expressions more when using English than when using machine translation.

Not only is abbreviation difficult, but we also expect that making an appropriate reference (that would be smoothly identified by the Matchers) is also difficult when participants rotate their Director roles. When participants rotate their Director roles, the new Director (previous Matcher) typically explains each referent based on what he believes he shares with others [4]. However, in machine-translation-mediated communication, participants are less able to distinguish between information that they do and do not share with others (i.e., condition (3) does not hold). Therefore, we expect that the new Director will not be able to formulate appropriate references that would be smoothly identified by the Matchers:

H3 (Improvements in Making Appropriate References): Participants are less able to improve their efficiency of formulating appropriate references when using machine translation than when using English.

METHOD

Design

Thirteen triads (total of thirty-nine participants) from different language communities—China, Korea, and Japan—participated in the experiment. Nine triads participated in a referential communication task using their native languages through machine translation; four triads participated in the same referential communication task using a common language (English, which is not their native language). The experimental design was a between-subjects design for comparing referential communications carried out using the above two language methods.

Participants

Participants consisted of thirteen Chinese, thirteen Korean, and thirteen Japanese living in Japan. None of the participants knew each other before the experiment. Their English proficiency levels varied, but all of the participants had studied English for more than six years, and they were able to read and write basic English. They frequently used e-mail and instant messaging, but only a couple of them had used machine translation before the experiment. Participants were paid for their participation.

[1] Since machine translation automatically translates all messages, there is no difference in delay between conversation in English and using native languages.

Procedure

Step(1): On arrival, participants were taken to a room and asked to complete experimental consent forms. Next, participants were taken to a room partitioned into three compartments with a computer in each, and asked to sit in front of one of the computers. Participants were then given explanations of how to use Langrid Chat and an overview of the experiment. Participants were told that a) each person has the same set of figures in different orders; b) there are three roles: one Director and two Matchers; c) the Director must explain each figure one by one until both Matchers arrange their figures in the Director's order; d) the matching task is repeated six times using the same figures but in different orders, and each time the role of Director is rotated.

Step(2): As a pre-study, the participants engaged in a short-term referential communication task using three tangram figures (different from those used in Step(3)). The pre-study was conducted to let participants familiarize themselves with Langrid Chat.

Step(3): Triads were presented with eight tangram figures (Figure 3) arranged in different sequences, and they were instructed to match the arrangements of figures using Langrid Chat.

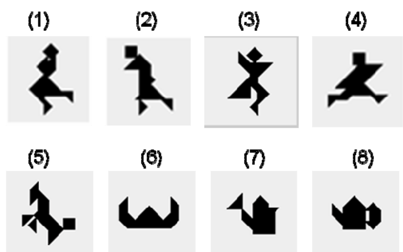


Figure 3. Eight tangram figures used in the experiment.

Rotating the role of Director for each trial, this matching task was repeated for six trials (i.e., two cycles) using the same figures but in different orders.

Step(4): Following the four matching tasks, participants were interviewed, as described below.

Please note that the experimental design was incomplete in that Director role was not counterbalanced for order; Japanese participants played the Director role for the first and fourth trial, Korean participants in the second and fifth trial, Chinese participants in the third and sixth trial.

Measures

Efficiency of Referential Communication. The triads were instructed to complete the task as efficiently as possible. We used the number of utterances (messages) per figure made by Directors to measure the efficiency of referential communication.

Abbreviation of Referring Expressions. We compared the length of referring expressions of the same Director between the first and second cycles and calculated the frequency of the Directors abbreviating their referring expressions. We did not compare the length of referring expressions between different Directors because the number of words differs among different languages even when they use the same expressions.

Improvements in Making Appropriate References. When Directors make appropriate references based on prior mutually accepted descriptions, Matchers should be able to identify the referents through the "*basic exchange* [7]" more frequently, where basic exchange is the most efficient way to identify a referent consisting of two steps: (a) the presentation of a referring expression and (b) its acceptance. To measure the appropriateness of each Director's reference, we calculated the proportion of basic exchange.

Interview. At the end of the experiment, we interviewed each participant separately using Japanese or English. When the participants had trouble understanding or speaking, bilingual translators translated our questions. There were no predetermined questions, but the topics covered the usefulness of the multilingual chat system (Langrid Chat), the ease of constructing and understanding utterances, and the strategies they used for effectively completing the task. The interview also helped to explain some specific incidents observed during the task.

RESULTS

Three groups were excluded from quantitative analysis since the members ran out of time and could not repeat the tasks for six trials using machine translation.

Efficiency of Referential Communication

Number of Utterances

Our first hypothesis *H1* stated that participants would more efficiently identify a referent when using English rather than machine translation. To test this hypothesis, the numbers of Director's utterances per figure were analyzed in a repeated measures ANOVA with Language Condition as a between-subjects factor[2]. Results indicated a significant main effect for Trial ($F[5, 40]=8.95$, $p<.001$) and a significant main effect for Language Condition ($F[1,8]=15.68$, $p=.001$) but no interactions.

As shown in Figure 4, the number of Director's utterances decreased over trials for both Language Conditions. As predicted by *H1*, however, it was proved that the Machine Translation condition yielded more utterances of a director compared to the English condition.

In forming our first hypothesis, we anticipated that participants would have trouble identifying referents through machine-translation-mediated communication due to the following two factors:

[2] Where ANOVA is carried out, the test for homogeneity of variance (Levene test) was also carried out. Unless reported, variances were equal between conditions ($p>.05$).

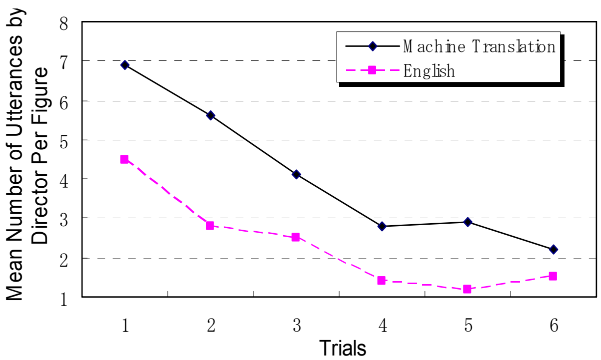


Figure 4. Mean number of utterances by a Director per figure.

- Two Matchers B and C will not be able to share the same Director A's utterance (i.e., condition (1) does not hold) because of the discrepancy in translation between A to B and A to C.

- Two Matchers B and C will not be aware of whether they share the same utterance of Director A (i.e., condition (2) does not hold).

To see how these factors actually affected referential communication, we examined the conversations in our experiment in further detail. In the following, we examine the impact of these factors one by one.

Places of Identifying Referents

When two Matchers do not share the same utterance of a Director (i.e., when condition (1) does not hold), Matchers may not be able to identify the referents based on the same Director's utterances. As expected, we found many cases in which Matchers identified the referents at different places in the conversation; specifically, one Matcher required more information and/or clarification than the other when using machine translation (Excerpt 1).

Excerpt 1. Matchers accepting Director's Proposal at Different Points of the Conversation (translated into English). Underline&Boldface indicates the originator of each message.

	Japanese Screen	Korean Screen	Chinese Screen
	<3rd trial> Director: Chinese		
1	C: A head is a square one.	C: The head is square.	**<u>C: Its head is square.</u>**
2	C: The edge run toward the right.	C: The vicinity is attached to the right.	**<u>C: It runs toward its right.</u>**
3	K: Is it the design to which you run?	**<u>K: Does it looks like running?</u>**	K: Is it after we assume that I compare and run?
4	**<u>J: I got it.</u>**	J: I got it.	J: I got it.
5	C: A lower back is the parallelogram.	C: A lower back is the parallelogram.	**<u>C: The lower back is the parallelogram.</u>**
6	K: I got it.	**<u>K: I got it.</u>**	K: I got it.

To understand what the participants were trying to communicate, we translated all messages into English. In addition, to share the automatically translated messages in this paper, we further translated the translated messages into English.

In the excerpt above, a Japanese Matcher and a Korean Matcher identified one of the Tangram figures based on a Chinese Director's explanation. In this trial, the Japanese Matcher identifies the figure in the 4th line, while the Korean Matcher identifies it in the 6th line. Although this was their third time to match the same figures, the Korean Matcher was late in identifying the figure, presumably because the Chinese Director's 2nd utterance made no sense to the Korean Matcher.

To see whether such a case (i.e., Matchers identifying a referent at different places in the conversation) occurred more frequently in machine-translation-mediated communication than in English, we counted the number of such cases for each trial and then performed a repeated measure ANOVA on those numbers.

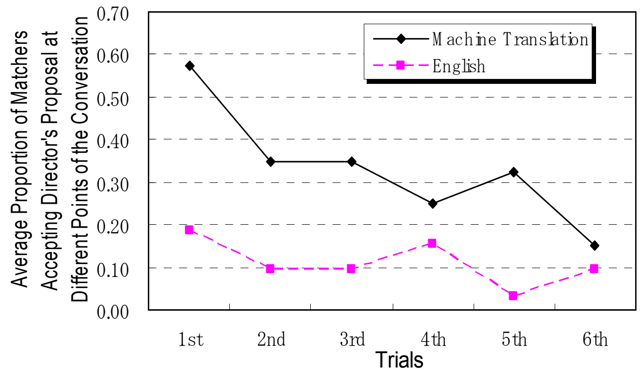


Figure 5. Average proportion of Matchers identifying a figure at different points in the conversation.

As shown in Figure 5, Matchers identified the referents at different points in the conversation more frequently in machine-translation-mediated communication than in English ($F[1,8]=15.99$, $p<.01$). We also found a significant main effect for Trial ($F[5, 40]=3.44$, $p<.05$) but no interactions.

Although back translation offered Directors the awareness of how their messages were translated into the other languages, it appeared from the interviews that rewriting their messages until the back translations of the two different languages reflected the meaning of the original message was difficult and time consuming. As a result, a Director's utterance was often translated differently to the two Matchers, leading them to identify the figures at different points in the conversations (i.e., based on different information). We speculate that such a tendency will increase as the number of languages increases in multiparty machine-translation-mediated communication.

Adaptation of References toward Others

From further observation, we found that referential communication using machine translation was even more inefficient because Matchers were not aware whether they shared the same Director's utterance (i.e., condition (2) did not hold).

Excerpt 2. Director not being able to coordinate his utterance toward the slow Matcher (translated into English). Underline&Boldface indicates the originator of each message.

	Japanese Screen	Korean Screen	Chinese Screen
<2nd trial> Director: Korean			
1	K: Looks like a pitcher.	**K: The shape of a pitcher.**	K: It's a financial aid person electron, an arm is done.
2	C: Sorry, not well understood.	C: Sorry. Not well understood.	**C: Sorry, I don't understand.**
3	K: The third one is swept when watering flowers.	**K: The third one is used when watering flowers.**	K: When giving water to a flower, the third is used.
4	**J: A sprinkler?**	J: A sprinkler?	J: Is this a sprinkler?
5	K: Yes.	**K: Yes.**	K: Yes.
6	C: The mouth was big.	C: The mouth became big.	**C: Its spout is big.**
7	K: The mouth is big.	**K: The mouth is big.**	K: The mouth is big.
8	**J: Is the mouth triangle?**	J: Is the mouth triangle?	J: Is the mouth triangle?
9	C: Got it, no problem.	C: Got it. No problem.	**C: Got it. No problem.**
10	K: Do you understand?	**K: Do you understand?**	K: Do you understand?
11	K: OK.	**K: OK.**	K: OK.
12	K: The mouth is triangle.	**K: The mouth is triangle.**	K: Mouth is triangle.
13	**J: I got it!**	J: I got it!	J: I got it!
<3rd trial> Director: Chinese			
14	C: A sprinkler.	C: A sprinkler.	**C: A sprinkler.**
15	C: Water was given and it was consumed.	C: Water was given and it was consumed.	**C: We use it for watering flowers.**
16	K: I got it.	**K: I got it.**	K: I got it.
17	C: The mouth is big.	C: The mouth is big.	**C: The spout is big.**
18	K: Yes, yes.	**K: Sure, sure.**	K: Nene.
19	K: It has a right triangle mouth, right?	**K: It has a right triangle mouth.**	K: You had a mouth of a right triangle, right?
20	**J: Sorry.**	J: Sorry.	J: Sorry.
21	**J: I got it.**	J: I got it.	J: I got it.

Matcher C, B often acquires knowledge of why C did not accept A's proposal concurrently with him or her by following the subsequent conversation between A and C. B makes use of such knowledge to coordinate his or her own utterances on the referent upon becoming the next Director [5]. However, such coordination was rarely observed in referential communication using machine translation.

In Excerpt 2, for example, a Japanese Matcher and a Chinese Matcher identify one of the Tangram figures based on a Korean Director's explanation. In this (second) trial, the Chinese Matcher identifies the figure in the 9th line, but the Japanese Matcher cannot identify it at the same timing. He asks the Director a question regarding the shape of the pitcher's spout (whether it is triangular) and manages to identify the figure in the 13th line. Although it is typically the case that the next Director coordinates his utterance (i.e., indicating that the pitcher's spout is triangular) so that the previous slow Matcher (i.e., the Japanese Matcher) can easily identify the referent, the Chinese Director in the consecutive trial did not do so. The Japanese Matcher finally manages to identify the figure with the help of the Korean Matcher.

Interestingly, the Korean Director's utterances were translated similarly to both Matchers in the second trial (from line 2). It is likely that the Chinese and the Japanese Matcher shared similar information regarding the Korean Director's utterance. Thus, if the Chinese Director (in the third trial) had coordinated his utterance indicating that the pitcher's spout was triangular, the Japanese Matcher would have been able to identify the figure more smoothly. We infer that the Chinese participant did not do so because he did not know whether he shared the same information with the Japanese Matcher in the second trial; maybe he could not understand why the Japanese Matcher could not accept the Korean Director's proposal concurrently with him in the second trial (whether because of translation error or other reasons), and thus he did not know what strategy to take. Similar cases were found elsewhere.

To examine whether such cases occurred more frequently in machine-translation-mediated communication than in English, we first extracted the cases in which Matchers differed in their places of accepting the Director's proposal. Then, for each case, two independent coders classified whether the next Director coordinated their utterances toward the previous slow Matcher. Since the coders only understood Japanese and English, they classified the transcripts of which Korean and Chinese utterances were translated into Japanese by bilingual translators. Agreement between the two coders was high (Cohen's Kappa values of the transcripts using English and machine translation were 0.91 and 0.95, respectively). We then calculated the rate of Directors coordinating their utterances toward the previous slow Matcher for each triad.

Overall, Directors coordinated their utterances toward the previous slow Matcher more when using English (Avg: 78.8%) than machine translation (Avg: 48.8%). A T-test showed a significant difference between the two language conditions ($t(8)=2.63$, $p<.05$). Since the previous slow Matchers often required further explanation when Directors did not coordinate their utterances toward them, we infer that such a lack of coordination of utterances was one reason leading them to inefficient communication requiring a large number of utterances to match the figures.

Abbreviation of Referring Expressions

Studies using referential communication tasks have shown that once a pair of communicators has entrained on a particular referring expression for a referent, they tend to abbreviate this expression on subsequent trials [2, 14]. However, we predicted in *H2* that abbreviation of referring expressions is difficult, particularly for triads using machine translation.

To examine *H2*, we compared the lengths of referring expressions of the same Director between the first and second cycles and classified for each referent whether the referring expression was (i) shortened (i.e., certain adjectives or/and explanations are eliminated), (ii) lengthened (i.e., certain adjectives or/and explanations are added), or (iii) other (identical or totally differentiated). For each participant, we calculated the rates of shortened and lengthened referring expressions.

Although the difference was not significant, participants shortened their referring expressions slightly more when using English (Avg: 45%) than machine translation (Avg: 31%) ($F[1,8]=3.98$, $p=.08$). As a more interesting finding, participants lengthened their referring expressions significantly more when using machine translation (Avg: 19%) than English (Avg: 6%) ($F[1,8]=5.21$, $p<.05$).

It seems that participants had trouble finding referring expressions that could be shared with all three members. Even in a case where a Director's reference was smoothly accepted by the Matchers in the first cycle, the Director sometimes lengthened his or her referent in the second cycle because the reference could not be used between the two Matchers (when one of the Matchers became the Director). The excerpt below captures this tendency.

In Excerpt 3, it appears that the Directors could not determine which terms to omit and which to leave (from 4th to 6th trial). We infer that Directors are reluctant to abbreviate their referring expressions once a new adjective or/and explanation is added during their mutual acceptance process, since they do not know which terms are translated correctly among all language pairs or why a new explanation has been added. To minimize their collaborative effort, it seems that they adopt a strategy of listing several references so that some parts of the list would be correctly translated in the translations of any language pair. We speculate that such difficulties in sharing the same reference will increase as the number of languages increases in multiparty machine-translation-mediated communication.

Improvements in Making Appropriate References

We hypothesized in *H3* that participants are less able to improve their efficiency in formulating appropriate references when using machine translation than when using English because they are less able to distinguish between information that they do and do not share with others (i.e., condition (3) does not hold).

We have already seen much evidence that making appropriate references is difficult. For example, coordinating their utterances toward the previous slow Matcher was difficult; finding a reference that could be shared between all members was also difficult.

Excerpt 3. Directors not being able to abbreviate their referring expressions (conversation is translated into English). Underline&Boldface indicates the originator of each message.

Japanese Screen	Korean Screen	Chinese Screen
<1st trial> Director: Japanese		
J: Number 2 is a horse.	J: Number 2 is a horse.	C: Number 2 is a horse.
<2nd trial> Director: Korean		
K: Number 4 is	**K: Number 4 is a person standing upside down.**	K: 4 times
--- *(snip)* ---		
J: Mr. B. Which number is the animal?	J: Mr. B. Which number is the animal?	J: Mr. B. Which number is the animal?
K: Animal?	**K: Animal?**	K: Animal?
--- *(snip)* ---		
J: Which number is the creature with a square tail?	J: Which number is the creature by which a tail is a square?	J: A tail, what number is a square creature?
C: An animal will be 8 days.	C: An animal is 8 days.	**C: Animal is number 8.**
K: I wouldn't know what to say, but something like an animal is 4 times most.	**K: I don't know what you are saying but the most animal like thing is number 4.**	K: Something like whatever animal says, is it wasteful, an unclear one is 4 times most.
<3rd trial> Director: Chinese		
C: It seems to be an animal.	C: It seems to be an animal.	**C: It looks like an animal.**
C: Horse	C: Horse	**C: Horse**
<4th trial> Director: Japanese		
J: Horse. Animal.	J: Horse. Animal.	J: Horse. Animal.
J: Tail is square.	J: A tail is square.	J: A tail is square.
<5th trial> Director: Korean		
K: It's an animal	**K: It's an animal.**	K: It's an animal.
K: It seems to be a word which raised its foreleg.	**K: It's a shape of a horse raising its front legs.**	K: A word is the design which entered a foreleg.
<6th trial> Director: Chinese		
C: Animal, it seems to be a horse.	C: Animal, it seems to be a horse.	**C: Animal, seems to be a horse.**
C: There is a square on the right side.	C: There is a square on the right side.	**C: There is a square on the right side.**

To see how much Directors improved in making appropriate references over trials, we calculated for each trial the rate of participants matching the figures through basic exchange (i.e., the most efficient way to match a figure: a Director proposing a reference and two Matchers accepting the reference immediately). Then, we performed a repeated measure ANOVA on those rates.

As shown in Figure 6, participants were able to match the figures more efficiently in English than in machine translation ($F[1,8])=61.43$, $p<.001$). We also found a significant main effect for Trial ($F[5, 40]=6.40$, $p<.01$) as well as a significant Language by Trial interaction ($F[5,40]=12.0$, $p<.001$). It appeared that Directors using machine translation had difficulty improving their references so that both Matchers could identify them immediately.

If Directors had used back translation more rigorously, the increasing rate of basic exchange could have been steeper. However, the problem does not lie only in the disinclination to use back translation. As previously mentioned, Directors were not aware which terms could be shared and which terms could not be shared with all of the members. Such unawareness impeded them from constructing appropriate references; even when they once

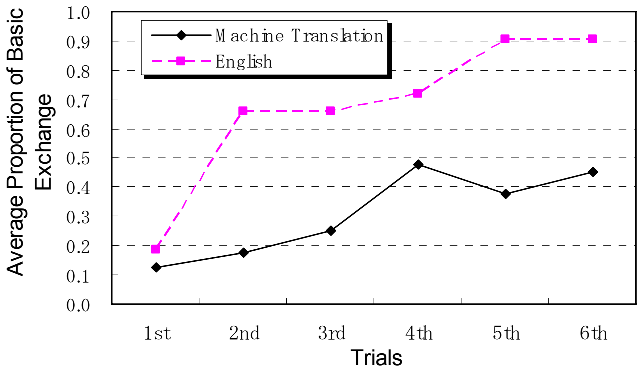


Figure 6. Average Proportion of Basic Exchange.

used a reference that could be shared among all of the members, they added redundant explanations when some problems occurred, and they were reluctant to shorten them because they were not aware which references could be shared among all members.

DISCUSSION

The goal of this study was to clarify why and how grounding conversations is difficult in machine translation-mediated multilingual triads.

Previous studies have documented the importance of satisfying the following conditions for communicators to successively build common ground: (1) they must share the same conversational content with others [4, 15]; (2) they must be aware that they are sharing the conversational content with others [4, 15]; (3) they must be able to distinguish between information they do and do not share with others [5, 12].

However, from our experiments, we found that satisfying these conditions was particularly difficult when the number of languages used in a group was larger than two. First, it appeared that condition (1) was often violated because of the discrepancy between translation from A to B and that from A to C. When condition (1) was violated, Matchers were not able to identify a referent at the same timing; one of the Matchers required more clarification for identifying the referent. Matchers tended to identify the referents based on different information. Furthermore, conditions (2) and (3) were often violated because participants using machine translation could not monitor how each utterance was translated into the other languages. Such a violation seemed to cause many problems in grounding references. In our experiment, we found three issues that seemed to arise from the violation of these conditions.

First, participants were not aware which parts of the conversational content they did and did not share with others. Under such a condition, we infer that Matchers had trouble understanding other Matchers' utterances (e.g., why a Matcher was asking for clarification) because they did not know the basis of their utterances. As a result, Directors were less likely to coordinate their utterances toward the previous slow Matcher. Second, participants were not aware which terms they could and could not share with all of the members. Under such a condition, it seemed that Directors could not determine which terms to omit and which terms to leave. As a result, Directors were less likely to abbreviate their referring expressions over trials. Finally, it appeared that participants using machine-translation-mediated communication had difficulty constructing appropriate (efficient) utterances because they could not distinguish between what they did and did not share with others. As a result, the participants' mutual acceptance process was inefficient and did not improve much compared to using English.

Although participants could always observe conversations between others through machine translation, it seemed that participants could not efficiently achieve mutual knowledge through indirect inferences. We speculate that one reason lies in the participants' behavior that they rarely provided back-channels or their status of understandings; when they had trouble understanding other participants' utterances, they ignored the utterance [22] or asked questions (instead of saying that they do not understand). This made them difficult to distinguish between shared and unshared information.

Theoretical Implications

Our study suggests the importance of not only grounding between speaker and addressee but also grounding between addressees in constructing effective machine-translation-mediated communication. When common ground is not well-established between addressees, communication is likely to become inefficient when they become a speaker. To successfully build common ground between addressees, it seems important for them to be able to monitor what is going on between a speaker and other addressees. By monitoring such conversation, they acquire knowledge of what others do and do not know. However, we speculate that being able to distinguish such knowledge is not sufficient for effective communication. When an addressee has trouble understanding a speaker's utterance, other addressees should be able to assess *why* the addressee fails to understand it by monitoring the conversation between speaker and the addressee (e.g., is it because of mistranslation or another reason?). When they are able to correctly assess the reason, they will be able to construct appropriate utterances that can be smoothly understood by others. We believe that knowledge of others (acquaintance relationships) and communicational context have a strong impact on participants' ability to assess such reasons.

Design Implications

Our findings and the above discussion suggest two recommendations for the design of future machine-translation-embedded communication systems to support group work.

- Provide speakers with an awareness of how their utterances are translated between addressees (i.e., whether the terms they are using can also be used between addressees).

- Provide addressees with an awareness of how a speaker's utterance is translated to other addressees using different languages (e.g., whether it is translated correctly or which part of the utterance is mistranslated).

One way of increasing mutual awareness among group members may be to share the video images of each participant's facial expressions. As shown in Veinott et al. study [21], video helps grounding between multilingual participants because it helps them assess other participants' level of understanding by providing their facial expressions.

For our future work, we are interested in investigating machine-translation-mediated communication which actually took place in the NPO that we have interviewed. In the long run, based on the findings from such investigations, we are hoping to contribute to the development of more effective machine-translation-mediated communication systems.

ACKNOWLEDGMENTS

This research was supported by the Kyoto University Global COE Program: Informatics Education and Research Center for Knowledge-Circulating Society. The authors would like to thank the Language Grid Project members, particularly Tomohiro Shigenobu for letting us use the multilingual chat system. We also thank the anonymous reviewers for their constructive comments.

REFERENCES

1. Aiken, M., Hwang, C., Paolillo, J., and Lu, L. A group decision support system for the Asian Pacific rim. *Journal of International Information Management*, 3, 1994, 1-13.
2. Brennan, S. E. Lexical Entrainment in Spontaneous Dialogue. *Proceedings of International Symposium on Spoken Dialogue*, 1996, 41-44.
3. Brennan, S. E. and Clark, H. H. Conceptual Pacts and Lexical Choice in Conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 6, 1996, 1482-1493.
4. Clark, H. H. *Using Language*. Cambridge, UK: Cambridge University Press, 1996.
5. Clark, H. H. and Haviland, S. E. Comprehension and the Given-New contract. *Discourse Production and Comprehension*, 1977,1-40.
6. Clark, H. H. and Marshall, C. E. Definite reference and mutual knowledge. *Elements of discourse understanding*, 1981, 10-63.
7. Clark, H. H. and Wilkes-Gibbs, D. Referring as a collaborative process. *Cognition*, 22, 1986, 1-39.
8. Climent, S., More, J., Oliver, A., Salvatierra, M., Sanchez, I., Taule, M., and Vallmanya, L. Bilingual Newsgroups in Catalonia: A Challenge for Machine Translation. *Journal of Computer Mediated Communication*, 9, 1, 2003.
9. Fussell, S., Krauss, R. Coordination of knowledge in communication: Effects of speakers' assumptions about what others know. *Journal of Personality and Social Psych*, 62, 3, 1992, 378-391.
10. Fussell, S., Kraut, R., and Siegel, J. Coordination of Communication: Effects of Shared Visual Context on Collaborative Work. *Proceedings of CSCW*, 2000, 21-30.
11. Garrod, S. and Anderson, A. Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, 27, 1987, 181-218.
12. Grice, H. P. Logic and conversation. *Syntax and Semantics, Vol. 3: Speech Acts*, Seminar Press, 1975, 113-127.
13. Ishida, T. Language Grid: An Infrastructure for Intercultural Collaboration. *IEEE/IPSJ Symposium on Applications and the Internet (SAINT-06)*, keynote address, 2006, 96-100.
14. Krauss, R. M. and Glucksberg, S. The development of communication: Competence as a function of age. *Child Development*, 40, 1969, 255-256.
15. Krauss, R. P. and Fussell, S. Mutual knowledge and communicative effectiveness. Intellectual Teamwork: Social and Technological Foundations of Cooperative Work, 1990, 111-146.
16. Langrid Chat: http://langrid.nict.go.jp/en/chat.html
17. NPO Pangaea: http://www.pangaean.org/
18. Ogden, B., Warner, J., Jin, W. and Sorge, J. Information Sharing Across Languages Using MITRE's TRiM Instant Messaging. 2003.
19. Shigenobu, T. Evaluation and Usability of Back Translation for Intercultural Communication. *International Conference on Human-Computer Interaction (HCII-07)*, 10, 2007, 259-265.
20. Takano, Y. and Noda, A. A temporary decline of thinking ability during foreign language processing. *Journal of Cross-Cultural Psychology*, 24, 1993, 445-462.
21. Veinott, S., Olson, J, Olson, G. and Fu, X. Video helps remote work: speakers who need to negotiate common ground benefit from seeing each other. *Proceedings of CHI*, 1999.
22. Yamashita, N. and Ishida, T. Effects of Machine Translation on Collaborative Work. *Proceedings of CSCW*, 2006.

Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence (IJCAI-09)

Context-Based Approach for Pivot Translation Services

Rie Tanaka
C&C Innovation Research Laboratories, NEC Corporation, Nara, 6300101, Japan
r-tanaka@ak.jp.nec.com

Yohei Murakami
National Institute of Information and Communications Technology (NICT), Kyoto, 6190289, Japan
yohei@nict.go.jp

Toru Ishida
Department of Social Informatics, Kyoto University, Kyoto 6068501, Japan
ishida@i.kyoto-u.ac.jp

Abstract

Machine translation services available on the Web are becoming increasingly popular. However, a *pivot translation service* is required to realize translations between non-English languages by cascading different translation services via English. As a result, the meaning of words often drifts due to the *inconsistency*, *asymmetry* and *intransitivity* of word selections among translation services. In this paper, we propose context-based coordination to maintain the consistency of word meanings during pivot translation services. First, we propose a method to automatically generate multilingual equivalent terms based on bilingual dictionaries and use generated terms to propagate context among combined translation services. Second, we show a multiagent architecture as one way of implementation, wherein a coordinator agent gathers and propagates context from/to a translation agent. We generated trilingual equivalent noun terms and implemented a Japanese-to-German-and-back translation, cascading into four translation services. The evaluation results showed that the generated terms can cover over 58% of all nouns. The translation quality was improved by 40% for all sentences, and the quality rating for all sentences increased by an average of 0.47 points on a five-point scale. These results indicate that we can realize consistent pivot translation services through context-based coordination based on existing services.

1 Introduction

Recently, the number of languages used in Web pages has increased rapidly. People using English on the Internet now comprise 30% of all Internet users; users of Asian languages comprise 26%; users of European languages excluding English comprise 25%; and users of all other languages comprise 20%.[1] This trend introduces the requirement for translations between non-English languages in addition to between English and non-English languages. Although the increase in the number of online translation services enables people to access machine translations easily, it is practically impossible to cover all combinations of *n* languages as the development of (n^2-n) direct translation services would be extremely costly. The pivot translation service generated by combining multiple translation services via a pivot language is a practical solution for such situation.

However, pivot translation often yields drifting for the meanings of words because of inconsistent word selection, making it difficult for users to continue communication. Establishing common ground among users in machine-translation-mediated communication is known to be difficult [Yamashita et al., 2009]; one of the causes of difficulty is inconsistent word selection [Yamashita and Ishida, 2006].

In phrase-based statistical machine translation (SMT), methods for pivot translation with no direct corpora between the source and target languages have been proposed [Utiyama and Isahara, 2007; Wu and Wang, 2007]. In their approach, the phrase-table required for SMT between the source and target languages is generated by combining phrase-tables between the source and pivot languages and the pivot and target languages. The phrase and lexical translation probabilities in the new table are estimated from original corpora, enabling more accurate selection of translated phrases. In the other approach for word selection problems, Kanayama and Watanabe [2003] proposed the linguistic annotation method. They embedded lexical and syntactical information for a source sentence into the intermediated sentence to assure the correctness of the pivot translation. However, the above approaches are not available immediately in practice because it is not easy to prepare the enormous and reliable corpora required to merge phrase tables or to apply the linguistic approach to all translation services. In contrast, we propose a method to realize consistent translation with available dictionaries and translation services.

To coordinate existing translation services, this study used the framework of service computing. In Web service composition, the *WS-Coordination* (Web Services Coordination)[2]

[1] The latest estimation of Internet users by language, carried out in May 2008 by Internet World Stats. See: http://www.internetworldstats.com/stats7.htm

[2] http://www-106.ibm.com/developerworks/library/ws-coor

specification enables the propagation of the service ID or port number as "CoordinationContext" to solve the semantic problems of service composition; it is also used to match input and output data types automatically [Hassine et al., 2006]. Moreover, the method of meta-level control for composite Web services in an open environment, known as "Service Supervision," has been proposed for designers who are not authorized to modify each component Web service [Tanaka et al., 2009]. In terms of improving the performance of composite Web services, a context-aware approach called situated Web service (SiWS) has been proposed to improve the performance of Web services with diverse interfaces and various clients [Matsumura et al., 2006]. We took this type of approach to coordinate word selection of whole component services with context from outside the Web services. In the development of machine translations or language resources, Bramantoro et al. [2008] proposed a method to combine language resources and middleware architecture to integrate deep and shallow natural language processing components. This approach uses both language resources and language processing component as Web services: our context-based coordination approach can contribute towards the improvement of combined services in such areas.

To solve the word selection problem in pivot translation services, we propose the *context-based coordination* method for translation services. We regard the internal translation processes of services as black boxes and realize the coordination outside the services instead of proposing a new machine translation technology. This study addresses the following issues.

Context-Based Coordination with Propagated Context
To ensure consistency in word selection, we propose the propagation of context across cascaded translation services by regarding the context as a set of multilingual equivalent terms. In the research area of bilingual dictionaries, methods to match the meanings of the words of different languages by combining multiple dictionaries are proposed. We refer to those methods and propose a method to generate the multilingual equivalent terms automatically based on commercially available bilingual dictionaries.

Multiagent Architecture for Coordination
This paper proposes a multiagent architecture as one way to implement context-based coordination, wherein the coordinator agent gathers and propagates the context from/to translation agents.

We implemented a coordinated Japanese-to-German-and-back translation service by cascading four translation services and obtained results indicating that the translation quality improved substantially. The advantage of this approach is that high-quality translations can be extracted from existing translation services with existing bilingual dictionaries without modifying their internal coding systems.

<Case 1>
Source sentence (English): Please add that picture in this paper.
□ Translation (Japanese): *douzo, sono shashin wo kono ronbun no naka ni tsuika shinasai.*
(Please add that picture in this thesis.)
<Case 2>
Source sentence (English): Please send me this paper.
□ Translation (Japanese): *douzo, kono kami wo watashi ni okuri nasai.*
(Please send me this paper.)

(a) Inconsistency in word selection

• Japanese user (Japanese): *kinou watashi tachi ha pa-thi wo sita.*
(We had a party yesterday.)
□ Translation (English): There was a party yesterday.

• English user (English): How was the party?
□ Translation (Japanese): *tou ha doudesita ka?*
(How was the political party?)

(b) Asymmetry in word selection

Source sentence (Japanese): *kanojo no ketten ha ookina mondai da.*
(Her fault is a big problem.)
□ Translation (English): Her fault is a big problem.
□ Translation (German): *Ihre Schuld ist ein großes Problem.*
(Her responsibility is a big problem.)

(c) Intransitivity in word selection

Figure 1. Issues in composite translation services

2 Overview of Context-Based Approach

2.1 Issues in Composite Translation Services

We conducted several experiments using the Language Grid [Ishida, 2006] and classified word selection errors into three categories: inconsistency, asymmetry, and intransitivity. *Inconsistency* is when translations of the same source word vary in different sentences. *Asymmetry* is when the back-translated word is different from the source word. The impact of these errors on communication has already been analyzed [Yamashita and Ishida, 2006]. Quantitative results with interview data show that lexical entrainment [Brennan and Clark, 1996] is disrupted by asymmetries in machine translations since they interfere with echoing. *Intransitivity* is when the word sense drifts across the cascaded machine translators.

Figure 1 presents examples of common problems encountered by cascaded translation services. All original Japanese and German sentences in this paper are italicized and their English translations are provided in parentheses. (a) is an example of inconsistency, wherein the English word "paper" is translated to the Japanese word *ronbun* (thesis) in Case 1, while the same word is translated into *kami* (paper) in Case 2. Asymmetry is presented in (b). In the first step of the machine translation-mediated communication, the Japanese word *pa-thi* (party), which means a social gathering, is translated into English correctly. However, when an English user echoes the word "party," it is translated into the Japanese word *tou* (political party). Intransitivity is presented in (c). The Japanese word *ketten* (fault), which means a weakness of character, is translated into English correctly, but mistranslated to the German word *Schuld* (responsibility). This is because the intermediate English word "fault" has several meanings, and the English-German translator does not have any knowledge of the context for the preceding Japanese-English translation.

2.2 Context-Based Pivot Translation Service with Multiagent Architecture

We propose a multiagent architecture for context-based pivot translation service, as shown in figure 2. The coordinator agent, which plays the role of controlling the whole translation, gathers and propagates context from/to the translation agents in addition to requesting them to translate the sentence. It possesses all possible contexts internally, selects all contexts that suit the context reported by the translation agent, and transfers them to the next translation agent. Translation agents possess the in-built functionality for the original translation service; they perform translations by taking into account the context provided by the coordinator agent, update the context, and transfer the result to the coordinator agent. They have knowledge of the languages and make language-specific processes or decisions. By using the agent framework, more advanced improvements are possible: for instance, adding the ability to interact with users in order to identify the context of the sentence.

Context can be represented in several ways, such as a set of characteristic words in a document, surrounding text, or talk of an expression. Since context in one language can be translated to other languages with multilingual equivalent terms, we represent context by sets of equivalent terms, not sets of terms in one language. In our architecture, we consider a set of terms in the source sentence as context in the source language and use equivalent terms as propagated context.

3 Generating Multilingual Equivalent Terms

The set of equivalent terms can be generated by analyzing generic bilingual dictionaries.[3] However, since it is costly and difficult to manually develop multilingual dictionaries that include all words in all languages, we require an automated method to develop such a dictionary. In previous work on this subject, the concepts for different languages were matched using bilingual dictionaries [Tokunaga and Tanaka, 1990]. We extended this idea to generate a set of trilingual equivalent terms (referred to hereafter as a *triple*). We represent mappings of words belonging to different languages in the form of a graph; a word is represented as a vertex, and a

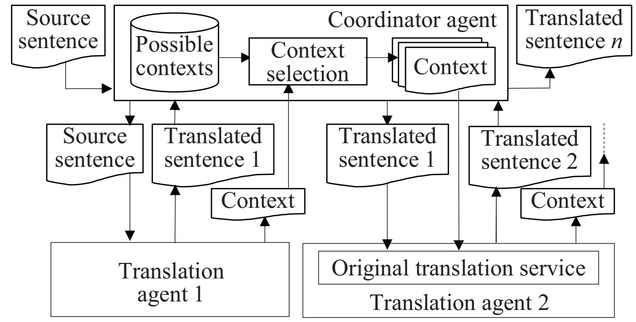


Figure 2. Multiagent architecture for context-based approach

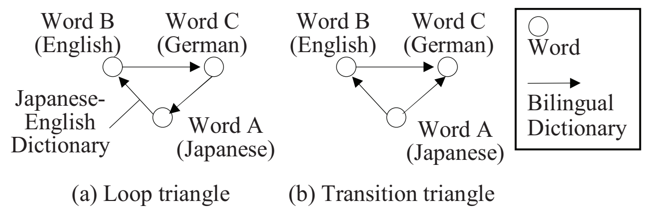


Figure 3. Two types of shapes of triangles

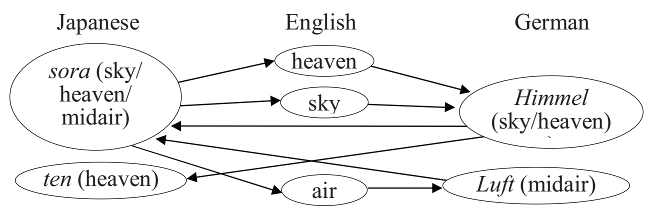


Figure 4. A loop triangle representing the sense of "sky•

mapping in bilingual dictionaries is represented as a directed edge. If the graph contains a triangle, the three words are considered equivalent terms. Figure 3 shows the two types of triangles: *loop* and *transition*. The loop triangle starts from a source language, looks up dictionaries three times, and returning to the source language. The transition triangle starts from a source language and looks up dictionaries to locate transitive and direct routes between the source and target languages. It is easy to generate a triple from such triangles. We call such triples generated from loop triangles loop-type triples hereafter.

Example 1 (A loop triangle representing "sky")

Figure 4 shows an example of a loop triangle, starting with the Japanese word *sora* (sky/heaven/midair). Words such as "sky" are extracted by looking up a Japanese-English dictionary. The German word *Himmel* (sky/heaven) is obtained by looking up the word "sky" in an English-German dictionary. Since the source Japanese word is extracted from a German-Japanese dictionary, {*sora* (sky/heaven/midair), sky, *Himmel* (sky/heaven)} is considered as a triple. Continuing this process further yields other triples.

[3] Multilingual equivalent terms can also be developed manually, as in the case of• EuroWordNet [Vossen, 1998].

Algorithm 1: COORDINATOR-AGENT *CA*

```
s_i /* Source sentence */
o_i /* A word in sentence s_i */
MTA /* An ordered list of translation agents
        (MTA = {MTA_1, MTA_2, ..., MTA_n}) */
MTA_i = {(s_i, s_i+1)} /* A translation agent; a set of pairs of
        sentence s_i and s_i+1 */
T_i /* A set of n-tuples (w_1, w_2, ..., w_n), where w_k is included in s_k
        (k□i); All n-tuples are n-lingual equivalent terms */
Q_k /* A set of pairs (o_i, m_i+1), where o_i∈s_i and m_i+1 is the
        modified translated word for o_i */
when received (ask, s_1) from user do
    T_1←{(w_1, w_2, ..., w_n)| w_1∈s_1};
    for each MTA_i in MTA do
        send (request, (s_i, T_i)) to MTA_i;
        when received (response, (s_i+1, Q_i)) do;
            T_i+1←SELECT-POSSIBLE-N-TUPLES (T_i, Q_i);
        end do;
    end loop;
    send (reply, s_n+1) to user;
end do;
```

Algorithm 2: SELECT-POSSIBLE-N-TUPLES (T_i, Q_i) **return** T_{i+1}

```
T_i+1←□;
for each pair (o_i, m_i+1) in Q_i do
    T_i+1←T_i+1 •  {(w_1, w_2, ..., w_n)|( w_1, w_2, ..., w_n)∈ T_i,
                    w_i=o_i and w_i+1=m_i+1};
end loop;
return T_i+1;
```

Figure 5. Algorithms of the coordinator agent *CA*

Algorithm 3: SERVICE-AGENT MTAi

```
t_i /* Translated sentence */
MT_i={(s_i, t_i)} /* A translation service; a set of pairs of s_i and t_i */
c_i+1 /* A word in sentence t_i */
P_i /* A set of pairs (o_i, c_i+1), where o_i∈s_i and c_i+1∈t_i */
when received (request, (s_i, T_i)) from CA do
    t_i←MT_i(s_i);
    P_i←GET-WORD-PAIRS-USED-BY-MT (s_i, t_i);
    Q_i←CREATE-WORD-PAIRS-TO-BE-USED (P_i, T_i);
    if Q_i•  P_i then
        s_i+1←MODIFY-TRANSLATED-SENTENCE (t_i,P_i, Q_i);
    else s_i+1←t_i;
    end if ;
    send (response, (s_i+1, Q_i)) to CA;
end do;
```

Algorithm 4: CREATE-WORD-PAIRS-TO-BE-USED (P_i, T_i) **return** *Qi*

```
Q_i←□;
for each pair (o_i, c_i+1) in P_i do
    for each n-tuple (w_1, w_2, ..., w_n) in T_i do
        if o_i∈(w_1, w_2, ..., w_n) and c_i+1∈(w_1, w_2, ..., w_n) then
            Q_i←Q_i •  {(o_i, c_i+1)};
        end if;
    end loop;
    if (o_i, c_i+1)∉Q_i then
        m_i+1←i+1th word in n-tuple selected from
                {( w_1, w_2, ..., w_n)|o_i∈(w_1, w_2, ..., w_n)};
        Q_i←Q_i •  {(o_i, m_i+1)};
    end if;
end loop;
return Q_i;
```

Figure 6. Algorithms of the translation agent *MTA*

This method can easily be extended to four or more languages by combining triples generated in each of the three languages similar to the extension approach proposed by Wu *et al*. [Wu *et al*., 2008]. For example, for Japanese, English, German, and French words, Japanese-English-German triples are obtained first followed by English-German-French triples. The quadruple is generated by combining two triples with identical English and German words. It is noteworthy that a triangle does not always imply equivalent terms. In the case where word A has word sense C_1 and C_2, word B has C_2 and C_3, and word C has C_3 and C_1, no shared sense exists between the three words. Assume that each word in a triple has n senses with uniform distribution, the probability of sharing the same sense is .83 for $n = 2$ and 0.91 for $n = 3$; this probability approaches 1 as n increases. In practice, the term frequencies of n senses are unequal, and the actual probability is higher than the calculated one. Thus we can obtain reliable equivalent terms by combining triples if the number of languages increases.

In related research on dictionary formulation, a method to construct a bilingual dictionary using a third language as an intermediate is proposed [Tanaka and Umemura, 1994]. This study takes the example of generating a Japanese-French dictionary by connecting Japanese-English and English-French dictionaries. It addresses the problem that a French word with a meaning different from that of the original Japanese word is obtained due to ambiguity in the intermediate English word; this problem is solved through inverse consultation with French-English and English-Japanese dictionaries. We focus on obtaining more reliable equivalent terms when dictionaries exist between each pair of languages and differ from the above research in terms of our assumptions and objectives. In order to realize coordination even when sufficient dictionaries are not available, methods such as inverse consultation are required to obtain equivalent terms.

4 Context-based Coordination Algorithms

Algorithms of the multiagent architecture for the context-based coordination are shown in figure 5 and 6. These algorithms are simple implementations of our multiagent model. Let machine translator MT_i input source sentence s_i and output translated sentence t_i. Let the translation agent MTA_i receives source sentence s_i, generate and modify t_i, and output s_{i+1}, which is a source sentence of MTA_{i+1}. Let the coordinator agent *CA* repeat the coordination process from MTA_1 to MTA_n and receive s_{n+1} as the final result in the target language. Multilingual equivalent terms in n languages are grouped into n-tuples. The context T_i is a set of n-tuples and the i-th word in each n-tuple in T_i is included in s_i. In a n-tuple (w_1, ..., w_n), the words w_2, ..., w_n have the same meaning as w_1 i.e. the same meaning as original sentence s_1, and their use assures the correct translation.

First, *CA* prepares the initial context T_1 from s_1 received from the user and starts translation. After MTA_i returns the translated sentence s_{i+1} and Q_i—representing word pairs of the source word in s_i and translated word in s_{i+1}—*CA*

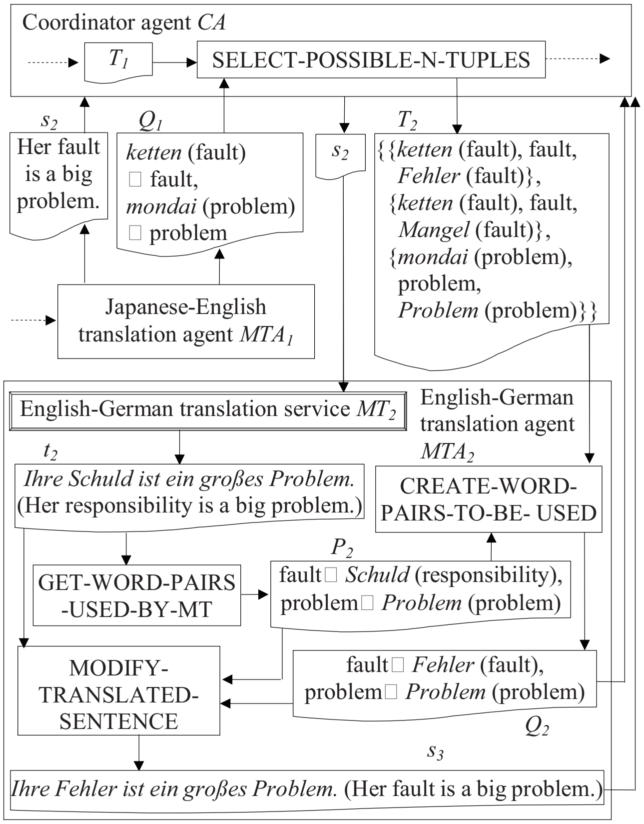


Figure 7. Example of Coordinated Translation Services

generates a new context T_{i+1} for the $i+1$-th translation by narrowing down T_i such that the $i+1$-th word in each n-tuple appears in s_{i+1} by the SELECT-POSSIBLE-N-TUPLES procedure. T_{i+1} may contain ambiguity in word selection for the $i+2$-th word, as more than two n-tuples containing the same j-th word ($1 \le j \le i+1$) can exist with different $i+2$-th words. If there are several candidates for the $i+2$-th word, the $i+1$-th translation agent MTA_{i+1} determines the most appropriate one. The choice is noted to *CA* by Q_{i+1}, and *CA* reflects it to the next translation.

MTA_i generates a translated sentence t_i using MT_i to create P_i—a set of word pairs of source word o_i and translated word c_{i+1}—using the GET-WORD-PAIRS-USED-BY-MT procedure. One way to implement this function is to divide s_i and t_i into morphemes and map between them using bilingual dictionaries. Then, MTA_i modifies words in P_i based on the context T_i using the procedure CREATE-WORD-PAIRS-TO-BE-USED and Q_i. Since T_i preserves the words used in the preceding i translations, the translated words excluded from T_i may have different meanings. Such words are replaced by words included in T_i, selected from among a few candidates if T_i contains ambiguity. Finally, t_i is modified by the procedure MODIFY-TRANSLATED-SENTENCE, wherein the words are replaced using P_i and Q_i. The word selection process can be improved through several methods: for instance, by referring to term frequency or priority of words, in case the translation agent possesses this information. If the entire document or conversation logs are available, this information can be utilized by *CA* to create an initial context T_1.

Example 2 (Context-based translation)
We show the translation process for the sentence shown in figure 1(c). In this example, the replacement of target words is limited to nouns. Figure 7 shows the process of the English-German translation agent MTA_2 after the Japanese-English translation agent MTA_1 completes its translation process. In the first step, the coordinator agent *CA* receives the Japanese source sentence s_1 = "*kanojo no ketten ha ookina mondai da* (Her fault is a big problem)," sets all possible n-tuples including the words in s_1 and transfers s_1 and T_1 to MTA_1. MTA_1 then translates s_1 into the English sentence t_1 = "Her fault is a big problem" using the Japanese-English translation service MT_1. MTA_1 obtains pairs P_1 of words in s_1 and t_1: P_1 = {{*ketten* (fault), fault}, {*mondai* (problem), problem}}. MTA_1 then examines the translated words. For example, if T_1 contains triples including both *ketten* (fault) and "fault," MTA_1 realizes that they share the same meaning. If that is not the case, the triples may remain incomplete, and MTA_1 has to abandon efforts to maintain context. If the triples are complete, then triples including both *ketten* (fault) and "fault" as well as those including both *mondai* (problem) and "problem" should be contained in T_1. Therefore, translated words are not modified: $Q_1 = P_1$ and $s_2 = t_1$. MTA_1 then sends s_2 and Q_1 to *CA* and *CA* generates the new context T_2. For example, both triples of T_1 including both *ketten* (fault) and "fault" are to be included in T_2, as shown in figure 7.

In the second step, s_2 and T_2 are sent to the second English-German translation agent MTA_2. MTA_2 translates s_2 to the German sentence t_2 = "*Ihre Schuld ist ein großes Problem* (Her responsibility is a big problem)." Pairs P_2 are then obtained: P_2 = {{fault, *Schuld* (responsibility)}, {problem, *Problem* (problem)}}. It appears that the word *Schuld* (responsibility) has semantically drifted, as there is no triple in T_2 that includes both "fault" and *Schuld* (responsibility). Thus it is replaced by a word that is included in a triple in T_2, which also includes "fault." If the first triple in figure 7 is selected, Q_2 would be {{fault, *Fehler* (fault)}, {problem, *Problem* (problem)}}. MTA_2 modifies t_2 to s_3: s_3 = "*Ihre Fehler ist ein großes Problem* (Her fault is a big problem)." s_3 is finally returned to the user.

5 Evaluation

We constructed Japanese-English-German triples limiting their parts-of-speech to nouns. Table 1 lists the dictionaries used and the number of triples obtained from them. Transition-type triples start with Japanese words. A total number of 21,914 triples were obtained. We first analyzed the effectiveness of the 21,914 triples in covering arbitrary Japanese documents. We used the term frequency of nouns in a Web corpus storing 470 million sentences containing 5000 million Japanese words [Kawahara and Kurohashi, 2006]. The triples

Table 1: Dictionary and generated triples

(a) Bilingual dictionaries used to obtain triples

Dictionary	Number of headwords
Genius Japanese-English dictionary	31,944 (noun)
Concise Japanese-German dictionary	38,487(all words)
Oxford English-German dictionary	31,180 (noun)
Crown German-Japanese dictionary	34,255 (noun)

(b) Number of triples of each type

Type	Number of triples
Loop	15,627
Transition (starting from Japanese)	13,757
Total (no overlaps)	21,914

Source sentence (Japanese; A):
torakku ga michi wo husaide ita.
(A truck was blocking the road.)
B: *torakku ha houhou wo samatageta.*
(A truck was blocking the method.)
C: *torakku ha michi wo samatageta.*
(A truck was blocking the road.)

Figure 8. Example of an improvement from 4 (Most) to 5 (All)

appeared to cover 58% of all nouns in the corpus and 40% of all parts-of-speech words. If the triples are used in descending order of term frequency, 6,000 triples can cover 50% of nouns and 38% of all parts-of-speech words. This implies that a relatively small number of triples can cover the majority of frequently used nouns.

We then conducted a preliminary evaluation of the quality of Japanese-German back translation using the cascade of Japanese-English, English-German, German-English, and English-Japanese translations. We compared the source Japanese sentence (A), back-translated Japanese sentence generated without context (B), and that generated based on context (C). For purposes for accuracy, we took the subjective evaluation by three Japanese subjects who were native speakers of Japanese. The subjects were asked to evaluate the translation quality on a five-point scale, how much of the original meaning of sentence A was conveyed through sentences B and C (5-All, 4-Most, 3-Much, 2-Little, 1-None). Source sentences were selected from the Machine Translation Test Set provided by the NTT Natural Language Research Group[4]. We randomly selected 100 samples in which B and C were different. The results of Welch's test show that there is a difference in quality between B and C with a confidence level greater than 98%.

On average, the translation quality improved for 41 sentences and the score increased by an average of 0.47 points using context-based coordination. For example, in figure 8, without context the Japanese word *michi* (road) is mistranslated to *houhou* (method). This error occurs because the intermediate English word "way" has several meanings. The quality improved in the case of 34% for the sentences that were previously assigned a rating of 4 when translated without coordination. Similarly, sentences with ratings of 3, 2, and 1 showed improvements for 32%, 49%, and 60% respectively with the context-based approach.

[4] http://www.kecl.ntt.co.jp/mtg/resources/index.php

6 Conclusion

This study proposes a method for context-based coordination to overcome mistranslations during pivot translation, which occurs because of inconsistent word selection. The major aspects are summarized below.

Context-based Coordination with Propagated Context

We took an approach to propagate context across combined translation services. Treating context as a set of multilingual equivalent terms used in translation, we propose to obtain all possible terms based on triangle forms formed by the relationships between words and translated words extracted from bilingual dictionaries. Our triangle method can be easily extended to four or more languages, and it is efficient in obtaining a sufficient amount of terms; the evaluation results show that the generated equivalent noun terms cover 58% of nouns and 40% of all parts-of-speech appearing in arbitrary sentences.

Multiagent Architecture for Coordination

We proposed a multiagent architecture as one way to implement coordination with propagated context, wherein the coordinator agent gathers and propagates context from/to translation agents. Evaluation results of the translation quality of the indicated improvements in 41% of the total 100 sentences used and that the quality rating increased by an average of 0.47 points on a five-point scale. This architecture offers the flexibility of extension and the possibility of constructing a more complex composition of translation services and other types of language resources.

By considering the translation services as black boxes, a substantial improvement in translation quality was realized. The advantage of our approach is that we can improve the translation quality without any corpora, training of translation services with training sentences, or changing the inner components of systems; we only use available language resources and add some components outside existing translation services. This improvement is not trivial in the intercultural collaboration domain [Ishida *et al*., 2007]. Context-based coordination approach will play an important role in the quality improvement of the component service itself making up the composite service, which is frequently considered an issue of the component technologies.

Acknowledgments

This collaborative research was conducted between NICT and Kyoto University when the author Rie Tanaka was a master's degree student at Kyoto University; it was supported by the Kyoto University Global COE Program: Informatics Education and Research Center for Knowledge-Circulating Society, Strategic Information and Communications R&D Promotion Programme from Ministry of Internal Affairs and Communications, and a Grant-in-Aid for Scientific Research (A) (21240014, 2009-2011) from the Japan Society for the Promotion of Science (JSPS).

References

[Bramantoro et al., 2008] Arif Bramantoro, Masahiro Tanaka, Yohei Murakami, Ulrich Schäfer and Toru Ishida. A Hybrid Integrated Architecture for Language Service Composition. *ICWS-08*, pages 345–352, 2008.

[Brennan and Clerk, 1996] Susan E. Brennan and Herbert H. Clark. Conceptual Pacts and Lexical Choice in Conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22(6):1482–1493, 1996.

[Hassine *et al*., 2006] Ahlem Ben Hassine, Shigeo Matsubara and Toru Ishida. A Constraint-Based Approach to Horizontal Web Service Composition. *ISWC-06*, pages 130–143, 2006.

[Ishida, 2006] Toru Ishida. Language Grid: An Infrastructure for Intercultural Collaboration. *SAINT-06*, pages 96–100, keynote address, 2006.

[Ishida *et al*., 2007] Toru Ishida, Susan R. Fussell and Piek Vossen. (Eds.): *Intercultural Collaboration*. Lecture Notes in Computer Science, 4568, Springer-Verlag, 2007.

[Kanayama and Watanabe, 2003] Hiroshi Kanayama and Hideo Watanabe. Multilingual Translation via Annotated Hub Language. *MT-Summit IX*, pages 202–207, 2003.

[Kawahara and Kurohashi, 2006] Daisuke Kawahara and Sadao Kurohashi. Case Frame Compilation from the Web using High-Performance Computing. *LREC-06*, 2006.

[Matsumura et al., 2006] Ikuo Matsumura, Toru Ishida, Yohei Murakami and Yoshiyuki Fujishiro. Situated Web Service: Context-Aware Approach to High Speed Web Service Communication. *ICWS-06*, pages 673–680, 2006.

[Tanaka and Umemura, 1994] Kumiko Tanaka and Kyoji Umemura. Construction of a Bilingual Dictionary Intermediated by a Third Language. *COLING-94*, pages 293–303, 1994.

[Tanaka et al., 2009] Masahiro Tanaka, Toru Ishida, Yohei Murakami, and Satoshi Morimoto. Service Supervision: Coordinating Web Services in Open Environment. *ICWS-09*, to be published, 2009.

[Tokunaga and Tanaka, 1990] Takenobu Tokunaga and Hozumi Tanaka. The Automatic Extraction of Conceptual Items from Bilingual Dictionaries. *PRICAI-90*, pages 304–309, 1990.

[Utiyama and Isahara, 2007] Masao Utiyama and Hitoshi Isahara. A Comparison of Pivot Methods for Phrase-based Statistical Machine Translation. *HLT-NAACL*, pages 484–491, 2007

[Vossen, 1998] Piek Vossen. (Eds.) EuroWordNet: A Multilingual Database with Lexical Semantic Networks. Dordrecht, Netherlands: Kluwer, 1998. See: http://www.hum.uva.nl/ ewn/.

[Wu and Wang, 2007] Hua Wu and Haifeng Wang. Pivot Language Approach for Phrase-Based Statistical Machine Translation. *ACL'07*, pages 856–863, 2007.

[Wu *et al*., 2008] Yanchen Wu, Fang Li, Rie Tanaka and Toru Ishida. Automatic Creation of N-lingual Synonymous Word Sets. *SKG-08*, pages 141–148, 2008.

[Yamashita *et al*., 2009] Naomi Yamashita, Rieko Inaba, Hideaki Kuzuoka and Toru Ishida. Difficulties in Establishing Common Ground in Multiparty Groups using Machine Translation. *CHI'09*, pages 679–688, 2009.

[Yamashita and Ishida, 2006] Naomi Yamashita and Toru Ishida. Effects of Machine Translation on Collaborative Work. *CSCW-06*, pages 515–523, 2006.

Constraint Optimization Approach to Context Based Word Selection

Jun Matsuno **Toru Ishida**
Department of Social Informatics,
Kyoto University, Kyoto 6068501, Japan
matsuno@ai.soc.i.kyoto-u.ac.jp ishida@i.kyoto-u.ac.jp

Abstract

Consistent word selection in machine translation is currently realized by resolving word sense ambiguity through the context of a single sentence or neighboring sentences. However, consistent word selection over the whole article has yet to be achieved. Consistency over the whole article is extremely important when applying machine translation to collectively developed documents like Wikipedia. In this paper, we propose to consider constraints between words in the whole article based on their semantic relatedness and contextual distance. The proposed method is successfully implemented in both statistical and rule-based translators. We evaluate those systems by translating 100 articles in the English Wikipedia into Japanese. The results show that the ratio of appropriate word selection for common nouns increased to around 75% with our method, while it was around 55% without our method.

1 Introduction

Activities are being conducted to improve the accessibility and usability of language services for intercultural collaboration to overcome language and cultural barriers with Language Grid [Ishida, 2006]. We are developing a multilingual environment for the translation of Wikipedia articles in cooperation with the Wikimedia Foundation. However, during this period, we have observed that output words selected by automatic machine translation systems, in both statistical machine translation (SMT) and rule-based machine translation (RBMT), are not consistent. For example, when machine translating the English Wikipedia article "George Washington" into Japanese, 18 nouns appear multiple times and are translated with different meanings. Although 5 of these nouns are context-dependent, the remaining 13 should have consistent Japanese equivalents. Inconsistency in word selection is a major problem since it prevents the user from recovering the meaning of the source text [Yamashita and Ishida, 2006; Tanaka *et al.*, 2009]. Take for example the machine translation of an English document that reads "The paper is excellent. I want to know about the author of the paper." into the Japanese "sono kami ha subarashii. watashiwa, ronbun no chosha wo siri tai. (The sheet of paper is excellent. I want to know about the author of the scientific paper.)" . The word "paper" should be translated into "ronbun (a scientific paper)" in both the first and the second sentences, but "paper" is translated into "kami (a sheet of paper)" in the first sentence. Richer contextual information is needed if we are to resolve inconsistency in word selection. In this example, the machine translation result of a single sentence was inadequate because of the failure to apply global contextual information.

Methods that improve statistical machine translation quality by using word sense disambiguation (WSD) have been proposed in the field of machine translation with contextual information [Carpuat and Wu, 2007; Chan *et al.*, 2007]. These methods, however, consider the contextual information of only neighboring sentences, and the contextual information available in the whole article is not used. Machine learning is the dominant approach in WSD, and huge features have to be treated if sentences other than neighboring sentences are used as the sources of contextual information. Moreover, it is difficult to prepare a sufficiently large training data set to give each feature an appropriate weight.

This paper proposes a word selection method based on constraint optimization. The constraint optimization problem demands that each constraint be weighted according to its degree of importance. A method that applies constraint optimization to word selection has been proposed, but it is unable to use the context of the whole article because constraint is based on single sentences [Canisius and Bosch, 2009]. As a result, consistent word selection can not be performed over the whole article. However, in the constraint optimization approach, it should be possible to use contextual information from the whole article because a variable is assigned to each word appearing in a document and word selection based on constraints between variables is performed. Thus, we propose the use of constraints between words in the whole translated article based on semantic relatedness and contextual distance between words; we resolve word sense ambiguity by using contextual information in the whole translated article. As far as we know, this study is the first to use the context of the whole article for ensuring word consistency.

2 Semantic Relatedness Between Translated Words in a Single Sentence

We formulate the word selection problem based on the weighted constraint satisfaction problem [Bistarelli *et al.*, 1997], one of the constraint optimization problems, to resolve inconsistency in word selection in the machine translation of a document. In this formulation, ambiguity in the sense of a noun in the original document is resolved by using the semantic relatedness between words in each translated sentence. That is, independent word selection is performed for each sentence by using contextual information in a single sentence. We enumerate the requirements for word selection below, and formulate the word selection problem so that it can meet those requirements.

1. The translation candidates of noun w in the original document are all translated nouns of w in the translated document
2. There is semantic relatedness between translated words in the same sentence
3. A solution is the assignment of translated words to the nouns in the original document that maximizes the sum of semantic relatedness between translated words

From requirement 1, one variable x is created for each noun w in the original document, and all translated nouns of w in the translated document are included in a domain D for each variable. From requirement 2, the constraint representing "there is semantic relatedness between translated words" is imposed between x_i and x_j if the original words of x_i and x_j co-occur in the same sentence ($1 \leq i < j \leq n$). This semantic relatedness is computed quantitatively by function SR.

We use the method of computing semantic relatedness, employed by Wikipedia [Gabrilovich and Markovitch, 2007], to compute function SR. In this method, the relative strengths between x_i and each Wikipedia article are determined by using the tf/idf score based on the number of occurrences of x_i in each article of Wikipedia in the translated language, and a translated word vector weighted for each article $v_{x_i} = (v_{x_i1}, v_{x_i2}, \ldots, v_{x_im})$ is obtained (m is the number of articles in Wikipedia in the translated language). Specifically, x_i appears $tf(i,k)$ times in the k th of the m articles, and appears in l articles. v_{x_ik} is computed as $v_{x_ik} = (1 + \log tf(i,k)) \log \frac{m}{l}$. A translated word vector v_{x_i} is obtained by performing this calculation for all articles. Semantic relatedness between translated words is expressed quantitatively by a value that is not less than 0 and not more than 1 by computing the cosine similarity between v_{x_i} and v_{x_j} which are, respectively, translated word vectors for x_i and x_j. Accordingly, $SR_{ij}(x_i, x_j)$ is determined as:

$$SR(x_i, x_j) = \frac{v_{x_i1}v_{x_j1} + \cdots + v_{x_im}v_{x_jm}}{\sqrt{v_{x_i1}^2 + \cdots + v_{x_im}^2}\sqrt{v_{x_j1}^2 + \cdots + v_{x_jm}^2}}$$

The average of the values of function SR for all pairs of variables in which the constraint is imposed is expressed as:

$$ASR(X) = \frac{\sum_{\{i,j\} \in V} SR(x_i, x_j)}{|V|}$$

(Set V consists of the pairs of indexes that correspond to the pairs of variables in which constraints are imposed.)
The larger the value of function ASR is, the larger the sum of semantic relatedness between translated words in each sentence is. Therefore, context-dependent word selection is performed for each sentence in the original document when the value of function ASR is largest. From requirement 3, the optimal solution for this problem is the tuple of translated words for the variables with maximum value of function ASR.

3 Semantic Relatedness Between Translated Words in a Document

It is thought that semantic relatedness between translated words which appear in the same sentence is really large. However, even if translated words appear in different sentences, there should be semantic relatedness between translated words according to the closeness between the contexts in which translated words appear in a document. It is expected that more accurate word selection will be realized by using the semantic relatedness between words in the translated document. We adopt this approach to formulate the word selection problem based on the weighted constraint satisfaction problem. Word selection using contextual information in the whole article is performed by solving this word selection problem. We enumerate the requirements that the word selection problem should meet below.

1. The translation candidates of noun w in the original document are all translated nouns of w in the translated document
2. There is context-dependent semantic relatedness between translated words in the same document
3. A solution is an assignment of translated words to the nouns in the original document that maximize the sum of context-dependent semantic relatedness between translated words

From requirement 1, one variable x is created for each noun w that appears in the original document, and all translated nouns of w in the translated document are included in domain D for each variable. From requirement 2, constraints representing "there is context-dependent semantic relatedness between translated words" are imposed between x_i and x_j if the original words of x_i and x_j co-occur in the same document ($1 \leq i < j \leq n$). This context-dependent semantic relatedness is computed quantitatively by function CSR which is based on function SR. Function CSR becomes important when applying machine translation to collectively developed documents like Wikipedia.

We now turn to the computational model of function CSR to compute context-dependent semantic relatedness between translated words tw and tw' whose original words are, respectively, w and w' in the same document. First, semantic relatedness SR(tw, tw') between tw, tw' is not less than 0 and not more than 1, and context-dependent semantic relatedness CSR(tw, tw') between tw, tw' does not exceed context-independent semantic relatedness SR(tw, tw'). Namely, the closer the contexts in which tw and tw' appear in a document are, the more the value of CSR approaches that of SR. In addition, we consider that the closeness of the contexts in which tw and tw' appear in the translated document is equivalent to the closeness of the contexts between the sentences in which w and w' appear in the original document. We call this contextual distance. The value of contextual distance is larger than 0, and the smaller the value is, the closer the contexts are. To express the requirements for the computational model of CSR, We describe tw and tw2 as the translations of the same two words, w, that appear in different locations of the original document, and describe tw' as the translated word of word w' in the same original document. Additionally, we describe s as a function that expresses the sentence in which the original word of the translated word appears by accepting a translated word as input, and describe DIS as a function which expresses contextual distance between these sentences upon receiving the two sentences as input. We use the following mathematical expressions to enumerate the requirements for the computational model of CSR.

1. $0 \leq SR(\text{tw,tw'}) \leq 1$
2. $0 \leq DIS(s(\text{tw}), s(\text{tw'}))$
3. $0 \leq CSR(\text{tw,tw'}) \leq SR(\text{tw,tw'})$
4. $DIS(s(\text{tw}), s(\text{tw'})) = 0$
   $\Longrightarrow CSR(\text{tw,tw'}) = SR(\text{tw,tw'})$
5. $DIS(s(\text{tw}), s(\text{tw'})) \leq DIS(s(\text{tw2}), s(\text{tw'}))$
   $\Longrightarrow CSR(\text{tw,tw'}) \geq CSR(\text{tw2,tw'})$

Our computational expression of CSR, shown in Figure 1, meets these requirements.

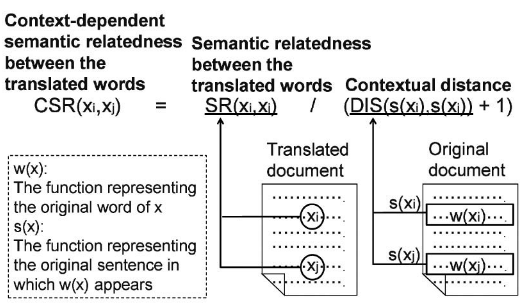


Figure 1: Computation of context-dependent semantic relatedness between translated words

We describe num as a function which expresses the order of the sentence in the article upon receiving an original sentence as input. The order of the sentence is the number of the sentence counting from the beginning of the article. Function DIS is simply based on the physical distance between original sentences as below.

$$DIS(s(x_i), s(x_j)) = num(s(x_j)) - num(s(x_i))$$

The average of the values of function CSR for all pairs of variables is expressed as below.

$$ACSR(X) = \frac{\sum_{j=i+1}^{j=n} \sum_{i=1}^{i=n} CSR(x_i, x_j)}{{}_nC_2}$$

Function $ACSR$ computes the average of the measurement of semantic relatedness between translated words in the whole translated article. The value of function $ACSR$ represents how a translated word which has a context-dependent meaning is selected for each noun in the original document. It also means that the value of function $ACSR$ represents how the same translated word that has the appropriate meaning is selected for the same nouns that have the same meaning in the original document. From requirement 3, the optimal solution for this problem is the tuple of translated words for the variables that maximize the value of function $ACSR$. Figure 2 formulates the word selection problem using semantic relatedness between translated words in a document.

Variable Set $X = \{x_1, \ldots, x_n\}$
(x_i:The translated word of the noun which appears in i th order in the original document)
Domain Set $D = \{D_1, \ldots, D_n\}$
(D_i:The set whose elements are all translated nouns of $w(x_i)$ in the translated document
$w(x)$:The function expressing the original word of translated word x)

The function expressing semantic relatedness between translated words
$SR_{ij}(x_i, x_j) = \frac{v_{x_i 1} v_{x_j 1} + \cdots + v_{x_i m} v_{x_j m}}{\sqrt{v_{x_i 1}^2 + \cdots + v_{x_i m}^2}\sqrt{v_{x_j 1}^2 + \cdots + v_{x_j m}^2}}$
($v_{x_k l}$:The weight of x_k for the l th of m articles in Wikipedia in the translated language
m:The number of articles in Wikipedia in the translated language)
The function expressing contextual distance between original sentences
$DIS(s(x_i), s(x_j)) = num(s(x_j)) - num(s(x_i))$
($s(x)$:The function expressing the sentence in which the original word of translated word x appears
$num(s(x))$:The function expressing the order of sentence $s(x)$ which appears in the document)
The function expressing context-dependent semantic relatedness between translated words
$CSR(x_i, x_j) = \frac{SR(x_i, x_j)}{DIS(s(x_i), s(x_j)) + 1}$

The function expressing how inconsistency in word selection is resolved
$ACSR(X) = \frac{\sum_{j=i+1}^{j=n} \sum_{i=1}^{i=n} CSR(x_i, x_j)}{{}_nC_2}$
Optimal Solution
The tuple of translated words for the variables with maximum $ACSR(X)$

Figure 2: Formulation of the word selection problem using semantic relatedness between translated words in a document

4 Example of the Word Selection Problem

We give an example of the word selection problem in Figure 3. Figure 4 and Figure 5 show the constraint networks yielded when this word selection problem is formulated by using the semantic relatedness between translated words in a single sentence and in a document, respectively.

Source document (English): Inuit people have their own peculiar language. However, peoples with different languages do not always have different cultures.
Translated document (Japanese): inuitto no hitobito ha karerajishin no tokuyuuna gengo wo motte imasu.
(Inuit folks have their own peculiar language.)
shikashi, kotonaru gengo wo motu minzoku ha tsuneni kotonaru bunka wo motte inai.
(However, ethnic groups with different languages do not always have different cultures.)

Figure 3: English-Japanese machine translated document in which inconsistency in word selection of “people” occurs

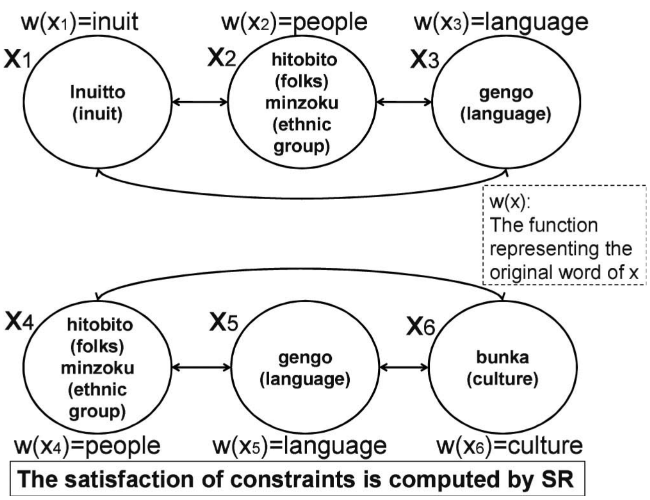


Figure 4: Constraint network representing the word selection problem of Figure 3 which is formulated using semantic relatedness between translated words in a single sentense

In Figure 4, the semantic relatedness between translated words in each sentence is computed, and word selection is independently performed for each sentence. The values of function SR for the pair of translated words are, for example, SR(“inuitto(inuit)”, “hitobito(folks)”) $= 0.0241$ and SR(“inuitto(inuit)”, “minzoku(ethnic group)”) $= 0.0524$. The value of function SR for the pair of “inuitto(inuit)” and “minzoku(ethnic group)” is more than twice that for the pair of “inuitto(inuit)” and “hitobito(folks)”.In Figure 5, context-dependent semantic relatedness between words in the translated document is computed, and word selection using contextual information in the whole document is performed. If $x_2 =$ “hitobito(folks)” and $x_4 =$ “minzoku(ethnic group)”, the values of function CSR for the pair of x_1 and x_2 and for the pair of x_1 and x_4 are calculated to be, respectively, CSR((“inuitto(inuit)”,“hitobito(folks)”) $=$ 0.0241 and CSR(“inuitto(inuit)”,“minzoku(ethnic group)”) $= 0.0262$. The original words of x_1 and x_2 appear in the same sentence,

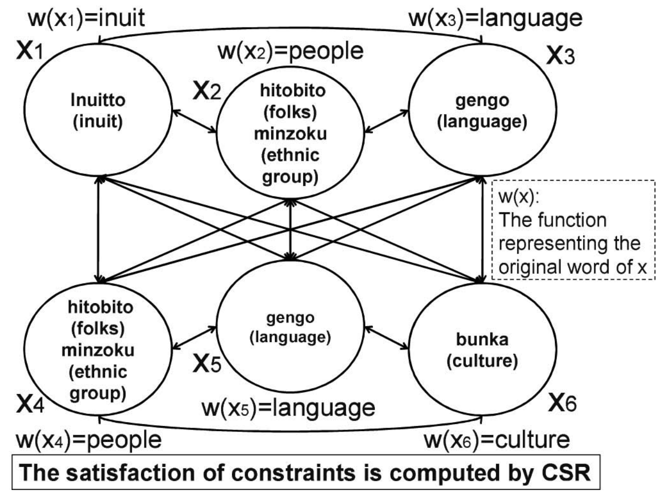


Figure 5: Constraint network representing the word selection problem of Figure 3 which is formulated using semantic relatedness between translated words in a document

but those of x_1 and x_4 appear in different sentences. Accordingly, the value of context-dependent semantic relatedness between “inuitto(inuit)” and “minzoku(ethnic group)” is not much larger than that between “inuitto(inuit)” and “hitobito(folks)”.

The translated word that should be selected for $w(x_2)$ and $w(x_4)$ is “minzoku(ethnic group)”. Although “minzoku(ethnic group)” and “hitobito(folks)” are selected for $w(x_2)$ and $w(x_4)$, respectively, in the word selection problem represented by the constraint network of Figure 4, “minzoku(ethnic group)” is selected for both $w(x_2)$ and $w(x_4)$ in the word selection problem represented by the constraint network of Figure 5. This is because the semantic relatedness between the translated word of $w(x_4)$ and “inuitto(inuit)” which has strong semantic relatedness with “minzoku(ethnic group)”, which is the appropriate translated word for $w(x_4)$, is used in the word selection problem represented by the constraint network of Figure 5.

5 Evaluation

5.1 Evaluation Settings

We implemented the systems of WSD/SR(sentence) and WSD/CSR(article) to formulate the word selection problem using semantic relatedness between translated words in a single sentence and a document, respectively, and resolved the word selection problem by applying the hill climbing approach. Furthermore, we implemented WSD/SR(article). WSD/SR(article) is different from WSD/CSR(article) in that function SR is used instead of CSR to compute the semantic relatedness between translated words. By comparing the evaluation results of WSD/SR(article) and WSD/CSR(article), we can better understand the effectiveness of using function CSR which becomes important when applying machine translation to collectively developed documents like Wikipedia. We used Google Translate[1] and J-Server[2] as examples of SMT and RBMT systems, and used 100 samples which were randomly selected from English Wikipedia articles whose bodies contained more than 500 words as the source documents.

5.2 Evaluation Results

Table 1 shows (a) "the total number of appearances of all common nouns" when translating the 100 samples by Google Translate and J-Server. The common nouns that were included in (a) had different meanings for the translated words selected by machine translation in each document. Table 2 and Table 3 show the number of nouns that were appropriately translated (a) when Google Translate and J-Server were used, respectively.

Table 1: Number of common nouns evaluated

	Google Translate	J-Server
(a)	427	369

(a)"the total number of appearances of all common nouns" (These common nouns had different meanings for the translated words selected by machine translation in each document)

The followings are shown from the evaluation results.

- Both Google Translate and J-Server performed appropriate word selection at the rate of about 55%.
- WSD/SR(sentence) improved word selection quality by 10 points by using contextual information in single sentences. However, the translations still had a word selection rate of about 35%.
- WSD/SR(article) selected the same translated word for the same nouns in the same document by computing semantic relatedness rather than contextual distance although WSD/SR(sentence) selected translated words independently in each sentence. Therefore, WSD/SR(article) consistently selected inappropriate translated words for nouns for which the same translated word should have been selected, and WSD/SR(article) decreased word selection quality more than WSD/SR(sentence) in some cases.
- WSD/CSR(article) yielded better word selection quality than WSD/SR(article) because it uses richer contextual distance to compute semantic relatedness. As a result, WSD/CSR(article) was the best system in terms of word selection quality.

However, we regarded the translation candidates of a word as all translated words which the machine translation system selected for the word in the same document. Therefore, WSD/CSR(article) sometimes failed to select appropriate translated words because appropriate translated words were not included in their translation candidates. Extracting translation candidates from bilingual dictionaries may improve word selection quality.

[1] http://translate.google.co.jp/

[2] http://www3.j-server.com/KODENSHA/contents/entrial/index.htm

Table 2: Comparative evaluation of word selection quality for Google Translate

System	The number of nouns that were appropriately translated
Google Translate	245(57.4%)
+ WSD/SR(sentence)	274(64.2%)
+ WSD/SR(article)	306(71.7%)
+ WSD/CSR(article)	313(73.3%)

Table 3: Comparative evaluation of word selection quality for J-Server

System	The number of nouns that were appropriately translated
J-Server	200(53.9%)
+ WSD/SR(sentence)	241(65.0%)
+ WSD/SR(article)	240(64.5%)
+ WSD/CSR(article)	271(72.9%)

6 Related Work

Existing WSD studies attempt to identify the correct meaning of a polysemous word by using context. Carpuat and Wu [2005] proposed a method that uses words selected by WSD to replace words in a machine translated sentence. They verified whether WSD could improve the translation quality of statistical machine translation (SMT) in the translation of a single sentence or not. The evaluation results using BLEU metric, which is an automatic evaluation method, showed that using WSD decreased the translation quality of SMT. This was because the word replacement degraded the fluency of the sentence. Our method also replaces translated words so we need to manually evaluate the translation quality of the resulting sentences.

In [Carpuat and Wu, 2005], it was shown that the direct use of WSD for SMT could not improve translation quality. Methods that improve the translation quality of SMT by coordinating a WSD model and statistical models of SMT have been proposed [Carpuat and Wu, 2007; Chan *et al.*, 2007]. However, in [Carpuat and Wu, 2007], contextual information from only the original sentence was used for WSD. In [Chan *et al.*, 2007], contextual information in multiple sentences was used for WSD, but sentences that were used as contextual information were limited to the original sentence and the immediately adjoining sentences. This is because a WSD method based on machine learning, such as a support vector machine, needs an impractically large training data set if sentences other than an original sentence and its neighboring sentences are used for WSD. In these methods, consistent word selection is not performed over the whole article because contextual information from the whole article is not used.

SMT methods select translation rules based on context by using the wealth of contextual information available in translation rules and syntax trees have been recently proposed [He *et al.*, 2008; Liu *et al.*, 2008; Shen *et al.*, 2009]D However, using contextual information obtained in the production process of sentences demands the existence of a large training data set. Moreover, these methods select translation rules based on context, while our method uses context to resolve word sense ambiguity.

Our method performs word selection based on the weighted constraint satisfaction problem. Canisius and Bosch [2009] proposed a method that improves the translation quality of SMT based on the weighted constraint satisfaction problem. In this method, constraints on the connections between translated words are initially obtained from a corpus. The line of translated words that maximizes the translation score while satisfying the constraints is produced as the translation output sentence. Therefore, imposing constraints between words in a translated sentence enables the use of contextual information in a translated sentence. In our method, constraints indicating that there is semantic relatedness between words are imposed between words throughout the whole translated article. In addition, constraints are weighted by the degree of importance of the contextual information according to semantic relatedness and contextual distance between words. This realizes word selection based on contextual information from the whole translated article.

7 Conclusion

Inconsistency in word selection is a problem that occurs when the instances of one source word are given different translations. Consistent word selection can be realized for the translation of documents like Wikipedia by resolving this problem. Contextual information taken from the whole article must be used to resolve this problem. We proposed a word selection method based on constraint optimization. Our method can suppress inconsistency in word selection by using contextual information from the whole article, not just single sentences.

Evaluations on Wikipedia articles showed that our method was effective for both statistical and rule-based translators. The ratio of appropriate word selection for common nouns was around 55% with previous approaches. However, it was around 75% with our method. Using contextual information from the whole document improves the word selection quality of machine translations. We will evaluate the translation quality in terms of fluency to highlight the benefits of our method.

Acknowledgments

This research was supported by Strategic Information and Communications R&D Promotion Programme (SCOPE) from Ministry of Internal Affairs and Communications of Japan and a Grant-in-Aid for Scientific Research (A) (21240014, 2009-2011) from Japan Society for the Promotion of Science (JSPS).

References

[Bistarelli *et al.*, 1997] Stefano Bistarelli, Ugo Montanari and Francesca Rossi. Semiring-Based Constraint Satisfaction and Optimization. *Journal of the Association of Computing Machinery(JACM), vol.44 no.2*, pages 201-236, 1997.

[Canisius and Bosch, 2009] Sander Canisius and Antal van den Bosch. A Constraint Satisfaction Approach to Machine Translation. *In Proceedings of the 13th Annual Conference of the European Association for Machine Translation(EAMT-09)*, pages 182-189, 2009.

[Carpuat and Wu, 2005] Marine Carpuat and Dekai Wu. Word Sense Disambiguation vs. Statistical Machine Translation. *In Proceedings of the 43th Annual Meeting of the Association of Computational Linguistics(ACL-05)*, pages 387-394, 2005.

[Carpuat and Wu, 2007] Marine Carpuat and Dekai Wu. Context-Dependent Phrasal Translation Lexicons for Statistical Machine Translation. *In Proceedings of Machine Translation Summit XI*, pages 73-80, 2007.

[Chan *et al.*, 2007] Yee Seng Chan, Hwee Tou Ng and David Chiang. Word Sense Disambiguation Improves Statistical Machine Translation. *In Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics(ACL-07)*, pages 33-40, 2007.

[Gabrilovich and Markovitch, 2007] Evgeniy Gabrilovich and Shaul Markovitch. Computing semantic relatedness using Wikipedia-based explicit semantic analysis. *In Proceedings of the 20th International Joint Conference on Artificial Intelligence(IJCAI-07)*, pages 1606-1611, 2007.

[He *et al.*, 2008] Zhongjun He, Qun Liu and Shouxun Lin. Improving Statistical Machine Translation using Lexicalized Rule Selection. *In Proceedings of the 22nd International Conference on Computational Linguistics(COLING-08)*, pages 321-328, 2008.

[Ishida, 2006] Toru Ishida. Language Grid: An Infrastructure for Intercultural Collaboration. *IEEE/IPSJ Symposium on Applications and the Internet(SAINT-06)*, pages 96-100, 2006.

[Liu *et al.*, 2008] Qun Liu, Zhongjun He, Yang Liu and Shouxun Lin. Maximum Entropy based Rule Selection Model for Syntax-based Statistical Machine Translation. *In Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing(EMNLP-08)*, pages 89-97, 2008.

[Shen *et al.*, 2009] Libin Shen, Jinxi Xu, Bing Zhang, Spyros Matsoukas and Ralph Weischedel. Effective Use of Linguistic and Contextual Information for Statistical Machine Translation. *In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing(EMNLP-09)*, pages 72-80, 2009.

[Tanaka *et al.*, 2009] Rie Tanaka, Yohei Murakami and Toru Ishida. Context-Based Approach for Pivot Translation Services. *In Proceedings of the 21st International Joint Conference on Artificial Intelligence(IJCAI-09)*, pages 1555-1561, 2009.

[Yamashita and Ishida, 2006] Naomi Yamashita and Toru Ishida. Effects of Machine Translation on Collaborative Work. *In Proceedings of International Conference on Computer Supported Cooperative Work(CSCW-06)*, pages 515-523, 2006.

2009 Fifth International Conference on Semantics, Knowledge and Grid

User-Centered QoS in Combining Web Services for Interactive Domain

Arif Bramantoro[1], Toru Ishida[2]

Department of Social Informatics, Kyoto University
Yoshida-honmachi, Kyoto, Japan
[1]arif@ai.soc.i.kyoto-u.ac.jp
[2]ishida@i.kyoto-u.ac.jp

Abstract **— The success of the emerging service oriented computing relies fully on the Quality of Service (QoS). However, existing QoS techniques do not accommodate users' skills and preferences. We propose user-centered QoS, which is a QoS defined by the interaction between skills/preferences of service user(s) and quality of service provider(s). By implementing user-centered QoS approach, the best service is delivered to users based on the calculation not only the quality of the services but also the skill/information of users. We proposed a novel two-stage approach for combining services in user-centered QoS, i.e. intra-workflow and inter-workflow service selection. Intra-workflow service selection is used to calculate the most optimal QoS value for each composite service. Inter-workflow service selection is used to search for the most optimal combination of composite services by utilizing the QoS values obtained from intra-workflow service selection. In this paper, we provide a concrete example of user-centered QoS in the language services domain. This problem arises when there are multi users with different quality of English using multilingual chat service.**

I. Introduction

We are already in mature era of service oriented computing, with a rapid progress into the complete philosophy or paradigm rather than merely technology. The visionary promise of delivering dynamic creation of loosely coupled information system is almost into reality. Both industrial and research efforts within the vision of service oriented computing are vastly spanning various disciplines, including Quality of Service (QoS).

Having QoS in any concepts and technologies of web service is inevitable; in fact the QoS is implicitly available in all applications and just need to be exploited. However, current QoS researches are not aligned with the definition of *service* in service oriented computing. Researchers tend to define QoS as a one-way concept from service provider to service user. QoS should be based on the interaction between service users and service providers as Zhang et al. define the concept of *service* in their book [22]. Based on this *service* definition, we propose a new concept of user-centered QoS in service oriented computing to emphasize the need of accommodating the interaction between service users and providers. We define user-centered QoS as a QoS that involves users more in the QoS calculation and control based on the interaction between users and the services or the providers of the services, not just one-way definition from providers.

The current QoS researches [3,6] in service oriented computing only take the concept of QoS from network domain for granted. QoS is actually not only about underlying network, but also the capability of service provider and at the same time the user skills or preferences. The current techniques of QoS based web service selection [10], [11] only accommodate few information regarding user's skills or preferences. For example, Chaari *et al*. in [23] provides consumer's requirements, however, the requirements are only limited to the given metrics (reliability, response time, etc) that the consumers do not have other options. The same problem exists in another paper in [24] that provides a capture to user preference (even for dynamically changing preference), but it lacks a flexibility to define new metrics based on the user's needs. Failure in satisfying these requirements will deliver to the user disappointment in using web service.

To address the importance of user-centered QoS in service oriented computing, we need a concrete and complete example of QoS problem. Recently, we faced a fundamental QoS related problem in a real application. This problem arose, when we used multilingual chat service that combines translation services and morphological analyzer services in different languages [14]. We found an interesting situation in this composite service. It started when there were initially two users using the service, Japanese and Chinese users. Japanese user was good in English, but Chinese user had no English capability. The chat service thus provided Japanese-Chinese translation service.

After a while, another user came and wanted to join the conversation. This user was from Indonesia who could speak English considerably enough. Since the Indonesian-English translation was available, the chat service was composed by multi-hop translation service for Japanese-English-Indonesian and Chinese-English-Indonesian. However, the QoS of these multi-hop translation services was not good enough [15]. The translation results were terrible. All users got disappointed of this irritating communication. This irritating problem can be avoided if the user-centered QoS aware service selection is available. The service selection should consider the QoS related multiuser condition and manage this information for QoS calculation together with QoS information from provider. Based on this new QoS calculation, the best combination of services can be selected and delivered to users.

Motivated by the aforementioned problem, we propose a new framework that introduces two-stage approach of service

978-0-7695-3810-5/09 $26.00 © 2009 IEEE
DOI 10.1109/SKG.2009.106

selection for user-centered QoS, i.e. intra-workflow and inter-workflow service selection. We use intra-workflow service selection to calculate the most optimal QoS value for each composite service and inter-workflow service selection to search for the most optimal combination of composite services by utilizing QoS values obtained from intra-workflow service selection. We argue that one-stage service selection is not enough to solve the problem of user-centered QoS, especially in multiuser environment.

The aim of this paper is to optimize a concrete problem of user-centered QoS by using a robust technique and a reliable architecture, even if the environment dynamically changes. We realize that there have been some breakthroughs of QoS researches in service oriented computing. However, we argue that none of these researches can solve the fundamental problems that we found in language services and most likely in other services. Hence, our contributions are as follows: (a) we give a new concept of user-centered QoS in service oriented computing; (b) we present a novel approach of two-stage service selection, i.e. intra-workflow and inter-workflow service selection, in user-centered QoS; (c) we provide a concrete example of user-centered QoS problem to show the importance of accommodating an interaction between users' skill/preference and the service being used.

The rest of this paper is organized as follows. Section 2 presents our concept of user-centered QoS in service oriented computing. Section 3 describes the approach of intra-workflow web service selection for user-centered QoS, while inter-workflow service selection is in Section 4. A complete description of user-centered QoS problem is described in Section 5. Section 6 shows the architecture of user-centered QoS. Finally, we summarize and conclude the paper in Section 7.

II. USER-CENTERED QOS IN SERVICE ORIENTED COMPUTING

We define user-centered QoS as a different approach of QoS that emphasizes the interaction between service users and service providers. This definition is aligned with the definition of *service* for service oriented computing written in Zhang *et al.*'s book [22] as follows:

> "*Services* represent a type of relationships-based *interactions* (activities) between at least one service provider and one service consumer to achieve a certain business goal or solution objective."

We argue that it is essential to adopt the concept of *interaction* from the definition of service in service oriented computing to the concept of QoS. Although original concept of QoS is from network domain, it is necessary to have distinct concept of QoS in service oriented computing.

In user-centered QoS, the interaction between service users and service providers has several key factors that influence the overall quality. We propose user preferences or skills that can be used as key factors in the interaction. In user-centered QoS framework, any users can give a preference of the service that they want to use or let their skills included in combining web services. This framework provides high flexibility for users to choose what QoS requirements of the services that they prefer to use or appropriate to their skills. Therefore, users will get what they want. For example, user skill of bidding (a combination between trust score, number of sold and bought product) should be considered as a key factor in deciding the best services of internet auction delivered to user.

Another example is a commonly used scenario in many service oriented computing examples, i.e. travel planner services. Suppose there are multi-national passengers who want to travel together. There is a user preference that related to these passengers, which is hospitality. For the users from Asia might consider the hospitality from the flight attendance is importance whereas their other colleges who from Europe and America do not consider this issue. So, there is a different level for hospitality between these users of the same travel service that we have to deal with.

The last example that we use to show that user-centered QoS is a real problem is language service, which exists in both single-user and multiuser environment. In single-user environment, there is a Japanese user who wants to use dictionary service. Since there are two dictionary services available, i.e. English-to-English dictionary service and English-to-Japanese dictionary service, the service selection should consider the QoS related condition of the user, i.e. mother tongue and English capability that can be indicated from language certificate. In multiuser environment, mother tongue and English certificate should be included also in combining different translation services for each user. The example of multiuser language service problem is already explained in introduction section. Due to the limited space of this paper, we use a multiuser based language services as a running example throughout this paper.

In addition to the previously mentioned research problem of QoS, it becomes a common sense amongst researchers in service oriented computing that QoS metrics is related to network domain and, therefore, they adopt the entire network metrics into service oriented computing, such as response time, reliability, availability, and so on. There are only few researches, to our knowledge, that propose a new metric related to particular domain and accommodate user requirements [13], [19]. However, these researches lack a real example in service oriented application and an integrated solution to calculate the metrics. This will cause inability to show the importance of accommodating users in QoS control. A special attention is given to the previous work [25] that provides a flexible framework to change QoS metrics based on user preference. However, this paper still uses network-domain QoS metrics or other QoS metrics, such as price, that is not related to network but is actually used by application.

To solve the problem of user-centered QoS, we need a robust technique and a flexible specification for user-centered QoS. We choose to use and extend constraint optimization technique [20], a well known AI technique to solve many sophisticated problems, such as scheduling, temporal reasoning, resource allocation, etc. Accordingly, the problem of web service selection can be modeled and solved by using constraint optimization technique. Previously, Ben Hassine *et al.* in [7] has formulized Web service composition problem based on a constraint optimization problem (COP), while Channa *et al.* in [8] has proposed the use of constraint satisfaction problem (CSP) in dynamic web service composition. However, these two papers did not include QoS management constraints and even can solve the user-centered QoS problem that we found.

Original constraint optimization problem is characterized with a triplet entities (*X, D, C*) plus objective function. *X* is a finite set of variables associated with finite domains *D* as a list of possible values for each variable, whereas *C* is a set of constraints. In our approach, it is possible to define conditional constraints [2] to accommodate the resource allocation, especially when there is a resource dependent to other resources. Lastly, the objective function is optimized to find a complete assignment of values to all variables and at the same time satisfying the constraints.

In the web service selection point of view, we extend the triplet of constraint optimization problem into quadruplet. A new variable, *P*, is created to accommodate user profile that defines user skills or preferences. As an example, *P* in the language service can be mother tongue and foreign language certification score. Hence, the extended constraint optimization formulization is as follows:

- $X=\{X_1,\ldots,X_n\}$ is a set of abstract web services, with $X_i.IN$ is a set of required input types, $X_i.OUT$ is a set of required output types, $X_i.QOS$ is a set of required QoS types. These requirements are defined as abstract service specifications.
- $D=\{D_1,\ldots,D_n\}$ where D_i a set of concrete web services X_i that can perform the task of the corresponding abstract web services.
  $D_i=\{s_{i1},\ldots,s_{ik}\}$ where s_{ij} is a concrete web service of the corresponding X_i with $s_{ij}.IN$ is a set of provided input types, $s_{ij}.OUT$ is a set of provided output types, $s_{ij}.QOS$ is a set of provided QoS types. In semantic matching of web service selection [4], every element of the input set in concrete service specification should be also an element of the input set in abstract service specification and every element of the output set in abstract service specification should be also an element of the output set in concrete service specification. We argue that in QoS based matching every element of the QoS set in abstract service specification should be also an element of the output set in concrete service specification. Therefore, we define semantically matched service specification as follows.
  - $D_i=\{s_{ij} \mid s_{ij}.IN \subseteq X_i.IN \wedge X_i.OUT \subseteq s_{ij}.OUT \wedge X_i.QOS \subseteq s_{ij}.QOS\}$
  - $P=\{P_1,\ldots,P_m\}$ is a set of user profile obtained from each user. P_i consists of profile values of user *i*.
  - $C=\{C_1,\ldots,C_p\}$ is a set of constraints which contains *CS* as a set of soft constraints with a penalty of $\rho C_i \in [0, 1]$, and *CH* as a set of hard constraints.
  - *f(R)* is the objective function to be maximized. The goal is to find the best assignment *R* for the variables in *X* while satisfying all the hard constraints. *R* is the resulted solution of a problem assigned by the instantiation of all variables of the problems. In the web service selection, we define the objective function *f(R)* by using penalty over soft constraints *ψ(R)* and QoS function *QoS(R)* as shown in Eq. 1.

$$f(R)=QoS(R)-\psi(R) \quad \textbf{(1)}$$

To solve web service selection problem, we have to find the best assignment of the variable *R** such that, all the hard constraints are satisfied while maximizing the following function in Eq. 2.

$$R^*=arg\ \max_{R\in Solution} f(R) \quad \textbf{(2)}$$

The penalty over soft constraints can be calculated by summing the penalties associated to all soft constraints as described in Eq. 3.

$$\psi(R)=\sum_{C_k\in CS} \rho C_k \quad \textbf{(3)}$$

The QoS functions consists of commonly used QoS metrics, such as price, reputation, reliability, availability; and other newly defined QoS metrics from users. The detail QoS function is described in the Eq. 4 where *Q(R)* is a QoS function obtained from existing known aggregation and/or newly defined function for customized QoS metrics and *m* is the number of QoS metrics.

$$QoS(R)=Q_1(R)+Q_2(R)+\ldots+Q_m(R) \quad \textbf{(4)}$$

To calculate each QoS function, we refer to the two papers [5], [13] that provide the aggregation functions of most QoS metrics in network domain, such as time, price, availability, reliability, reputation and success rate. Zeng *et al*. in [5] gives a foundation for QoS aggregation function. Canfora *et al.* [13], on the other hand, provides specific aggregation functions for each workflow constructs and additionally domain-dependent attribute. Our approach handles user-specified attribute differently to what proposed in [13]. We argue that QoS aggregation function for user-specified attribute should be defined freely by users (or third parties, such as service brokers) based on particular domain.

III. Intra-Workflow Service Selection

In this section, we give a detail explanation of intra-workflow service selection whereas inter-workflow service selection will be explained in the next section. As introduced partly in the first section, we provide a concrete problem of user-centered QoS in the multiuser environment. Our approach in solving user-centered QoS problem in multiuser environment is based on the two-stage service selection, i.e.: intra-workflow and inter-workflow service selection. Intra-workflow service selection is used to calculate the most optimal QoS value for each composite service. Inter-workflow service selection is used to search for the most optimal combination of composite services by utilizing QoS values obtained from intra-workflow service selection. To see the relation between these two service selections, we provide an interaction model as described in Fig. 1.

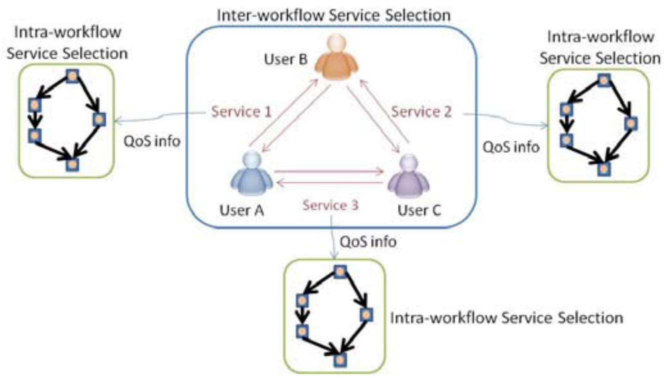


Fig. 1. Interaction model between inter-workflow and intra-workflow service selection

It is clearly seen from Fig. 1 that each service in inter-workflow service selection has QoS value resulted from intra-workflow service selection. In a real world, the service used by each user might be in the form of composite service. In case it is composite service, we need to calculate QoS based on service workflow. The calculation of QoS in each workflow is performed in intra-workflow service selection. Since there are some possible services for each users, QoS of each possible service should be calculated separately in intra-workflow service selection. In intra-workflow service selection, QoS calculation for each workflow is based on the most optimal solution of concrete services. In other words, intra-workflow service selection calculates the total QoS value of all concrete services composed in one workflow.

As an example of intra-workflow service selection, let us take a part of user-centered QoS problem in the language services. In the language service, we can compose a translation service with the community dictionary service to increase the quality of translation [1]. One of the workflow for possible concrete composite service between Japanese user and Indonesian user is ja-id translation service as described in Fig. 2. The detail calculation of QoS based on objective function will be explained in Section 5.

The formulization for this workflow is as follows:

- $X=\{X_1, X_2, X_3, X_4, X_5\}$, where:
  - X_1: Morphological analyzer service;
  - X_2: ja-en translation service;
  - X_3: en-id translation service;
  - X_4: Community dictionary service;
  - X_5: Term replacement service;
- $D=\{D_1, D_2, D_3, D_4, D_5\}$, where
  - D_1: {mecab at NTT, ICTCLAS, KLT at Kookmin University, treetagger at IMS Stuttgart};
  - D_2: {JServer at Kyoto-U, JServer at NICT, WEB-Transer at Kyoto-U, WEB-Transer at NICT};
  - D_3 : {ToggleText at Kyoto-U, ToggleText at NICT};
  - D_4: {Life Science Dictionary, Natural Disasters Dictionary, Kyoto Tourism Dictionary at NICT, Academic Terms Dictionary at NII};
  - D_5: {TermRepl service};

  (For *the* sake of simplicity, we omit the input and output parameters of D_i)
- $C=CS \cup CH$, in this intra-workflow service selection, however, we only employ hard constraints so that the objective function focuses on calculating the aggregated QoS values, where:
  - CH including (due to page limitation, only example constraints are shown)
    - C_1: For multi hop translation, $X_2.OUT=X_3.IN$;
    - C_2: For composite service which involves X_2 and X_4 (translation service and multilingual dictionary), serverLocation(X_2)=serverLocation(X_4);
    - C_3: For morphological analysis used together with community dictionary services, partialAnalyzedResult($X_1.OUT$) $\in X_4.IN$.

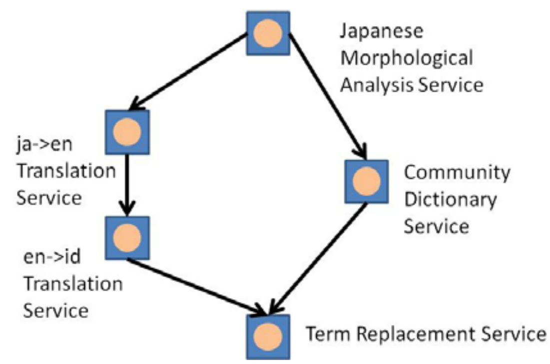


Fig. 2. A workflow of Japanese-Indonesian translation service

IV. INTER-WORKFLOW SERVICE SELECTION

In inter-workflow service selection, there is a combination of services between users in multiuser environment. One user can have different service from the service used by other users. This combination is not necessarily related to the control of workflow, such as sequence, split, choice and loop. The relation of services used by each user is more likely in the form of constraints. The main task of inter-workflow service selection is to find the best combination of services that meet the QoS constraints based on QoS related condition of users and the quality of the service itself.

To solve our formulization of user-centered constraint optimization problem for QoS, we use a simple search algorithm for constraint optimization problem. Our algorithm is based on the basic search algorithm for constraint optimization, branch-and-bound algorithm [20]. The aim of using this algorithm is to find the best solution by extending backtracking search to traverse the search space seeking all solutions. It maintains the value of objective function so far, which is so called a lower bound. In addition, for each partial solution, the algorithm also computes an upper bound using a bounding evaluation function, which overestimates the best-solution in objective function that can extend the partial solution. Therefore, when the upper bound of the partial solution is less than the lower bound, the partial solution can be aborted, and the algorithm backtracks, pruning the subtree below the partial solution. The algorithm returns to the previous partial solution and attempts to find a new assignment to *X*.

We have to slightly modify this algorithm to incorporate user-centered QoS in constraint optimization. The modification is related to the checking whether the QoS information of current domain's workflow is already calculated or not. If the QoS information is not yet calculated in intra-workflow service selection, then the algorithm will call intra-workflow function to calculate the QoS of the current domain. The intra-workflow function is similar to the search algorithm for inter-workflow service selection. The difference is that the intra-workflow function delivers the optimized QoS information of particular domain, not the optimized solution.

As any other search algorithms in constraint optimization technique [9], our algorithm produces the complexity of NP-Hard. Here, we argue that the function of intra-workflow is rarely executed. This is due to that the workflow does not easily change over the time and a new service is not added frequently. Furthermore, in our architecture this function can be executed in offline processing. Therefore, the number of constraints and services is fixed and we can maintain the complexity of this algorithm in polynomial time not NP-Hard anymore, unless for a worse case when the workflow changes or there is a new service added in the set of concrete web services frequently.

As an example of inter-workflow service selection, let us take a part of user-centered QoS problem in the language services. The problem of multilingual chat service can be formulized as follows (the detail service selection with objective function will be explained in Section 5):

- $X=\{X_1, X_2, X_3, X_4, X_5, X_6\}$, where
  - X_1: service from Japanese user to Chinese user;
  - X_2: service from Chinese user to Japanese user;
  - X_3: service from Japanese user to Indonesian user;
  - X_4: service from Indonesian user to Japanese user;
  - X_5: service from Chinese user to Indonesian user;
  - X_6: service from Indonesian user to Chinese user;
- $D=\{D_1, D_2, D_3, D_4, D_5, D_6\}$, where
  - D_1: {ja-en, ja-zh, en-zh, no translation service};
  - D_2: {zh-en, zh-ja, en-ja, no translation service};
  - D_3: {ja-en, ja-id, en-id, no translation service};
  - D_4: {id-en, id-ja, en-ja, no translation service};
  - D_5: {zh-en, zh-id, en-id, no translation service};
  - D_6: {id-en, id-zh, en-zh, no translation service};
- $P=\{P_1, P_2, P_3\}$, where
  - P_1 is a user profile of Japanese user.
    P_1.mother_tongue=Japanese, P_1.english_writing_skill=0.8, P_1.english_reading_skill=0.9;
  - P_2 is a user profile of Chinese user.
    P_2.mother_tongue=Chinese, P_2.english_writing_skill=0.1, P_2.english_reading_skill=0.2;
  - P_3 is a user profile of Indonesian user.
    P_3.mother_tongue=Indonesian, P_3.english_writing_skill=0.6, P_3.english_reading_skill=0.6;
- $C=CH \cup CS$ (we will present the soft constraints CS in Section 5), where
  - Hard constraints CH, where each user should type in one language (although it is possible to type more than one languages in chat services, we assume that the user preference of one language is a hard constraint), including
    - C_1: X_1=ja-en => (X_3=ja-en $\vee$ X_3=ja-id);
    - C_2: X_1=ja-zh => (X_3=ja-en $\vee$ X_3=ja-id);
    - C_3: X_1=en-zh => X_3=en-id;
    - C_4: X_2=zh-en => (X_5=zh-en $\vee$ X_5=zh-id);
    - C_5: X_2=zh-ja => (X_5=zh-en $\vee$ X_5=zh-id);
    - C_6: X_2=en-ja => X_5=en-id;
    - C_7: X_4=id-en => (X_6=id-en $\vee$ X_6=id-zh);
    - C_8: X_4=id-ja => (X_6=id-en $\vee$ X_6=id-zh);
    - C_9: X_4=en-ja => X_6=en-id;
      (For simplicity, we omit the other way around of the constraints C_{10} to C_{18})
    - C_{19}: X_1=no_translation => (X_3=no_translation $\vee$ X_3=en-id);
    - C_{20}: X_2=no_translation => (X_5=no_translation $\vee$ X_5=en-id);
    - C_{21}: X_4=no_translation => (X_6=no_translation $\vee$ X_6=en-zh).
      (For simplicity, we omit the other way around of the constraints C_{22} to C_{24})

The complete set of the hard constraints from C_1 until C_{24} is described in Fig. 3.

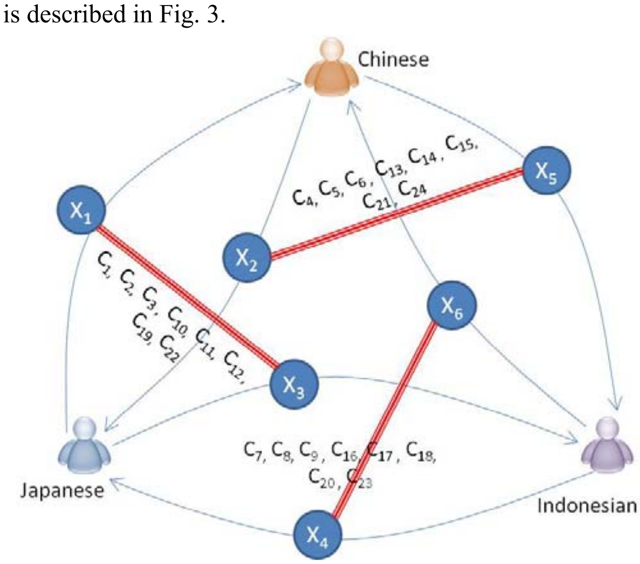


Fig. 3. Simplified constraint graph for hard constraint examples in intra-workflow service selection

V. USER-CENTERED QOS IN MULTIUSER ENVIRONMENT

In this section, we present a real scenario that shows the problem of user-centered QoS in detail. This scenario involves a complete set of web services and frequently used by real users, i.e. the Language Grid [16]. The Language Grid is a service oriented collective intelligent platform to collect and share language services. Delivering QoS on the Language Grid is challenging because there are many applications with different characteristics and requirements compete for all language resources [17].

QoS metric applicable to language service is accuracy which consists of the combination between fluency and adequacy [15]. Fluency refers to well-formed grammar, contains correct spellings, adheres to common use of terms, titles and names, is intuitively acceptable and can be sensibly interpreted by a native speaker. Adequacy refers to the degree to which information present in the original is also communicated in the translation.

In the case of multilingual chat service, user-centered QoS approach is needed to calculate the information of user's ability in language and the accuracy of translation. When initially there were two users, a Japanese user with good English and a Chinese user with no English, the composition should automatically select the translation service from Chinese to Japanese and vice versa. After an Indonesian user who can speak English a little joined the conversation, the composition should recalculate QoS information from each translation service and compare it with each user's language capability. In this case, the chat service should include Chinese-English translation for communicating Chinese and Indonesian users; but no translation service (English only communication service) for Indonesian and Japanese users. This is due to the poor quality of Japanese-Indonesian and Chinese-Indonesian translation services, which use multi-hop translation services with English as a pivot language [14]. We provide Fig. 4 to clearly understand this problem.

In intra-workflow service selection, the objective function is used to retrieve the optimized QoS value of each workflow. Hence, the aim of this objective function is not to find the best solution but rather than to retrieve the QoS value of composite services that can be used by inter-workflow service selection. We use the same objective function in Eq. 1 modified to compromise with the characteristic of language service's quality of service. The cascaded translation service represented with sequential workflow reduces the overall quality. The multi-hop translation service represented by two translation services in sequence workflow gives the most significant influence to the overall quality and therefore should be given the biggest weight amongst others, i.e. 0.6 for ja-en and en-id translation services. However, we use multiplication for these services since the quality becomes much decreasing if we combine two translation services as in the following Eq. 5.

$$f(R) = 0.2 \times s_{1j}.accuracy + 0.6 \times s_{2j}.accuracy \times s_{3j}.accuracy + 0.1 \times s_{4j}.accuracy + 0.1 \times s_{5j}.accuracy \quad \textbf{(5)}$$

We assume that the accuracy value from each language service in this implementation is available from language evaluation system utilizing human evaluation system or automatic one such as BLEU [12]. As a result of intra-workflow service selection, the most optimal QoS accuracy value for ja-id translation service is delivered by the combination of {mecab at NTT, WEB-Transer at NICT, ToggleText at NICT, Kyoto Tourism Dictionary at NICT, TermRepl service}.

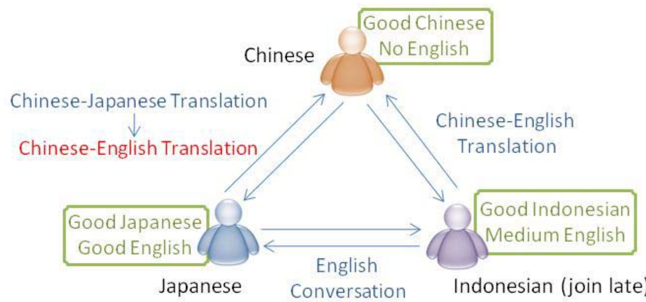


Fig. 4. Multilingual chat service problem

Inter-workflow service selection can use the resulted QoS values obtained from intra-workflow service selection. We introduce a new function to estimate the quality of message (QoM) that calculates each possible abstract translation service between two users (represented by users' profile). In this case, we consider mother tongue of user, English writing skill and reading skill as user profile. We define (QoM) function sent by one user represented by user profile P_i and received by another user represented by user profile P_j that uses translation service X_n in Eq. 6.

$$QoM(P_i, X_k, P_j) = Accuracy(P_i.writing\_skill(X_k.input\_language)) \times X_k.accuracy \times Accuracy(P_j.reading\_skill(X_k.output\_language)) \quad \textbf{(6)}$$

In inter-workflow service selection, the objective function is used to find the best solution. This function consists of penalty over soft constraints *ψ(R)* and QoS function *QoS(R)* as described in Eq. 1. Since QoS function in this case is calculated based on user-defined QoS metrics, i.e. translation accuracy values of each service, the QoS function is modified from Eq. 4 as the summation of QoM function in Eq. 6 which is described in the following Eq. 7.

$$QoS(R) = \sum_{\substack{X_k \in R \\ ServiceInBetween\ (P_i,P_j)=X_k}} QoM(P_i, X_k, P_j) \quad \textbf{(7)}$$

The most optimal result for this problem is {en-zh translation service, zh-en translation service, no translation service, no translation service, zh-en translation service, en-zh translation service}.

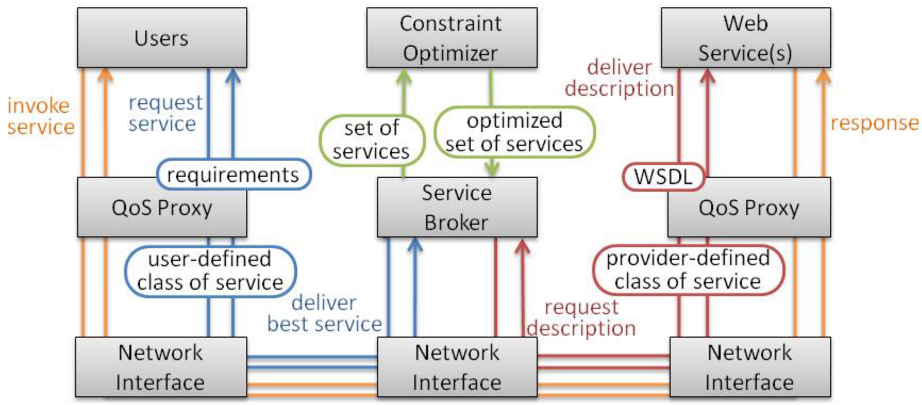


Fig. 5. User-centered QoS architecture

VI. USER-CENTERED QOS ARCHITECTURE

In this section, we implement user-centered QoS in a real system by designing the user-centered architecture for web service selection. To support user-centered QoS framework, we extend the original version of QoS proxy as been previously introduced in [21]. In our architecture, the job of QoS proxy is to translate user requirements of web services and QoS into user-defined class of service. Another job of QoS proxy is to translate WSDL into provider-defined class of service. These two classes of service can be evaluated in constraint optimizer sent by service broker. Fig. 5 illustrates a complete architecture between web service user(s), service broker, and web service provider(s).

In this architecture, each provider can offer different classes of service for different QoS and each class of service can be utilized by more than one user. By having these two kinds of class of service, there is flexibility for users to (re)define their own QoS metric with their own QoS value. This architecture also has an advantage of allowing users to create a new QoS metric based on their needs if the existing class of service is not suitable for them.

The scenario in our architecture is as follows. Initially user requests a service by defining her requirements through QoS proxy in which translating the requirements into class of service and sending it to service broker. Service broker then requests service descriptions based on broker's own database or third party, such as UDDI, to service provider. Getting a description request by service broker, a service provider sends his class of service that is previously translated by QoS proxy from WSDL. The next step is running the constraint optimization algorithm based on the constraints inside user-defined and provider-defined class of service. The constraints together with a set of potential services sent by service broker fed into constraint optimizer to produce a number of feasible services which then can be ranked to find the optimal solution. The final step is the service invocation from user after receiving the best service from service broker.

VII. CONCLUSION

In this work, we proposed a new concept in service oriented computing, i.e. user-centered QoS in combining web services. User-centered QoS is a QoS defined by the interaction between service user(s) and the service itself. The previous concept of QoS in service oriented computing is a QoS that is delivered by service provider to service user. This is contradicted to the concept of *service* in service oriented computing that should be based on the provider and user interaction. This is also against the fact that the best practices of most service oriented applications, especially in multiuser environment, need the QoS interaction between user skills / preferences and provider. Three examples are given in this paper, QoS of travel planner service used by multi-national passengers use with different judgment on hospitality factor, QoS of multimedia services decided on user's behaviour, and QoS of language service based on language capability of each user. In this paper, we gave a complete explanation of user-centered QoS problem to the last example, i.e. language service.

In this paper, we presented a fundamental QoS related problem. This problem arose when we used multilingual chat service that combines several language services, such as translation services and morphological analyzer services in different languages. It started when there were two users using the service. They were Japanese and Chinese users. The Japanese user was good in English, but the Chinese user had no English capability. The chat service thus should automatically provide Japanese-Chinese translation service. After a while, another user from Indonesia who could speak English considerably enough joined the conversation. Since the Indonesian-English translation was available, the chat service was composed by multi-hop translation service for Japanese-English-Indonesian and Chinese-English-Indonesian. However, the QoS of these multi-hop translation services was not good enough. All users got disappointed of this irritating communication. In user-centered QoS aware, the chat service should automatically provide no-translation chat service between Japanese and Indonesian since they have a quality of

English much better than the QoS of multi-hop translation services.

In our experiment, the problem of user-centered QoS cannot be solved in one-stage of service selection. Therefore, we proposed a novel two-stage approach for combining services, i.e. intra-workflow and inter-workflow service selection. Intra-workflow service selection is used to calculate the most optimal QoS value for each possible workflow. Inter-workflow service selection is used to search for the most optimal solution by utilizing the QoS values obtained from intra-workflow service selection. This two service selections utilize the modified technique of constraint optimization and a reliable architecture based on user-defined and provider-defined class of service.

Acknowledgment

This research was partially supported by a Grant-in-Aid for Scientific Research (A) (21240014, 2009-2011) from Japan Society for the Promotion of Science (JSPS), and also from Global COE Program on Informatics Education and Research Center for Knowledge-Circulating Society.

References

[1] Y. Murakami, T. Ishida, T. Nakaguchi, "Infrastructure for Language Service Composition," in *Proc. SKG'06*, 2006.

[2] S. Mittal, B. Falkenhainer, Dynamic constraint satisfaction problems. in *Proc. AAAI'90*, 1990, pp. 25–32. MIT Press.

[3] S. Ran, "A Model for Web Services Discovery with QoS," *ACM SIGecom Exchange*, vol. 4, issue 1, 2003.

[4] M. Paolucci, T. Kawamura, T.R. Payne, K. Sycara, "Semantic Matching of Web Services Capabilities," in: *Proc. ISWC'02*, 2002.

[5] L. Zeng, B. Benatallah, M. Dumas, J. Kalagnanam, Q.Z. Sheng, "Web Engineering: Quality Driven Web Service Composition," in *Proc. WWW'03*, 2003, ACM Press.

[6] A. Sahai, J. Ouyang, V. Machiraju, K. Wurster, BizQoS, "Specifying and Guaranteeing Quality of Service for Web Services through Real Time Measurement and Adaptive Control," Hewlett-Packard Labs Technical Report HPL-2001-134, 2001.

[7] A.B. Hassine, S. Matsubara, T. Ishida, "Constraint-based Approach to Horizontal Web Service Composition," in *Proc. ISWC'06*, 2006, LNCS, vol. 4273, pp. 130-143. Springer-Verlag.

[8] N. Li, S. Channa, A.W. Shaikh, X. Fu, "Constraint Satisfaction in Dynamic Web Service Composition," in *Proc. DEXA'05*, 2005, pp. 658-664.

[9] L. Li, J. Wei, T. Huang, "High Performance Approach for Multi-QoS Constrained Web Services Selection," in *Proc. ICSOC'07*, 2007, pp. 283-294.

[10] L. Zeng, H. Lei, H. Chang, "Monitoring the QoS for Web Services," in *Proc ICSOC'07*, 2007, pp. 132-144.

[11] C. Zhang, R.N. Chang, C. Perng, E. So, C. Tang, T. Tao, "QoS-Aware Optimization of Composite-Service Fulfillment Policy," in *Proc ICSOC'07*, 2007, pp. 11-19.

[12] K. Papineni, S. Roukos, T. Ward, W. Zhu, "BLEU: a method for automatic evaluation of machine translation," in *Proc. ACL'02*, 2002, pp. 311-318.

[13] G. Canfora, M.D. Penta, R. Esposito, M.L. Villani, "A framework for QoS-aware binding and re-binding of composite web services," Journal of Systems and Software, vol 81, issue 10, pp. 1754-1769, 2008.

[14] M. Tanaka, T. Ishida, Y. Murakami, S. Morimoto, "Service Supervision: Coordinating Web Services in Open Environment," in *Proc. ICWS'09*, 2009.

[15] R. Tanaka, Y. Murakami, T. Ishida, "Context-Based Approach for Pivot Translation Services," in *Proc. IJCAI'09*, 2009.

[16] T. Ishida, "Language Grid: An Infrastructure for Intercultural Collaboration," in *Proc. SAINT'06,* 2006, pp. 96-100.

[17] A. Bramantoro, M. Tanaka, Y. Murakami, U. Schäfer, T. Ishida, "A Hybrid Integrated Architecture for Language Service Composition," in *Proc. ICWS'08*, 2008, pp. 345-352.

[18] I. Matsumura, T. Ishida, Y. Murakami, Y. Fujishiro, "Situated Web Service: Context-Aware Approach to High Speed Web Service Communication," in *Proc. ICWS'06*, 2006, pp. 673-680.

[19] V. Deora, J. Shao, W.A. Gray, N.J. Fiddian, "A Quality of Service Management Framework Based on User Expectations," in *Proc. ICSOC'03*, 2003, pp. 104-114. Springer, Heidelberg.

[20] R. Dechter, *Constraint Processing*, Morgan Kaufmann, San Francisco, 2003.

[21] M. Tian, A. Gramm, T. Naumowicz, H. Ritter, J. Jchiller, "Efficient Selection and Monitoring of QoS-aware Web services with the WS-QoS Framework," in *Proc. WI'04*, 2004, pp. 152-158. IEEE Press.

[22] L. J. Zhang, J. Zhang, H. Cai, *Services Computing*. Springer-Verlag, 2007.

[23] S. Chaari, Y. Badr, and F. Biennier, "Enhancing web service, selection by qos-based ontology and ws-policy," in *Proc. SAC '08, 2008*, pp. 2426–2431, New York, NY, USA, ACM.

[24] H. Q. Yu and S. Reiff-Marganiec, "A method for automated web service selection," in *Proc. SERVICES'08, 2008*, pp. 513–520, USA.

[25] S. Lamparter, A. Ankolekar, R. Studer, and S. Grimm, "Preference-based selection of highly configurable web services," in *Proc. WWW'07*, 2007, pp. 1013–1022, USA, ACM.

2010 IEEE International Conference on Services Computing

Market-Based QoS Control for Voluntary Services

Yohei Murakami
Language Grid Project,
National Institute of Information and Communications Technology (NICT), Kyoto, Japan
Email: yohei@nict.go.jp

Naoki Miyata, Toru Ishida
Department of Social Informatics,
Kyoto University, Kyoto, Japan
Email: miyata@ai.soc.i.kyoto-u.ac.jp,
ishida@i.kyoto-u.ac.jp

***Abstract*—With the development of services computing technology, more and more voluntary services have been available on the Internet. When using voluntary services, users tend to demand higher QoS (e.g., throughput of the services) than they actually need because there is no cost. To control QoS of the voluntary services appropriately, it is necessary to design resource allocation mechanism using utilities on both service users and providers. Therefore, we have proposed market-oriented resource allocation where users and providers exchange system resources and QoS based on their utilities. In our proposed approach, service users obtain more utilities if higher QoS is allocated according to their preferences in using the services, while service providers get more utilities if their services are more effectively used by their preferred users. In order to validate the proposed method, we have compared market-based approach with demand-based approach by simulation. The simulation results show that our approach motivated users to give true demands more than demand-based approach.**

***Keywords*-QoS Control; Voluntary Services; Market-Oriented Model;**

I. Introduction

In open-source development, knowledge sharing and voluntary services by community members lead to innovation. This trend is reaching to services computing domain[1]. To innovate services computing technology, more and more voluntary services have been available on the Internet. The Language Grid Project [2] also aims to develop a system where language resources (e.g., machine translator, dictionaries, and so on) are voluntarily provided as a Web service; users can then compose new language services using the existing language services. The language service providers provide services by utilizing their language resources and computational resources of the system. Users can employ the language services for free only for non-profit objectives. We call these services where service providers volunteer resources that can be used by other users for free, as *voluntary services*.

The objective of voluntary services is to contribute to certain communities. For example, an NPO that assists foreign tourists provides voluntary services with the expectation that these services will be used by the tourists during their visit to a country. An academic organization provides voluntary services with the expectation that they will be used by students studying a particular subject.

In order to prevent such systems from overloading, it is necessary to suitably allocate computational resources to users. This resource allocation is based on the preferences of the providers as well as those of the users. Since the service providers cannot obtain any profits from providing their services, it is necessary to motivate them by reflecting their preferences in the system. However, users tend to input " pseudo-demands", that users demand more computation resources than they need and they actually do not use the allocated computation resources. Pseudo-demands decrease both the utility of other users using the same services and that of the service providers.

In this research, we consider dynamic resource allocation to control QoS of voluntary services. The problems involved in allocating computation resources are as the follows.

- Establishment of resource allocation in voluntary services Users in voluntary services have no cost constraints since the services are free. The objective of the providers is that their resources are effectively utilized by users. In order to realize the suitable resource allocation, it is necessary to clearly define the purposes and constraints of voluntary services and the characteristics that the allocation methods should have.
- Suitably allocate resources in large-scale systems There are many users and providers in open Internet services. The greater the number of users and providers that exist in the system, the greater is the computational time required to allocate resources. Therefore methods that can suitably allocate resources within an appropriate time in such large-scale open systems and that also have the necessary characteristics for resource allocation are required.

In this research, we model a resource allocation problem for voluntary services to control QoS of voluntary services. Then, we apply to this problem a market-based approach using a heuristic in order to solve the problem within an appropriate time.

II. QoS Control

QoS control methods have been proposed for suitably allocating resources in order to efficiently utilize the finite

978-0-7695-4126-6/10 $26.00 © 2010 IEEE
DOI 10.1109/SCC.2010.66

computation resources in large-scale systems. Zeng, L. et al. [3] formulated the problem of web service composition in terms of QoS and proposed AgFlow; this approach selects appropriate services using integer programming. AgFlow has a service quality model to evaluate the overall quality of composite web services. AgFlow also selects services based on each task or the global allocation of tasks using integer programming for composite service execution. Menasce, D. A. et al. [4] proposed an architecture that allocates QoS based on user utilities in a service oriented architecture. In their proposed approach, users provide QoS brokers using the utility functions of the users and the cost constraints for the required services. Service providers register with the broker by providing service demands for each of the resources used by the provided services and the cost function for each of the services. The QoS broker uses analytic queuing models to predict the QoS values of the various services that can be selected under varying workload conditions. Buyya, R. et al. [5] describe an approach for introducing a market model to general grid systems. There exist various users and providers. There also exist various objectives, strategies and patterns of demand and supply. They introduce a competitive market model in order to realize a system where users and providers can maximize their utility. As a result, resources are allocated to users based on the various utilities of users and providers. In grid service areas, other researches on resource allocation employ economic approach and reinforcement learning [6], [7].

On the other hand, we assume that the voluntary services are free. That is why users may input pseudo-demands if systems simply allocate computation resources based on the demands of users. Moreover, the approach that charges a fee for the voluntary services are not appropriate because the objective of providers is to not obtain any profit from providing services. We propose an approach that is applicable to voluntary services.

III. QoS Control for Voluntary Services

A. Voluntary Services

The overview of voluntary services is shown in Figure 1. Voluntary service delivery platform provides a finite computational resource for common use. Service providers offer web service using the shared computational resources. The objective of the providers is that their services are more effectively used by their preferred users. On the other hand, users select the necessary services from the available services according to their preferences. Administrators monitor the platform and manage the access rights so that the entire system is suitably utilized. In this paper, we call the shared computational resources as "system resource" and the throughput of provided services as "QoS" [8].

Voluntary services become overloaded due to burst access since the shared computational resources are finite. Service providers do not consider the system resources when they limit the use of their QoS. In order to prevent the system from overloading, system resources must be suitably allocated to the service users based on the preferences of the service users and service providers. The objective of the providers is that their services are more effectively used by their preferred users, while that of users is to satisfy their requirements using the services.

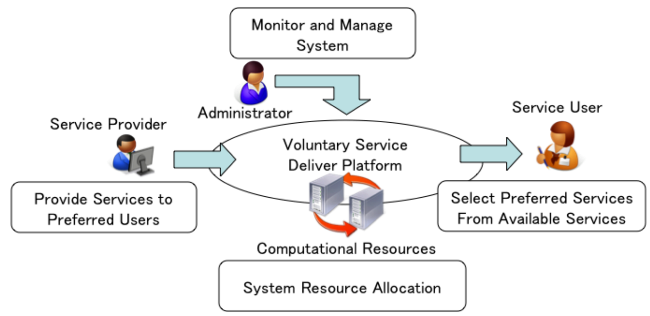


Figure 1. Stakeholders in voluntary services

B. Stakeholders

We now describe, the models of users, service providers and the administrator in voluntary services. Table 1 shows the motivation and problems of the stakeholders. Providers provide their services using the shared system resources. The objective of providers is to contribute their services to certain communities and users. The utility of providers increases when their services are utilized by the targeted users. In other words, providers have preferences for users and their utility is determined based on these preferences and the amount of consumed QoS.

Users use services in order to complete their tasks. There are multiple interchangeable services for a task in a system. Users select services among the available services. Users have multiple requirements, each of which is assigned with a weight. A requirement has a maximum amount of allocated QoS and a set of services. In this research, we assume that users know their future requirements.

The objective of an administrator is to motivate more users and providers to participate in the system and activate the system. In order to achieve this, the administrator must allocate system resources to users based on the preferences of users and providers. That will motivate providers to offer their resources and make more QoS available. An increase in the number of QoS will lead to an increase in the number of users. Finally, the opportunities that users can utilize the offered resources will motivate more providers to offer their resources, thereby activating the system.

C. Resource Allocation Problem in Voluntary Services

The purpose of resource allocation in voluntary services is to realize suitable resource allocation based on the pref-

Table I
OBJECTIVES AND PROBLEMS OF STAKEHOLDERS

	Incentives	Problems
Providers	Contribute to preferred users	Access control of non-preferred user
Users	Utilize preferred services	Pseudo-demands
Administrators	Motivate users and providers to participate	Suitable resource allocation

erences of users and providers. There are two restrictions in allocating resource in these systems.

- A fee cannot be charged for system resources
  Charging a fee for system resources does not lead to suitable resource allocation in voluntary services since users may not have the required amount of money; further, the purpose of the system is to not obtain a profit.
- True demands and pseudo-demand cannot be differentiated
  It is difficult to determine whether the demands of users are true or not, either beforehand or posteriori in voluntary services. If the system determines whether the demand is true or not based on whether the resource is actually used, it will motivate users to waste the resource in order to avoid a penalty. Therefore, this approach is not effective.

When voluntary services are free, users are not penalized even if they input greater demands than their true demands or unnecessary demands. If the excessively allocated resources are actually not used, it decreases the utility of other users who want to use the same services and that of providers who want their services to be used. We term such demand as " pseudo-demand".

The proposed resource allocation mechanism should have the following three characteristics due to these restrictions.

- Motivate users to input true demands
  Since the system is unable to differentiate between true demands and pseudo-demands, the mechanism should motivate users to input true demands.
- Suppress the effects of pseudo-demands
  It is impossible to completely eliminate pseudo-demands in large-scale systems. The mechanism should suppress the effects of pseudo-demands.

D. Modeling Resource Allocation Problem

The model of users and providers is shown in Figure 2. The provided service has a product function that determines the amount of utilizable QoS based on the allocated system resources. Let x_t denote the system resource allocated to a service at t. Let $Q_s(x_t)$ denote the amount of QoS that service s can provide at time t. $Q_s(x_t)$ is calculated as follows.

$$Q_s(x_t) = Q_s^0 - \frac{Q_s^0}{1+\gamma_s x_t} \quad (1)$$

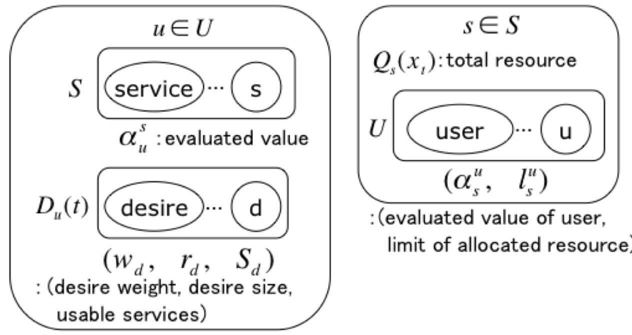


Figure 2. Provider and user model

Q_s^0 is the upper limit of the resources that provider s can produce. γ_s is a variable indicating how the production is increased by adding system resource.

Service providers decide the QoS allocation to users. Let $\alpha_s^u \quad (0 \le \alpha_s^u \le 1)$ denote service s provider 's evaluation for each user u, $l_s^u(t)$ denote the upper limit of the QoS of service s allocated to user u at time t, and $q_s^u(t)$ denote the QoS of service s allocated to user $u \in U$ at time t. On the other hand, service users decide the QoS to consume from the allocated QoS. Let $y_u^s(t)$ denote the QoS of service s that user u uses. $y_u^s(t)$ is less than $q_s^u(t)$. $q_s^u(t)$ is less than the upper limit $l_s^u(t)$. The sum of the QoS of service s allocated to users at time t is less than $Q_s(x_t)$.

Let $\alpha_u^s(0 \le \alpha_u^s \le 1)$ denote evaluation by user u for service s. Let $D_u(t)$ denote a set of requirements of user u at time t. Let w_d denote the importance of each requirement d in $D_u(t)$. Let r_d denote the ceiling amount of QoS for each requirement d. Let S_d denote the set of services used to satisfy the requirement d. In this problem, it is assumed that users know their future requirements. For the sake of simplicity, we assume that sets of services that a user uses at a time is independent. Namely, the equation $S_{d_1} \cap S_{d_2} = \phi$ $(d_1, d_2 \in D_u(t), d_1 \ne d_2)$ holds.

The utility of users is defined as weighted sum of satisfaction degree of each requirement d. Namely, the utility of user u at time t is determined as follows.

$$utility_u(\times_{s\in S_d} y_u^s(t)) = \sum_{d\in D_u(t)} w_d \frac{\sum_{s\in S_d} \alpha_u^s y_u^s(t)}{r_d}$$
$$\text{s. t.} \quad y_u^s(t) \le q_s^u(t), \sum_{s\in S_d} y_u^s(t) \le r_d \quad (2)$$

The utility of providers is defined as sum of product of evaluation for the users and QoS consumed by the users. Namely, the utility of provider of service s at time t is defined as follows.

$$utility_s(\times_{u\in U} y_u^s(t)) = \sum_{u\in U} \alpha_s^u y_u^s(t)$$
$$\text{s. t.} \quad \sum_{u\in U} y_u^s(t) \leq Q_s(t) \tag{3}$$

The goal of the users and providers is to maximize the utilities at each moment.

The unit of QoS allocated to users varies depending on the type of service. For example, in dictionary services, where the cost is almost constant for each invocation, the unit of QoS is the number of service invocation. In translation services or morphological analysis services, where the cost varies depending on the input argument, the unit of QoS is the length of the translated sentence or the size of the analysis result.

IV. Market-Oriented Resource Allocation

We propose a market-oriented approach that deals with system resources and QoS as goods for the above-described resource allocation problem. There are three reasons for introducing the market-model.

- Allocate resources suitably by the market mechanism
  The market mechanism can realize the suitable resource allocation based on the preferences of users and providers, enables users to utilize the finite system resources.
- Motivate users to input true demand
  In the market model, the finite system resource available for obtaining QoS is allocated to users in advance. Since pseudo-demands waste system resources that could be used for obtaining QoS for true demands, it motivates users to input true demands.
- Suppress effects of pseudo-demand
  The proposed method allocates system resource to users without considering their demands. In the market model, the system resources are equably allocated to users. Even if a user inputs longer and larger pseudo-demands than other true demands, the amount of system resource that the pseudo-demands can use is less than that used by the other true demands. This can suppress the effects of pseudo-demands.

We introduce a consumer-producer model and a current-future model proposed by Yamaki H. et al. [9]. We extend these models so that they can be applied to voluntary services since these models assume that the objective of the providers is to maximize profits.

In the consumer-producer model, the finite system resources are allocated to users as consumers. Users decide the system resource allocation to providers based on their demands. Providers produce QoS and allocate them to users. This model equally allocates system resources to users in order to equalize the opportunity of users to obtain QoS.

In the current-future model, the goods exchanged in the market are classified into current and future goods. Users can exchange their goods with each other according to their demands. This model allows users to exchange system resource based on their demands.

Users in this model decide the system resource allocation based on the current utility, expected future utility, and exchange ratio between current and future system resources. They attempt to obtain QoS in order to efficiently increase their utility. It is more efficient to obtain QoS that have a small demand from a few users than those having large demand from many users.

Providers allocate the produced QoS based on the demands of users and their evaluation for users. If a provider allocates all QoS only to highly evaluated users, the amount of system resources allocated to the provider decrease. Providers have to decide a suitable allocation in which the amount of allocated system resources is sufficient and their utility is high.

A. Consumer-Producer Model

The consumer agent corresponds one-to-one with a user of the system. Users evaluate the QoS and not system resources. In other words, instead of determining how many system resources they have, users must determine how their demands are satisfied by the QoS. The preference of user u for QoS is implied in the utility function. The function is based on the evaluation for the services to be used.

The producer agent in the market model corresponds with the service providers. The agent converts a system resource allocated by a user into a QoS and allocates the produced QoS to users so that their utility would be maximized. In this model, the consumer agent initially has no QoS. All QoS have to be produced by producer agents. Providers allocate their QoS based on the amount of produced QoS, the demands of users, and evaluation for users.

B. Current-Future Model

In the current-future model, time is divided into equal intervals. A unit of time for the present time is defined as current. A certain period $(T-1)$ after the current is defined as future. When the total amount of current system resources is β, the total amount of current system resources that users possess equals β. The total amount of future system resources equals $(T-1)\beta$.

The procedure of dealing with goods in the current-future model is shown in Figure 3. Let e_u^c, e_u^f denote the initial current and future system resources of u. Users exchange system resources between each other (1 in the figure) to decide x_u^c, x_u^f, the current and future system resource allocation, respectively. Users obtain QoS by using the system resources (2 in the figure). Next, when a unit

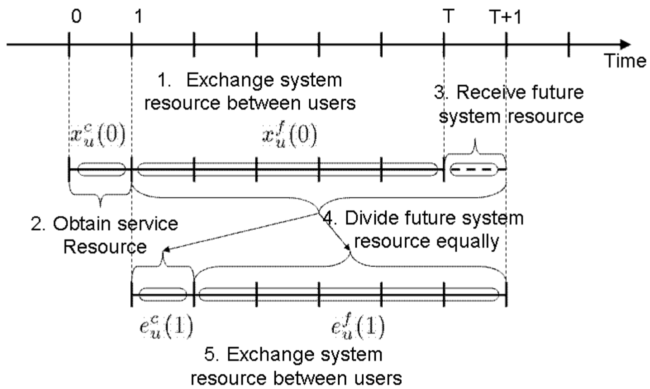


Figure 3. Current-Future model [9]

time elapses, the future goods are reflected to the system resource allocation in the next time unit. $\beta/|U|$ is allocated to each user (3 in the figure). Then, the resources that the users have are updated for a new time slice. Users divide their future resources $1:(T-1)$ to refrect to the resource allocation at the next time unit (4 in the figure). Namely, the current and future system resources that user u possesses at time t are determined as follows.

$$e_u^c(t) = \frac{1}{T}\left(x_u^f(t-1) + \frac{\beta}{|U|}\right) \tag{4}$$

$$e_u^f(t) = \frac{T-1}{T}\left(x_u^f(t-1) + \frac{\beta}{|U|}\right) \tag{5}$$

After the above procedure for initial resource allocation, the utility function and the demands of users and providers are updated and the resources are then allocated (5 in the figure). Then, the resource allocation is decided for the new period.

In the current-future model, providers have to expect future QoS allocations based on the allocated future system resources. The amount of future QoS that a provider can provide is determined through the following procedures. Initially, the allocated future system resources divided by the considered period is substituted in the product function to calculate the future QoS per unit time. The amount of future QoS is equal to the product of the QoS per unit time and the considered period. Namely, when $(T-1)$ future is considered and x_f future system resources are allocated, the amount of future QoS is given as follows.

$$(T-1)Q_s\left(\frac{x_f}{T-1}\right) = (T-1)\left(Q_s^0 - \frac{Q_s^0}{1+\gamma_s\frac{x_f}{T-1}}\right) \tag{6}$$

C. Resource Allocation Using Sensitivity Factor

We propose a heuristics approach that decides the resource allocation for the above-mentioned market model since it requires considerable time to calculate an optimal or Pareto-optimal resource allocation. We extend the technique proposed by Kuwabara, K. et al. [10] and apply it to the model.

Algorithm 1 Release current and future system resources

1: α: sensitivity factor
2: $p(i)$: exchange ratio of current system resource to future ones at i
3: e_u^c, e_u^f: initial current/future system resources
4: $x_u^c(i), x_u^f(i)$: current/future system resources at i
5: $g_u^c(i), g_u^f(i)$: released current/future system resources
6: $U_u^c(i), U_u^f(i)$: current/future utility per unit system resource at i
7: θ: threshold of released system resources
8: **if** $p(i-1)U_u^f(i-1) < U_u^c(i-1)$ **then**
9: $\quad (g_u^c(i), g_u^f(i)) = \begin{cases} (0, g_u^f(i-1) + \alpha(e_u^f - g_u^f(i-1)) & (g_u^c(i-1) < \theta) \\ ((1-\alpha)g_u^c(i-1), 0) & \text{(otherwise)} \end{cases}$
10: **else if** $p(i-1)U_u^f(i-1) > U_u^c(i-1)$ **then**
11: $\quad (g_u^c(i), g_u^f(i)) = \begin{cases} (g_u^c(i-1) + \alpha(e_u^c - g_u^c(i-1)), 0) & (g_u^f(i-1) < \theta) \\ (0, (1-\alpha)g_u^f(i-1)) & \text{(otherwise)} \end{cases}$
12: **else**
13: $\quad (g_u^c(i), g_u^f(i)) = (g_u^c(i-1), g_u^f(i-1))$
14: **end if**

Users decide the amount of current and future system resources to release based on the current and future utilities and the exchange ratio between the current and future system resources. The released system resources are reallocated to users following a rule of the market. Then, users decide the amount of current and future system resources allocated to providers. Providers decide the allocation of QoS to users. Finally, users decide the amount of QoS to use. The utility of users and providers is determined. The above-mentioned procedure is repeated until the utilities of the users and providers converge.

In the repetition of demand and supply, users adjust the amount of released system resources based on the resource allocation at the previous iteration. The initial amount of released system resources is determined so that the ratio of current system resources to future ones equals the ratio of current requirements weights to future ones. They adjust the amount based on the exchange ratio of current system resources to future ones and the current and future utilities. The adjustment algorithm is described in Algorithm 1.

When the utility per unit future system resource is larger than that by the current system resources that users can obtain by releasing a unit of future system resources, users intend to increase their current utility by releasing more future system resources or less current ones (line 8, 9), or vice versa (line 10, 11).

The system resources released by users are reallocated to users according to a rule of the market. In the market model, the amount of future system resources reallocated to a user is derived from the ratio of the current system resources released by the user to the sum of the released

Algorithm 2 System resource allocation to services

1: α: sensitivity factor
2: $S_d \subseteq S$: set of services that d uses
3: $D_u^c \subseteq D_u$: set of currently active requirements
4: $x_u^c(i)$: u's current system resources allocated at i
5: $d_{best}(i) \in D_u^c$: demand having the best utility per unit resource in D_u^c at i
6: $s_{best}(i) \in S_d$: service providing d the best utility per unit resource in S_d at i
7: $rate_d(i)$: rate of resources allocated to d at i
8: $rate_d^s(i)$: rate of resources allocated to s by d at i
9: **for all** $d \in D_u^c$ **do**
10: $\quad rate_d(i) = \begin{cases} rate_d(i-1) + \alpha(1 - rate_d(i-1)) \\ \qquad (d \text{ is } d_{best}) \\ (1-\alpha) rate_d(i-1) \quad \text{(otherwise)} \end{cases}$
11: $\quad m_d(i) = rate_d(i) x_u^c(i)$: resources allocated to d at i
12: $\quad$ **for all** $s \in S_d$ **do**
13: $\quad\quad rate_d^s(i) = \begin{cases} rate_d^s(i-1) + \alpha(1 - rate_d^s(i-1)) \\ \qquad (s \text{ is } s_{best}) \\ (1-\alpha) rate_d^s(i-1) \quad \text{(otherwise)} \end{cases}$
14: $\quad\quad m_d^s(i) = rate_d^s(i) m_d(i)$: allocate on s by d at i
15: $\quad$ **end for**
16: **end for**

current system resources and vice versa.Let b_u^c, b_u^f denote the current or future system resources reallocated to user u. b_u^c, b_u^f are given as follows.

$$(b_u^c, b_u^f) = \left(\frac{g_u^f}{\sum_{u' \in U} g_{u'}^f} \sum_{u' \in U} g_{u'}^c, \frac{g_u^c}{\sum_{u' \in U} g_{u'}^c} \sum_{u' \in U} g_{u'}^f\right) \quad (7)$$

After the system resources are reallocated, users allocate their system resources to providers. The behavior that allocates system resources to providers is shown in Algorithm 2. Although the allocation of only current system resources is described here, that of future system resources is determined in a similar manner. The amount of system resources allocated for the requirements is adjusted based on the resource allocation at the previous iteration. The requirements that increase the utility most efficiently in the current requirements have more system resources allocated to them in the current iteration than in the previous one. Other requirements have less system resources. The initial amount of system resources allocated to the requirements is determined based on the weights of these requirements. (line 10)

Then, the requirements allocate the given system resources to services. The initial allocation of system resources is determined based on the evaluated values of the services. The amount of system resources allocated to services is adjusted in the same manner as the allocation of requirements. (line 13)

In a manner similar to the above-mentioned procedure, the allocation of future system resources is coordinated

Algorithm 3 Service resource allocation to users

1: U_s : users that can use s
2: $q_s^u = 0 (u \in U_s)$: resource allocated to u by s
3: $U_{left} = U_s$: users unsatisfied with resource
4: $c_s^u = \min(r_d, l_s^u)$: ceiling amount of resource for u
5: $q_{left} = Q_s$: remaining resources that s has
6: **while** $U_{left} \neq \phi$ and $left > 0$ **do**
7: $\quad q_{given} = 0$
8: $\quad$ **for all** $u \in u_{left}$ **do**
9: $\quad\quad q = \min(q_{left} m_u^s \alpha_s^u / \sum_{u \in U_{left}} m_u^s \alpha_s^u, c_s^u - q_s^u)$
10: $\quad\quad (q_{given}, q_s^u) = (q_{given} + q, q_s^u + q)$
11: $\quad\quad$ **if** $q_s^u == c_s^u$ **then**
12: $\quad\quad\quad U_{left} = U_{left} \setminus \{u\}$
13: $\quad\quad$ **end if**
14: $\quad$ **end for**
15: $\quad q_{left} = q_{left} - q_{given}$
16: **end while**

based on the utility gained by providers at the previous iteration. The weight of the future requirements equals the weight of the requirements multiplied by the period that the requirement is active in the considered future. That is, w_d^f, which is the weight of the future requirement, equals $w_d(\min(t_d^{end}, t+T-1) - \max(t+1, t_{start}))$. $rate_d(0)$ equals $w_d^f / \sum_{d' \in D_u^f} w_{d'}^f$

The providers allocate QoS to users based on the allocated system resources and the evaluated values of the users. The algorithm is shown in Algorithm 3. Providers treat the amount of allocated system resources multiplied by the evaluated values of the users as the ratio of the QoS allocated to users. The smaller value between the calculated amount and the ceiling of the user is allocated to the user. This procedure is repeated until there are no unsatisfied users or no QoS.

Finally, users decide the amount of allocated service resource to use. Users select the QoS that maximize their utility from among the allocated QoS.

V. Simulation of Resource Allocation

In this section, the settings and results of the simulations conducted to verify the market model and the behaviors of users and providers are described.

A. Simulation Settings

We conduct simulations to verify the resource allocation based on the preferences of users and providers using the above-mentioned market model and the behaviors of users and providers. In this simulation, a random number is identically distributed. The number of users is 100 ($|U| = 100$), and the number of services is 100 ($|S| = 100$). The simulated period is 200. The number of the requirements that a user has in the given period is a random number between 6 and 10. The period of a requirement is a random number

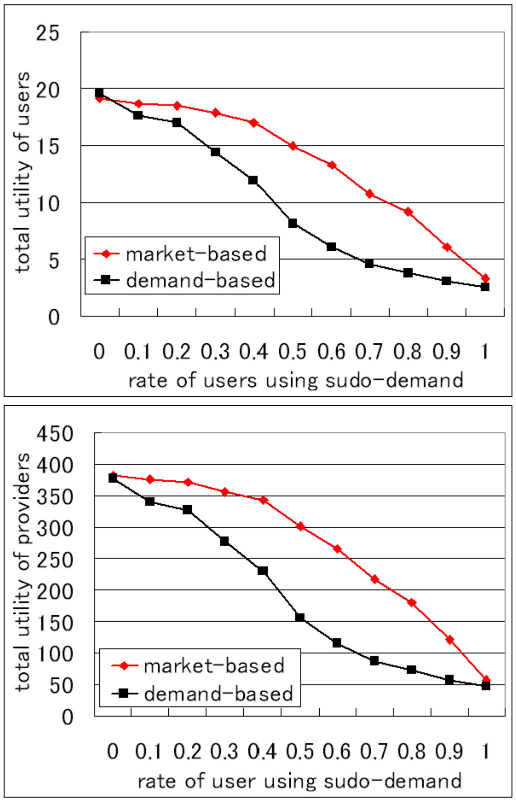


Figure 4. Utility of users and providers in changing the rate of pseudo-demands

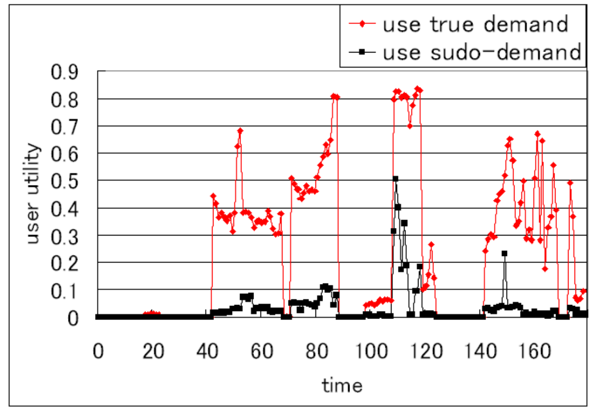


Figure 5. Utility of a user using pseudo-demands and true demands

between 10 and 30. r_d, which is the size of a requirement is a random number between 10 and 20. w_d, which is the weight of a requirement is a random number between 0 and 1. $|S_d|$, which is the number of services a requirement uses, is a random number between 3 and 7. The service has a quality value that is a random number between 0 and 1. The evaluated value of a service is normalized based on the quality value. Q_s, which is the largest amount of QoS a provider can provide, is a random number between 10 and 20. The period considered as future is 20 ($T = 20$). α, which is the sensitivity factor used by the user for adjusting system resource allocation, is 0.01. The sum of the system resources is 1000 ($\beta = 1000$). This implies that each user receives 10 system resources every time.

B. Simulation Results

We compare the utility of users and providers by changing the ratio of users who input pseudo-demands. The users input their true demands as a pseudo-demand between 0 and 200. The result is shown in Figure 4. The horizontal axis shows the rate of the users who input pseudo-demands. The vertical axis shows the average sum of the utilities.

When the rate of users using pseudo-demands increases, the utility of the users and providers decreases in both the market-based and demand-based approaches. The QoS allocated to pseudo-demands are not used. Since the amount of QoS allocated to true demands decreases, the utility of the users and providers decreases.

The decrease in the utility of users and providers in our approach is smaller than that in the demand-based approach. In the market-based approach, the pseudo-demands consume the system resources of a user at every time slice. Then, since the amount of system resources that a pseudo-demand can use is relatively smaller than that which a true demand can use in a certain time, the amount of QoS allocated to pseudo-demands is smaller than that allocated to true demands. On the other hand, pseudo-demands in the demand-based approach can obtain as many system resources as other true demands can. As a result, our approach can decrease the effect of pseudo-demands on other users to a greater extent than the demand-based approach.

The system needs to motivate users to input their true demands since pseudo-demands decrease the social surplus in the system. Here the utility of a user using true demands is compared to that of a user using pseudo-demands. In this simulation, other users input true demands.

The utility of a user using pseudo-demands is almost the same as that using only true demands in the case of the demand-based approach. Even if the user has wasted considerable QoS previously, the system resources are used for the user in a manner similar to that for other users. It is difficult for the system to motivate users to input true-demands.

In our approach, the utility of a user using true demands as compared to that of the same user using pseudo-demands is shown in Figure 5. When the user inputs pseudo-demands, the user uses his current system resources to obtain QoS for the pseudo-demands. As a result, the utility of the user using pseudo-demands is much smaller than that which the user gains by using true demands since the amount of system resources that the user can use for his true demands is small. This implies that our approach can motivate users to use true demands.

VI. Conclusion

In this research, we consider resource allocation for voluntary services. In such systems, users and providers have preferences for each other. The system resources should be allocated based on these preferences. Additionally, since users have no cost constraints, users may input pseudo-demands that do not actually use the allocated resources and therefore prevent suitable resource allocation. In order to realize suitable resource allocation in such systems, we model voluntary services and propose a market-based approach. This research makes the following two contributions.

- Model the resource allocation problem
  We clarify and model the requirements of resource allocation in voluntary services based on actual systems. In such systems, users and providers have preferences for each other. Providers decide the allocation of their resources to users. We also describe the characteristics that allocation methods should have.
- Propose a resource allocation method using heuristics
  We propose a market model comprising the current-future model and the consumer-producer model, and an approach for allocating resources using the market-based model in order to realize suitable resource allocation in voluntary systems. This approach can suitably allocate resources in an applicative time even in large-scale systems. We demonstrate that our approach can allocate resources based on the preferences of users and providers; further, it has the characteristics necessary for resource allocation.

The above contributions realize an suitable resource allocation for voluntary services by considering the user' and providers' preferences. It can also motivate users to input true demands and decrease the effects of pseudo-demands on other users.

Acknowledgment

A part of this works was supported by Strategic Information and Communications R&D Promotion Programme from Ministry of Internal Affairs and Communications, and a Grant-in-Aid for Scientific Research (A) (21240014, 2009-2011) from Japan Society for the Promotion of Science (JSPS).

References

[1] L.-J. Zhang, "TSC cloud: Community-driven innovation platform," *IEEE Transactions on Services Computing*, vol. 2, no. 1, pp. 1–2, 2009.

[2] T. Ishida, "Language grid: An infrastructure for intercultural collaboration," in *SAINT '06: Proceedings of the International Symposium on Applications on Internet*. Washington, DC, USA: IEEE Computer Society, 2006, pp. 96–100.

[3] L. Zeng, B. Benatallah, A. H. Ngu, M. Dumas, J. Kalagnanam, and H. Chang, "Qos-aware middleware for web services composition," *IEEE Transactions on Software Engineering*, vol. 30, no. 5, pp. 311–327, 2004.

[4] D. A. Menascé and V. Dubey, "Utility-based qos brokering in service oriented architectures," in *ICWS '07: Proceedings of the IEEE International Conference on Web Services (ICWS'07)*. IEEE Computer Society, 2007, pp. 422–430.

[5] R. Buyya, D. Abramson, and S. Venugopal, "The grid economy," *Proceedings of the IEEE*, vol. 93, no. 3, pp. 698–714, 2005.

[6] C. Weng, M. Li, X. Lu, and Q. Deng, "An economic-based resource management framework in the grid context," in *CCGRID '05: Proceedings of the Fifth IEEE International Symposium on Cluster Computing and the Grid (CCGrid'05) - Volume 1*. Washington, DC, USA: IEEE Computer Society, 2005, pp. 542–549.

[7] A. Galstyan, K. Czajkowski, and K. Lerman, "Resource allocation in the grid using reinforcement learning," in *AAMAS '04: Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems*. Washington, DC, USA: IEEE Computer Society, 2004, pp. 1314–1315.

[8] S. Ran, "A model for web services discovery with QoS," *ACM SIGecom Exchanges*, vol. 4, no. 1, pp. 1–10, 2003.

[9] H. Yamaki, M. P.Wellman, and T. Ishida, "Controlling application qos based on a market model," *The transactions of the Institute of Electronics, Information and Communication Engineers*, vol. 81, no. 5, pp. 540–547, 1998. [Online]. Available: http://ci.nii.ac.jp/naid/110003315712/en/

[10] K. Kuwabara, T. Ishida, Y. Nishibe, and T. Suda, *An equilibratory market-based approach for distributed resource allocation and its applications to communication network control*. River Edge, NJ, USA: World Scientific Publishing Co., Inc., 1996, pp. 53–73.

2010 IEEE International Conference on Services Computing

Service Supervision for Service-oriented Collective Intelligence

Masahiro Tanaka*, Yohei Murakami*, Donghui Lin*, Toru Ishida†*
*Language Grid Project
National Institute of Information and Communications Technology,
3-5 Hikaridai, Seika-cho, Kyoto, Japan
{mtnk, yohei, lindh}@nict.go.jp
†Department of Social Informatics, Kyoto University
Yoshida-honmachi, Sakyo-ku, Kyoto 606-8501 Japan
ishida@i.kyoto-u.ac.jp

***Abstract*—Service-oriented collective intelligence, which creates new value by combining services provided by various organizations via services computing technologies, has been gaining in importance with the development of services computing technologies. Because collective intelligence needs many participants, it is crucial to build a framework where a wide variety of policies of service providers are satisfied. In this paper, we propose an architecture which handles a comprehensive process of service selection, adaptation, and coordination to satisfy policies of service providers. First the system selects services, and then adapts the services to the given policies if any of available services cannot satisfy the policies. To achieve this, we formalized this problem as an extension of constraint satisfaction problem and showed a solution. Moreover, the system often needs to force a composite service to follow protocols given by service providers. Therefore we proposed a method which uses meta-level control functions for composite services in order to change order of service execution.**

***Keywords*-Service Supervision; Web service; coordination; collective intelligence;**

I. Introduction

Various programs and data have become available as Web services with the development of services computing technologies. To realize service-oriented collective intelligence, which integrates the services provided by various organizations, it is required to build a new framework satisfying the service providers' policies. Because collective intelligence needs many participants and it is crucial to convince service providers that their policy will be certainly satisfied on a framework where their services are deployed.

Service providers may have a wide variety of policies, such as limitation of transferred data, constraints on combinations of services and so on. From this aspect, service-oriented collective intelligence is quite different with traditional collective knowledge which relies on single license e.g. Wikipedia. For example, the Language Grid[1] is a platform for service-oriented collective intelligence, which allows to combine various language services such as machine translators and dictionaries. On the Language Grid, the service providers can set permission to use their service for each user and limit amount of transferred data.

To give service providers more incentive to join a platform of service-oriented collective intelligence, a composite service which combines multiple services needs to be executed satisfying policies of all the service providers concerned.

Some previous works have proposed methods to find a combination of services which satisfy given requirements. They select appropriate services from a set of services which are functionally equivalent[2], [3]. However, the previous works are not enough from the following aspects because they focused on finding a combination of services which give the best QoS, not on satisfying policies of service providers.

- There is often no combination of services which satisfy policies of all service providers concerned because the number of services is limited in reality.
- Policies of service provider may include constraints on execution of a composite service, which cannot be resolved by service selection.

To solve the problems, we propose an architecture which performs a comprehensive process of service selection, adaptation and coordination. First it tries to find a combination of services which satisfy policies of service providers in the similar fashion to the previous works. But it adapts the services to given policies by changing attributes of the services if there is no service required to satisfy the policies. Moreover, using meta-level control functions for composite services, it monitors and changes execution state of composite services in order to satisfy constraints on execution.

Some dynamic adaptation methods for composite services have been proposed. Most of them adopt aspect-oriented programming (AOP)[4] or a proxy between the composite service execution engine and invoked services[5]. The method proposed in [6] uses meta-level control functions to control execution of composite service. However, these works focus on adaptation without changing model of composite services. Therefore they do not handle a comprehensive process including service selection and coordination. We call the architecture Service Supervision because it continuously monitors or manages execution of services, and controls the

978-0-7695-4126-6/10 $26.00 © 2010 IEEE
DOI 10.1109/SCC.2010.64

behavior of services.

The rest of this paper is organized as follows. In Section II, we show a scenario which describes our research problems and goals of the architecture we propose. From Section III to Section V, we explain three layers in our architecture. Each of layer is responsible for service selection, adaptation, and coordination respectively. In Section VI, we briefly discuss the scalability of the proposed architecture. After we introduce related works in Section VII, we conclude this paper in Section VIII.

II. Design Goal

In this section, we first describe a scenario of a composite service which runs on a service-oriented collective intelligence platform. Then we show an overview of the architecture we propose.

A. Scenario

We take a composite service for translation deployed on the Language Grid[1] as an example. The composite service combines a morphological analyzer, a machine translator, and technical term dictionaries. This service improves translation quality of technical documents by translating technical terms in the given sentences using the technical term dictionaries, not the machine translator.

Figure 1 shows the overview of the composite service. A square which contains a circle represents a service invocation activity. First the given sentences are divided into morphemes by the morphological analyzer. Next dictionaries find technical terms which consist of the morphemes and return the translation of the technical terms. Finally the translator translates the whole sentences.

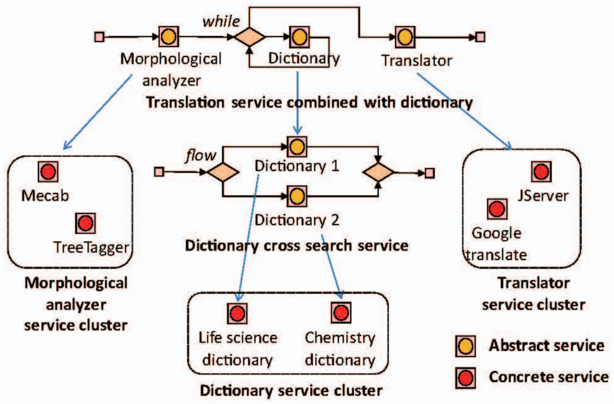


Figure 1. A composite service for translation

As shown in Fig. 1, we assume that a composite service is defined in a workflow description language such as WS-BPEL[7]. In a composite service, the constituent services define only the interface and are not bound to any endpoint. We refer to such a service as an *abstract service*. For example, on the Language Grid, an abstract service is defined for each service type such as translators and dictionaries. Endpoints for the services are determined when the composite service is invoked. A service to which an endpoint is bound is called a *concrete service*.

On the Language Grid, all the information on concrete services and composite services are managed by Language Grid Service Manager[1] shown in Fig. 2. Currently more than 70 concrete services and 12 composite services are registered and various information of the services including lisence and WSDL are managed on Language Grid Service Manager.

Service Name	Service Type	Languages (in Language Code)	Provider	Status
Google Translate	TRANSLATION	(sq<->ar), (sq<->bg), (sq<->zh-CN), (...	Google, Inc.	Run
Hamamatsu City Documents Related to Foreign Students	PARALLEL TEXT	(ja<->en), (ja<->pt), (ja<->es), (en<...	Language Grid Operation Ce...	Run
HTML Text Extractor	OTHER	(*)	Language Infrastructure Gr...	Run
ICTCLAS	MORPHOLOGICAL ANALYSIS	(zh)	NLP Group, Institute of Co...	Run
J-Server (Kyoto-U)	TRANSLATION	(ja<->en), (ja<->ko), (ja<->zh)	Ishida and Matsubara Labor...	Run

Figure 2. A list of services registered on Service Manager.

A set of concrete services which is bound to an abstract service is called a service cluster. In Fig. 1, there are two concrete services in each service cluster. In our example, we bind one of the two concrete services to the morphological analyzer and the machine translator. For the dictionary, we first bind a composite service for cross search and then bind two concrete services to abstract services in the cross search composite service. The Language Grid allows users to specify the hierarchical binding by describing the binding in SOAP header of the request message.

We show the process of execution of this composite service below. First we select concrete services which satisfy the user's request and service providers' policies. In our example, we assume that the request is translation from Japanese to English and that the user specifies the life science dictionary and the chemistry dictionary for the two abstract dictionary services.

When receiving the request, the system selects concrete services for the rest of the abstract services. Suppose it first tries Mecab, which is a morphological analyzer for Japanese, and Google Translate, which provides translation from Japanese to English. However, the combination may have the following three problems.

The first problem is constraints on a combination of services. Assume the provider of the chemistry dictionary prohibits use of its service with Google Translate. In such a case, the system needs to select JServer, which is another

[1]http://langrid.org/service_manager/

machine translator and supports Japanese-English translation.

The second problem is constraints on execution of a constituent service. The provider of the translator JServer may limits the length of the input to 1000 characters in order to reduce server load. In this case, we need to introduce an adaptation process which divides a long input into sentences before translation.

The third problem is constraints on execution of a composite service. Assume that both the life science dictionary and the chemistry dictionary are provided by the same provider and that the provider prohibits concurrent access to the two services to prevent a user from giving too much load. In this case, two dictionaries which are defined to be executed in parallel should be controlled to be executed sequentially.

B. Architecture Overview

We propose an architecture to solve the problems described in the previous scenario.

Our architecture consists of three layers which are connected to each other and solve the three types of constraints respectively. Figure 3 shows the overview of the architecture.

The roles of layers are as follows:

Selection Layer
This layer selects concrete services from service clusters to satisfy constraints on combination of services. The constraints are retrieved from service providers' policies and stored in the repository in this layer. If no combination which satisfies all constraints is found, this layer delegates control to Adaptation Layer.

Adaptation Layer
This layer adapts a constituent service in a composite service to the given constraints by changing attributes of the service. The adaptation is achieved by weaving processes before/after invocation of the target service based on AOP. More than one adaptation process can be combined by the planner in this layer. If the constraints are not satisfied in this layer, this layer delegates control to Coordination Layer.

Coordination Layer
This layer controls the order of services execution in a composite service. Using meta-level control functions for composite services proposed in [6], this layer forces execution of a composite service to follow a protocol defined by the given choreography in WS-CDL. If this layer cannot satisfy the constraints, execution of the composite service fails.

Based on this architecture, execution of a composite service proceeds as described below.

First a user sends a request to the composite service execution engine. The user can specify concrete services which are bound to abstract services if he/she needs.

The composite service execution engine invokes the service selection process before invocation of each service. It continuously needs to confirm that the constraints are satisfied because the state of services is changing and the output of a service may violate the constraints.

Selection Layer retrieves information from Service Repository and Service Provider Directory and stores the information in it. According to the information, it also finds a combination of services which satisfy constraints in response to the request from composite service execution engine.

Adaptation Layer stores various adaptation processes, such as dividing an input / merging outputs, and adding an authentication process. It decides which adaptation processes should be applied and the order of adaptation processes. After it finds a set of adaptation processes to be applied, the adaptation processes are registered to AOP Manager. AOP Manager is an extension of composite service execution engine. It monitors execution of a composite service and "weaves" adaptation process.

Coordination Layer has a repository of choreography, which is created from the model of composite services and service providers' policies on execution. After performing an execution control function of meta-level controller, it immediately finishes and returns control to the composite service execution engine.

III. Selection Layer

Selection Layer finds a combination of concrete services which satisfy constraints by service providers' policies.

In this paper, we formalize selection of services as constraint satisfaction problem (CSP). Given service profiles, a set of available concrete services and policies of service providers, the selection of services are defined as follows:

Variable
A variable represents an abstract service in a composite service. If a composite service invokes other composite services, the set of variables contains variables of abstract services in all composite services.

Domain
A domain represents a set of concrete services in a service cluster. A domain is described as $D = \{x_1, x_2, ...\}$, where $x_1, x_2, ...$ are concrete services. Attributes of concrete services are described as $x_i.attr_1, x_i.attr_2, ...$.

Constraints
Constraints represent policies of service providers and requirements of a user. Constraints are described as predicates based on attributes of concrete services or input/output of services.

Domains are retrieved from Service Repository according to the service type of each abstract service. Constraints are obtained by transforming service providers' information.

In the initial state, the set of variable contains variables which correspond to all abstract services. However, if it is ensured that a service will not be execute due to a conditional branch during execution, the variable corresponding to the service is removed.

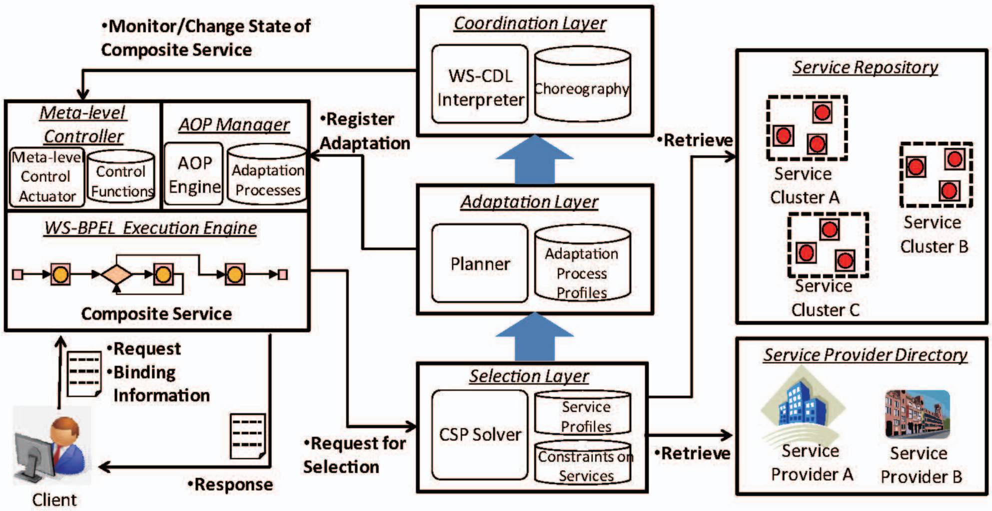


Figure 3. Layers of Service Supervision system.

We show an example of the definition for the translation composite service described in the previous section. As the composite services consist of a morphological analyzer, two dictionaries and a translator, variables X_{ma}, X_{dic1}, X_{dic2}, X_{trans} are defined. The variables have the corresponding domains as follows:

$$\begin{aligned}
D_{ma} &= \{x_{ma_mecab}, x_{ma_treetagger}\} \\
D_{dic1} &= \{x_{dic_life}, x_{dic_chem}\} \\
D_{dic2} &= \{x_{dic_life}, x_{dic_chem}\} \\
D_{trans} &= \{x_{trans_google}, x_{trans_jserver}\}
\end{aligned}$$

We also show some of attributes of concrete services below. The following definition represents that a morphological analyzer Mecab accepts only Japanese, and TreeTagger accepts English, German, French, Italian and so on. For a translator JServer, the supported language pairs and limitation of input sentences are defined.

$x_{ma\_mecab}.sourceLang = \{Japanese\}$
$x_{ma\_treetagger}.sourceLang =$
$\quad \{English, French, German, Italian, ...\}$
$x_{trans\_jserver}.langPair =$
$\quad \{(Japanese\ to\ English)(English\ to\ Japanese)...\}$
$x_{trans\_jserver}.maxInputLength = 1000$ ...

The constraints are shown below.

$C = \{$
$\quad C1 : X_{trans}.maxInputLength >$
$\qquad length(Request.input),$
$\quad C2 : exclusive(x_{trans\_goole}, x_{dic\_chem}),$
$\quad C3 : parallel\_prohibited(x_{dic\_life}, x_{dic\_chem})$ ... $\}$

$C1$ represents requirements based on the process of the composite service and shows that the limitation of input length of a translator must be longer than input sentences. $C2$ prohibits the combination of GoogleTranslate and the life science dictionary. $C3$ prohibits parallel execution of the chemistry dictionary and the chemistry dictionary.

In our example, we assume the following request from a user. The request consists of sentences to be translated, the source language, and the target language.

$Request = \{$
$\quad input$:"Sentence to be translated"
$\qquad$ (actually 1500 characters in Japanese),
$\quad source\_language : Japanese,$
$\quad target\_language : English\}$

User can specify the concrete services bound to any abstract services in a composite service when he/she invokes the composite service. For the abstract services which the user does not specify concrete services bound to, the process in Selection Layer selects concrete services. A user usually specifies concrete services when he/she knows how attributes of concrete services affect the result. Otherwise, a user delegates the selection to the system.

In our example, the user specifies the life science dictionary and the chemistry dictionary for two dictionaries in order to translate sentences in the area of biochemistry. On the other hand, the user does not specify concrete services for the morphological analyzer and the translator because he/she does not know which service can get a better result.

The CSP which formalizes the above conditions has no solution. Therefore it is impossible to satisfy the user's requirements and service providers' policies using the available services. The combinations which have the least number of violations are as follows. The former violates $C1$ and $C2$, and the latter violates $C2$, $C3$.

$$(X_{ma}, X_{dic1}, X_{dic2}, X_{trans}) =$$
$$(x_{ma\_mecab}, x_{dic\_life}, x_{dic\_chem}, x_{trans\_jserver})$$
$$(x_{ma\_mecab}, x_{dic\_life}, x_{dic\_chem}, x_{trans\_google})\}$$

The attributes of services which cause violation are changed in Adaptation Layer in order to resolve the violation. To determine which service should be adapted, we extended CSP based on the ideas of Open CSP[8] and Partial CSP[9].

In Open CSP, a new value for a domain is obtained when any combination of existing values cannot satisfy constraints. The paper shows that the domain of a variable which is located in the deepest leaf of a search tree of backtrack search certainly needs to be extended to resolve violations. Therefore, we need to adapt one of concrete services which correspond to such a variable.

Moreover, to determine which concrete service should be adapted, we also need to know a combination of services which gives the least violation count. Therefore we applied the idea of Partial CSP, which finds a solution which gives the least violation when the problem is over-constrained.

Figure 4 shows the algorithm which applies the ideas to the problem of finding a combination of services and determining a service to be adapted.

The algorithm begins with depth-first search (line 1-22). The index of the deepest variable in the search tree is recorded as k. During backtrack search, the least violation count and the combination of services which gives the count are also record as M and $currentSelection$ respectively.

Then it extends the domain D_k by giving k and $currentSelection$ to Adaptation Layer and searches for solution again (line 23-30).

```
function: searchCombination(X, D, C)
Inputs:
  X: Variables ({X1, ..., Xn}),  D: Domains ({D1, ..., Dn}),
  C: Constraints

 1:  i←1, k←1, M←∞
 2:  while(i > 0)
 3:    if all values in Di are checked
 4:      reset xi,  i←i-1
 5:    else
 6:      xi←next value of Di
 7:      if ({x1, ..., xi}) satisfies C
 8:        k← max(k, i+1)
 9:      end if
10:      i←i+1
11:      if i > n
12:        if ViolationCount({x1, ..., xn})  = 0
13:          return {x1, ..., xn}
14:        end if
15:        if N > ViolationCount({x1, ..., xn})
16:          N← ViolationCount({x1, ..., xn})
17:          currentSelection ←{x1, ..., xn}
18:        end if
19:        i←i+1
20:      end if
21:    end if
22:  end while
23:  xk'← findAdaptation(k, currentSelection)
24:  if xk' is failure
25:    return failure
26:  end if
27:  Dk ← Dk ∪ {xk'}
28:  Change order of variables (Move Xk
29:    to first  as X1)
30:  return searchCombination(X, D, C);
```

Figure 4. Algorithm for finding a combination of services

IV. Adaptation Layer

I If Selection Layer cannot find a combination of services which satisfy all constraints, Adaptation Layer adapts a service to constraints by changing attributes of services. The adapted service can be considered as a new service in Selection Layer.

To find appropriate adaptation processes, profiles of adaptation processes are stored in Adaptation Layer. Figure 5 shows an example of the profile of an adaptation process. The adaptation process first divides an input string into sentences before a service is invoked, and then invokes the target service for each sentence. Finally it merges the results after the target services finishes. This process is used to adapt a translation service which has the limit of input length.

An adaptation process is identified by the name described in "adaptation" tag. Preconditions and effects are described in a profile. In the description of preconditions and effects, "Request" represents a request given by the user and "Service" represents a service to which this adaptation process is applied.

The profile in Fig. 5 shows that the input must consist of more than one sentence and the longest sentence must be shorter than the limit of input length of the target service. This also shows the adaptation process removes the limit of input length if applied.

As shown in Fig. 3, profiles of adaptation processes are stored in Adaptation Layer, but the implementation is in AOP manager, which is an extension of the composite service execution engine. It is difficult for the standard framework to change the model of a composite service and to deploy the new model if adaptation processes applied are frequently changed. This is the reason the adaptation is realized by using AOP techniques.

Various adaptation methods for composite services using AOP have been proposed in previous works(e.g. [4], [5]). One of the most flexible methods[5] weaves a process described in WS-BPEL into the target composite service. We assume we adopt this method for our architecture and show the implementation of adaptation process in Fig. 6.

The adaptation process divides an input and merges the

```
<adaptation name="DivideAndLoop">
  <precondition>
    <expression language="javascript">
      numOfSentences(Request.input) > 1
      && Service.maxInputLength(
        length(getLongestSentence (
              Request.input)))
    </expression>
  </precondition>
  <effect>
    <expression language="javascript">
      Service.maxInputLength = INF
    </expression>
  </effect >
</adaptation>
```

Figure 5. Profile of an adaptation process.

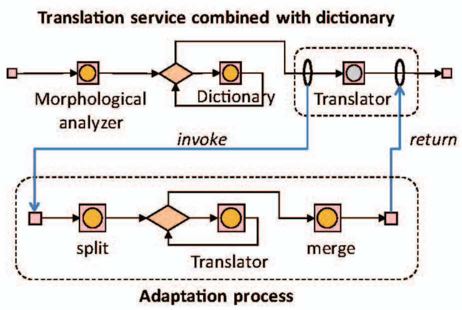


Figure 6. Adaptation process for division and merging

```
function: findAdaptation(k, currentSelection)
Inputs:
 k: Index of target service
 initialSelection: A combination of services which
   give the least violation count without adaptation

 1:  currentSelection, tempSelection ← initialSelection
 2:  leastViolation ← ViolationCount(currentSelection)
 3:  for MAX times
 4:   PossibleAdaptation = {a | a ∈ Set of adaptation
 5:    processes and all preconditions of a are satisfied}
 6:   for each a in PossibleAdaptation
 7:     nextSelection ← Apply a to k th service of
 8:                        currentSelection
 9:     if(leastViolation > ViolationCount(nextSelection))
10:       tempSelection ← nextSelection
11:     end if
12:   end for
13:   if tempSelection = initialSelection
14:     coordinateOrder()
15:     return failure
16:   end if
17:   if tempSelection = currentSelection
18:     return k th service of currentSelection
19:   end if
20:   currentSelection ← tempSelection
    end for
```

Figure 7. Search algorithm for adaption.

output before/after the target service is invoked. The invocation of target service is located in loop in the adaptation process. The invocation which is originally defined in a composite service is skipped.

Using adaptation method described above, more than one adaptation process can be applied to one target service. We search for a sequence of adaptation processes by hill climbing taking the count of violation as the evaluation value. Figure 7 shows the algorithm for selecting adaptation processes. This algorithm is called by the algorithm in Fig. 4.

The algorithm takes the index of the target service and a combination of services which gives the least count of violation as an input. It tries to apply various adaptation processes in order to find a sequence of adaptation processes which can reduce the count of violations. If this algorithm cannot reduce the violation count, it invokes Coordination Layer and returns failure to Selection Layer (line 14-15). The service to which some adaptation processes are applied is returned to Selection Layer and added to the domain in CSP (line 18).

We refer to $x_{trans\_jserver}$ which the adaptation process in Fig. 7 is applied to as $a_{div\_merge}(x_{trans\_jserver})$. The domain of translation services is extended as follows:

$$D_{trans} = \{x_{trans\_google}, x_{trans\_jserver}, \\ a_{div\_merge}(x_{trans\_jserver})\}$$

We can obtain the following combination whose count of violations is reduced to 1 with the adapted service.

$$(X_{ma}, X_{dic1}, X_{dic2}, X_{trans}) = \{ \\ (x_{ma\_mecab}, x_{dic\_life}, x_{dic\_chem}, a_{div\_merge}(x_{trans\_jserver}))$$

V. Coordination Layer

Some service providers have policies about execution, e.g. constraints on order of service execution. Such constraints cannot be solved by adaptation of a service. Therefore we control the order of service execution in Coordination Layer.

The standard framework for composite services such as WS-BPEL does not provide a method for changing order of service execution at runtime. This is the reason we adopt meta-level control functions for composite services proposed in [6]. Table I shows some of the meta-level control functions. Using the functions, we can obtain and change state of a running composite service. The functions are implemented by modifying the composite service execution engine.

Table I
META-LEVEL CONTROL FUNCTIONS.

API	Effect
suspend	Suspend an activity whose state is ‘Ready’. The state of the suspended activity will be changed to ‘Suspended’
resume	Resume an activity whose state is ‘Suspended’. The state of the suspended activity will be changed to ‘Running’.
getProcessState	Get the current state of all activities in a process of the composite service.
terminateProcess	Terminate the process of the composite service.

The constraints on order of service execution can be defined by a choreography description language such as WS-CDL. We can check if an order of service execution satisfies the constraint or not using WS-CDL interpreter. Figure 8 shows the procedure for checking the order of service execution.

```
Function coordinateOrder()

choreography ← Protocol of service execution
  based on model of composite service and
  policies of service providers
currentState ← getProcessState()
queue ← getReadyActivity(currentState)
        ⊔ getSuspendedActivity(currentState)
for each act in queue
  if accept(choreography, currentState, act)
    resume(act)
    return
  else
    suspend(act)
  end if
end for
if getRunningActivity(currentState) is empty
  terminateProcess()
end if
```

Figure 8. Procedure for coordinating services.

We define states for an activity in a composite service. ‘Ready’ represents that service is ready to be executed. ‘Suspended’ represents that the service is suspended after it once becomes ‘Ready’. ‘Running’ represents that the service is being executed and ‘Finished’ represents that the execution of the service finished. The composite service execution engine sets the state of activities as execution of a composite service proceeds.

The algorithm in Fig. 8 first gets states of all activities in the composite service. Next it puts activities whose states are ‘Ready’ or ‘Suspended’ into a queue. Then it checks each activity can be accepted by the given choreography using the WS-CDL interpreter. If an activity is accepted, the activity is invoked. Otherwise, the state of the activity is changed to/keeps being ‘Suspended’.

If all activities in the queue are in ‘Suspended’ and there is no ‘Running’ activity, execution of the composite service is terminated because no activity can be executed and the state will not be changed.

In our example, concurrent access to the chemistry dictionary and the life science dictionary is prohibited. However, the composite service does not define the order of execution of the two services. Therefore the system monitors the state of execution and prohibits invocation of one of the dictionary services while another one is running.

VI. DISCUSSION

In this section, we discuss the characteristics and the operation of the architecture proposed in this paper.

The algorithm shown in Fig. 4 for Service Selection Layer performs simple depth first search for the simplicity of description. This leads the computational complexity of the process to $O(n^b)$ where n is the number of abstract services and b is the size of service clusters. This can be unacceptable for some applications.

In [9], however, more efficient algorithms for Partial CSP such as extensions of branch and bound method and backmarking method are proposed. The ideas of these methods can be easily applied to the algorithm 4. For example, we can introduce branch and bound method by back-tracking when the current number of violations exceeds the least recorded number of violations. This can generally make this algorithm more efficient.

The procedure of seaching for adaptaiton processes proposed in this paper is not complete. First a combination of concrete services which gives the least violation count cannot always reach the best solution. Other combination may give less violation count by applying adaptation processes. Secondly, search by hill-climbing can reach local minimum.

However, the proposed architecture and ideas of Open CSP and Partial CSP are not limited to the search methods we showed in this paper. We can easily introduce complete algorithms by revising algorithms in Fig. 4 and Fig. 7 in order to keep completeness. When adopting the architecture, we need to choose appropriate methods depending on the required realtime property of applications.

VII. RELATED WORKS

In the architecture proposed in this work, service selection, adaptation and coordination are performed during execution of composite service.

Many previous works have proposed methods for service selection. For example, the method proposed in [2] focuses on finding a combination of services which gives the best QoS. The method proposed in [3] select services considering interfaces of services in addition to QoS.

These works assume that vast amount of services are stored in a service cluster. In reality, however, the number of services which have equivalent functions is limited. That is the reason the previous works often cannot find a combination of services. Moreover, to handle the policies of service providers, we need not only finding a static combination of services but also dynamic adaptation and meta-level control of composite services.

Also in the area of dynamic adaptation, there have been some previous works. Most of the works can be classified into three types: weaving a new process based on AOP (Aspect-oriented Programming), using a proxy to monitor/change messages exchanged between a composite service and invoked services, and transforming the model of a composite service based on definition of additional processes.

AO4BPEL[5] is one of the framework for realizing AOP of composite services. It allows a user to define a pointcut in a WS-BPEL process and weave a process described in WS-BPEL as an advice. This can add processes for adaptation without changing the model of a composite service.

For service-oriented collective intelligence, however, it is required to satisfy various policies of service providers. This needs a comprehensive process of service selection, adaptation and coordination. AOP is suitable for adaptation as described in the previous section, but it is not flexible enough to coordination such as controlling order of service execution.

The work proposed in [4] adopts a framework using a proxy. It checks if messages exchanged among a composite services and the constituent services satisfy the given conditions when the composite service execution engine invokes the constituent services. If any of conditions is not satisfied, it performs some recovering processes, retries invocation, or changes the service to an alternative. But this focuses on adaptation of single service and does not deal with policies of all service provider concerned.

The methods which transform the model of composite services such as [10] does not need to modify the composite service execution engine. It suits to adding exception handling processes, but the composite service needs to be deployed if the adaptation process changed.

Meta-level control functions for Web services are proposed in [6]. The work provides more flexible controls than other works described above. But the flexibility often gives too much load on operators of the platform. The architecture proposed in this work partially adopted the work focusing on satisfying constraints on execution.

VIII. Conclusion

In this paper, we proposed an architecture which performs a comprehensive process for service selection, adaptation and coordination. The architecture is designed to satisfy policies of service providers.

The contributions of this paper are as follows:

- We proposed a method for service selection and adaptation in an integrated fashion by formalizing the problem as an extension of constraint satisfaction problem.
- We showed a framework which controls order of service execution using meta-level control functions for composite service.

To realize service-oriented collective intelligence, it is required to satisfy various policies of service providers. This paper is the first work which handles a comprehensive process of service selection, adaptation and coordination from the aspect.

In future work, we will apply the proposed architecture to a real environment, such as the Language Grid, where many service providers have joined.

Acknowledgment

This work was supported by a Grant-in-Aid for Scientific Research (A) (21240014, 2009-2011) and a Grant-in-Aid for Young Scientists (Start-up) (21800095) from Japan Society for the Promotion of Science (JSPS).

References

[1] T. Ishida, "Language Grid: An infrastructure for intercultural collaboration," in *IEEE/IPSJ Symposium on Applications and the Internet (SAINT-06)*, 2006, pp. 96–100.

[2] L. Zeng, B. Benatallah, A. H. Ngu, M. Dumas, J. Kalagnanam, and H. Chang, "Qos-aware middleware for web services composition," *IEEE Transactions on Software Engineering*, vol. 30, pp. 311–327, 2004.

[3] A. B. Hassine, S. Matsubara, and T. Ishida, "A constraint-based approach to horizontal web service composition," in *the 5th International Semantic Web Conference (ISWC 2006)*, vol. 4273, 2006, pp. 130–143.

[4] L. Baresi, S. Guinea, and P. Plebani, "Policies and aspects for the supervision of BPEL processes," in *the 19th International Conference on Advanced Information Systems Engineering (CaiSE07)*, 2007, pp. 340–354.

[5] A. Charfi and M. Mezini, "AO4BPEL: An aspect-oriented extension to bpel," *World Wide Web*, vol. 10, no. 3, pp. 309–344, 2007.

[6] M. Tanaka, T. Ishida, Y. Murakami, and S. Morimoto, "Service supervision: Coordinating web services in open environment," in *IEEE International Conference on Web Services (ICWS-09)*, 2009, pp. 238–245.

[7] "Business process execution language for web services (BPEL), version 1.1." http://www.ibm.com/developerworks/library/ws-bpel/, 2003.

[8] B. Faltings and S. Macho-Gonzalez, "Open constraint programming," *Artificial Intelligence*, vol. 161, no. 1-2, pp. 181–208, 2005.

[9] E. C. Freuder and R. J. Wallace, "Partial constraint satisfaction," *Artificial Intelligence*, vol. 58, no. 1-3, pp. 21–70, 1992.

[10] A. Mosincat and W. Binder, "Transparent runtime adaptability for bpel processes," in *the 6th International Conference on Service Oriented Computing (ICSOC 08)*, 2008, pp. 241–255.

2011 IEEE International Conference on Services Computing

Reputation-Based Selection of Language Services

Shinsuke Goto
Department of Social Informatics, Kyoto University, Kyoto 6068501, Japan
s-goto@ai.soc.i.kyoto-u.ac.jp

Yohei Murakami
National Institute of Information and Communications Technology (NICT), Kyoto, 6190289, Japan
yohei@nict.go.jp

Toru Ishida
Department of Social Informatics, Kyoto University, Kyoto 6068501, Japan
ishida@i.kyoto-u.ac.jp

***Abstract*—Quality of Service (QoS) can be used to select desired services from among those offering the equivalent function. In language services such as machine translation, one of the QoS metrics is translation accuracy. However, the problems are that evaluating the translation accuracy is too expensive, that the translation accuracy varies with the difficulty of the task, and that the usefulness of the translation to the user depends on the abilities of the user. In this paper, we propose a framework that selects a useful service for a specific user and task by using reputation information of users, which can be obtained at low cost. First, hypothetical reasoning is used to estimate the partial order relation between the accuracy of the language services, the language ability of the users, and the difficulty of the tasks. Second, deductive reasoning is applied to recommend useful services given the user and the task. We propose a reputation-based language service selection system that combines a partial order acquisition system with a service selection system.**

***Keywords*-service selection, QoS, hypothetical reasoning, reputation information**

I. Introduction

In services computing, a key user demand is selecting one of the available services from among those with equivalent functionality. If the right service can be found automatically, composite services can be developed more easily. To date, Quality of Service (QoS), which is a quantitative measure of service evaluation, is the most commonly applied technique for service selection. Language Grid [1] is a multilingual service infrastructure based on services computing technologies. It has various language services such as machine translation services and multilingual dictionary services. For language services, translation accuracy can be used as a QoS metric.

Using humans to evaluate translation accuracy is not feasible. Also, the accuracy of the translation varies depending on the task. For example, machine translation trained by a corpus in one domain has lower accuracy in translating out-of-domain texts than in-domain texts [2]. Additionally, users with different language abilities have different evaluation scores for the same machine translation. There is a negative correlation between the user's TOEIC test score and the user's evaluation score of English-Japanese machine translation [3]. These facts make it difficult to select the most useful service for a specific user and task.

To address this problem, this paper proposes a language service selection method based on reputation information. User reputations can be obtained more easily and at lower cost than human-rated translation accuracy. We assume that reputations involve only the user, task, and service. Moreover, we presume that the accuracy of the language service, the language ability of the user, and task difficulty have partial order relations. If user reputations and the partial order relations are sufficient, useful services for a specific user and task can be inferred by deductive reasoning. However, a user can't input the order relation between users, services, or tasks. Reputation information itself is not capable of recommending services to the user.

Our solution to these problems is to propose two methods: one obtains the order relations from reputation information, and the other is to select the service using them. These methods differ from the service selection by general QoS in that they don't use numeric values. To realize these methods, we faced several issues.

Order Relations Acquisition
: The order relations cannot be determined from just reputation information. Therefore, a formalization method to acquire the order relations from reputation information is needed.

Integration for Service Selection Platform
: In order to construct the service selection platform, we need to integrate an order relation acquisition engine, based on reputation information, and the execution engine, which invokes the service.

II. Service Selection

Various methods have been proposed for service selection. Among them, the most popular approach is QoS-based service selection. This section details QoS-based service selection and the extension of user-centered QoS.

QoS was originally developed in the field of computer networking. It employs numerical values for service evaluation. Zhang [4] enumerated four QoS standards for web services: security, transaction, reliability, and lifetime. Examples of applying these objective metrics to service selection are given in [5] [6] [7].

978-0-7695-4462-5/11 $26.00 © 2011 IEEE
DOI 10.1109/SCC.2011.111

Zeng et al. [5], proposed the basic QoS aggregation function. In [5], each QoS metric from the service provider is normalized, and aggregated for service selection.

Xu et al. [6] proposed the QoS metric called Reputation Score. It is combined with objective QoS metrics, and the result is used for service selection. Our study is related to [6] in that QoS is evaluated by the reputations of the users.

Note that collaborative filtering and social filtering are important methods to utilize the reputation information of other users. Shao et al. [8] proposed QoS-based service selection via collaborative filtering. Shao et al. finds user(s) with similar QoS evaluation from records when the actual QoS reputations of the users differ from that provided by the service registry. The prediction is made for the user.

The above works didn't mention the QoS metric defined by the users. This paper proposes a service selection model that can deal with the QoS of language services dependent on the user and task.

User-centered QoS [9] results from making the service evaluation depend on the user. Bramantoro et al. [9] proposed user-centered QoS, and showed a method for service selection in a multilingual chat system. The main advantage of user-centered QoS is its ability to include QoS metrics dependent on the domain of the service or the preference of the user. In [9], the accuracy of machine translation, and the foreign language ability of the user are regarded as the QoS factors for machine translation services. This is the basis of our work.

Our approach is innovative in that we tackle the problem of user-centered QoS metrics when they are not measured. The proposed service selection method is derived from reputation information by utilizing hypothetical reasoning.

III. Reputation-Based Selection

A. Overview

We describe the definitions necessary for service selection in this section. We assume that only the service, user, task impact the judgment of usefulness. Below we detail the elements that describe service selection:

Services
: Koehn [2] uses accuracy to evaluate machine translation services. The more accurate a service is, the more often the user will judge the service as useful.

Users
: Each user has some level of foreign language ability. He/she compares his/her own ability to the accuracy of the service to judge service usefulness. The lower the user's ability, the more often the user judges a service as useful.

Tasks
: Tasks represent the purpose behind the use of the translation service, and each task has a level of difficulty. The easier the task, the more often the user judges a service as useful.

Order Relations
: From the definition of services, users, and tasks, we can consider the partial order relation between the translation accuracy of the services, the foreign language ability of the users, and task difficulty. An example of an accuracy relation is "Translation Service A is more accurate than Translation service B"

Reputations
: The reputation is judged useful or useless by the triplet (service, user, task).

From these elements, we can consider the example of the service selection using reputation information and order relations. If a certain service is useful for a specific user and a task, a more accurate service is also useful for the same user and task. For such service selection to be available, this paper proposes the partial order relation acquisition using hypothetical reasoning.

Here is a concrete example. There are three Japanese-to-English translation services: Translution, Google Translate, and J-Server. The user Alice is looking for a useful translation service for translating a Japanese news article. Table I lists the reputation information known to the system. We acquire the order relations and select the service based on these reputations.

Table I
Reputation Information

No	User	Service	Task	Reputation
1	Alice	Google Translate	Chat	Useful
2	Alice	Translution	Chat	Useful
3	Bob	Translution	Chat	Useless
4	Bob	Google Translate	Chat	Useful
5	Carol	Google Translate	Chat	Useless
6	Carol	Translution	Chat	Useless
7	Carol	J-Server	News	Useful

Figure 1 is the concept image of the two systems yielding reputation-based service selection. Hypotheses acquisition obtains a set of consistent order relations from the reputation information of previous users. Its algorithm is based on hypothetical reasoning, in which we regard relation orders are hypotheses. When executing hypotheses acquisition, we assume all reputations are right. On the other hand, service selection receives the user's query, and then offers useful services to the user. It uses deductive reasoning to evaluate all services from the relation orders and reputations.

B. Hypotheses Acquisition from Reputation

Following the definitions in Section III-A, we propose a method to acquire order relations between the ability of the users, the difficulty of the tasks, and the accuracy of the language services. In this study, hypothetical reasoning is used as the basis of a partial order relation acquisition system. Hypothetical reasoning is a well-known inference

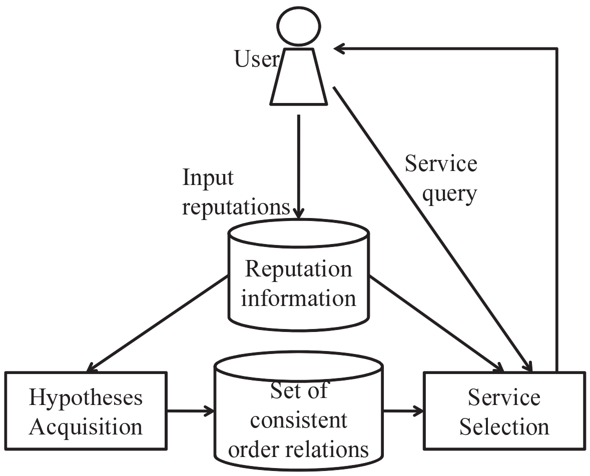


Figure 1. Concept of Reputation-Based Service Selection

method. It tries to prove an observation from background knowledge and hypotheses, and if the observation is proved, the hypotheses are regarded as right [10]. Hypothetical reasoning is formulated as the elements below:

- Σ The set of background knowledge, which is always valid
- H The set of the hypotheses which may not be true
- O The set of the observations

The schema of hypothetical reasoning is as follows. First, hypothetical reasoning tries to prove O from Σ by deduction. If O can't be proved from just Σ, hypothetical reasoning finds $H' \subset H$ satisfying the following condition.

$$\begin{cases} H' \cup \Sigma \vdash O \\ H' \cup \Sigma \ is\ consistent \end{cases}$$

The expression above means O can be proved from H' and Σ. The expression below indicates that $H' \cup \Sigma$ doesn't involve a contradiction. Namely, when only background knowledge can't prove O, hypothetical reasoning extracts consistent H', combines H' and Σ, and proves O.

By applying reputation information to hypothetical reasoning, (Σ, H, O) can be defined as follows:

- Σ Inference rules, integrity constraints, and domain dependent knowledge on service reputations
- H The order relations between the ability of the users, the difficulty of the tasks, and the accuracy of the
- O The reputation information obtained by questionnaires

Background knowledge Σ includes 6 inference rules. They are listed on Table II. The ground for these rules consists of the definitions of users, tasks, and services in Section III-A. We assume that the language ability of users, the accuracy of the services, and the difficulty of the tasks form partially ordered sets. Therefore, these inference rules indicate that the lower the language ability of the user is, the more accurate the service is, and the easier the task is, the more useful is the reputation. For example, inference rule 1 states that "The service that is judged useful for the same task by a user who has higher ability than him/her is useful for him/her". The left part of the condition clause means reputation, and the right part means order relation, which is a hypothesis. $Useful(user_i, service_j, task_k)$ means $service_j$ was judged useful by $user_i$ for $task_k$. The order relation clause means the first argument is higher/lower than the second argument. For example, $LowerAbility$ $(user1, user2)$ represents $user1$ has lower foreign language ability than $user2$. In addition, background knowledge includes an integrity constraint about the order relation. Moreover, the rule for background knowledge is that if $HigherAccuracy(service1, service2) \cap HigherAccuracy$ $(service2, service3)$ is true, $HigherAccuracy(service1,$ $service3)$ is also true due to transitivity.

Also, there are integrity constraints:

- $LowerAbility(user1, user2) \cap$ $LowerAbility(user2, user1) \rightarrow Conflict$
- $HigherAccuracy(service1, service2) \cap$ $HigherAccuracy(service2, service1) \rightarrow Conflict$
- $EasierTask(task1, task2) \cap$ $EasierTask(task2, task1) \rightarrow Conflict$
- $Useful(user, service, task) \cap$ $Useless(user, service, task) \rightarrow Conflict$

The first to third constraints are derived from irreflexivity of the order relation. The last constraint means the reputation of specific triplet must be either useful or useless.

The above constitutes the framework used to apply hypothetical reasoning to reputation information. However, there is a problem when using this framework directly: When proving one reputation, other reputations are not included in the set of knowledge. This is a problem because no inference rule can be applied if there is no reputation information in background knowledge.

We propose an approach toward this problem: when proving one reputation, all the other reputations are taken as domain dependent knowledge. This process is repeated until all reputations are proved. Then, the sets of hypotheses are merged to yield a set of hypotheses that can prove each reputation. This approach is based on the premise of the correctness of the reputations.

Algorithm 1 shows the hypothesis acquisition algorithm for service selection. We assume all reputation information is correct when acquiring the hypotheses. Also, we assume that only background knowledge and the set of hypotheses can be used to prove reputations. This is based on the closed world assumption, so if a predicate is not proved to be true, it set as false [11]. In algorithm 1, HypotheticalReasoning (K, H, O) in line 12 is the body of hypothetical reasoning; it outputs the set of proving hypotheses $P = \{H'_1, ..., H'_m\}, H'_i \subset H$ from background knowledge K, the set of hypotheses H,

Table II
INFERENCE RULES

No	Condition	Consequent	Type
1	$Useful(user1, service, task) \cap LowerAbility(user2, user1)$	$Useful(user2, service, task)$	Analogy from another user
2	$Useless(user1, service, task) \cap LowerAbility(user1, user2)$	$Useless(user2, service, task)$	Analogy from another user
3	$Useful(user, service1, task) \cap HigherAccuracy(service2, service1)$	$Useful(user, service2, task)$	Analogy from accuracy
4	$Useless(user, service1, task) \cap HigherAccuracy(service1, service2)$	$Useless(user, service2, task)$	Analogy from accuracy
5	$Useful(user, service, task1) \cap EasierTask(task2, task1)$	$Useful(user, service, task2)$	Analogy from difficulty
6	$Useful(user, service, task1) \cap EasierTask(task1, task2)$	$Useless(user, service, task2)$	Analogy from difficulty

Algorithm 1 AcquireHypotheses(K, H, R)

K /* The set of background knowledge */
H /* The set of hypotheses */
$R = \{r_1, ..., r_n\}$ /* The list of reputations */
CH /* The list of the set of hypotheses which is consistent and can prove each reputations */
P_i /* The list of the set of hypotheses which can prove reputation r_i */
$p_i \in P_i$ /* A subset of the H which can prove r_i */
$IC \subset K$ /* Integrity Constraint */
A /* The direct product of $P_1, ..., P_n$. A means the list of the set of hypotheses which can prove each reputations */
$a_k = \{p_1, ..., p_n\} \in A$ /* N-tuple which can prove each reputations */
PH /* The set of hypotheses which can prove each reputation */
for all r_i **in** R **do**
  $P_i \leftarrow \text{HypotheticalReasoning}(K \cup (R - \{r_i\}), H, r_i)$
end for
$A \leftarrow \prod_{i=1}^{n} P_i$
$CH \leftarrow \emptyset$
for all a_k **in** A **do**
  $PH \leftarrow \emptyset$
  for all p_i **in** a_k **do**
    $PH \leftarrow PH \cup p_i$
  end for
  if CheckConsistency$(K \cup R, PH, IC)$ **then**
    $CH \leftarrow CH \cup \{PH\}$
  end if
end for
return CH

Algorithm 2 CheckConsistency (K, H, IC)

1: **if** $K \cup H \vdash IC$ **then**
2:   **return** $false$
3: **end if**
4: **return** $true$

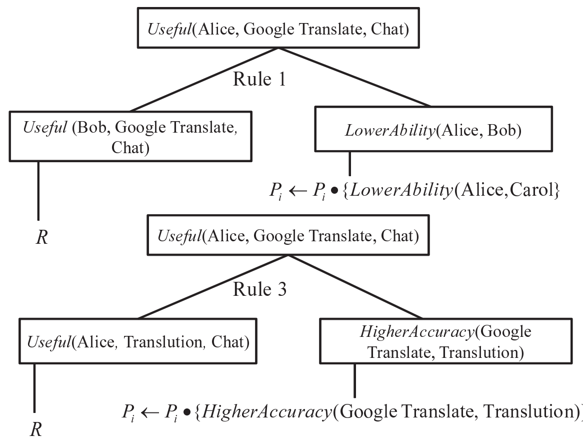


Figure 2. The Proof Tree of $Useful$(Alice, Google Translation, Chat)

and observation O. Each H'_i is consistent and can prove O. In addition, $\vdash$ in line 1 of CheckConsistency means the left side can prove the right side.

We explain the process of acquiring the set of hypotheses using the reputations in the example of Table I. The explanation of the top reputation in Table I, $Useful$(Alice, Google Translate, Chat) is shown in Figure 2. The goal is to prove the reputation $Useful$(Alice, Google Translate, Chat). First, by using inference rule 1, hypothesis acquisition tries to prove $Useful$(Bob, Google Translate, Chat) $\cap$ $LowerAbility$(Alice, Bob). Since $LowerAbility$(Alice, Bob) doesn't cause conflicts with background knowledge, the set of hypothesis $LowerAbility$(Alice, Bob) is added to a answer set which can prove $Useful$(Alice, Google Translate, Chat). This means if Alice has a lower language ability than Bob, the reputation $Useful$(Alice, Google Translate, Chat) is proved because $Useful$(Bob, Google Translate, Chat) is in the domain-knowledge. In the next step, Next, using inference rule 3, $Useful$(Alice, Google Translate,

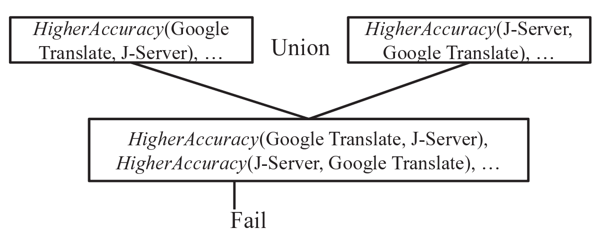


Figure 3. The Inference Tree of Checking Consistency

Chat) is expanded to $Useful$(Alice, Translution, Chat) $\cap$ $HigherAccuracy$(Google Translate,Translution). Similarly, the set $\{HigherAccuracy$(Google Translate,Translution) $\}$ is added. Therefore, HypotheticalReasoning returns $\{$ $\{LowerAbility$(Alice, Bob)$\}$, $\{HigherAccuracy$(Google Translate,Translution)$\}\}$.

Figure 3 represents the partial process of checking if a set of hypotheses is consistent. This is the process of the line 16 to line 24 of AcquireHypotheses. First, there are two set of hypotheses that can prove each reputation: one includes $HigherAccuracy$(Google Translate, J-Server), the other includes $HigherAccuracy$(Google Translate, J-Server). These hypotheses are merged, and becomes one set of reputation which includes both $HigherAccuracy$(Google Translate, J-Server) and $HigherAccuracy$(Google Translate, J-Server). However, these two order relations conflict because of the integrity constraint. Therefore, this set of hypotheses cannnot be nominated for the result of AcquireHypotheses.

Below is a concrete process of AcquireHypothesis in Table I. First, HypotheticalReasoning obtains the sets of hypotheses $P_1, ..., P_7$ that can prove each reputation. P_i means the set of the sets of the hypotheses that can prove reputation i in Table I. It outputs P_1 = $\{\{LowerAbility$(Alice, Bob)$\}$, $\{$ $HigherAccuracy$(Google Translate, Translution)$\}$ as the set of hypotheses that can prove $Useful$(Alice, Translution, Chat) when it is the argument of O. Also, HypotheticalReasoning tries to prove P_2, which represents $Useful$(Alice, Translution, Chat). However, P_2 can't be proved by any hypotheses. So P_2 becomes the empty set $\{\}$. Similarly, $P_1, ..., P_7$ can be determined by hypothetical reasoning.

- P_1 = $\{\{LowerAbility$(Alice, Bob)$\}$, $\{HigherAccuracy$(Google Translate, Translution)$\}\}$
- P_2 = $\{\}$
- P_3 = $\{\{LowerAbility$(Bob, Carol)$\}\}$
- P_4 = $\{\{LowerAbility$(Bob, Carol), $HigherAccuracy$ (Google Translate, J-Server), $EasierTask$(Chat, News)$\}$ $\}$
- P_5 = $\{\}$
- P_6 = $\{\{HigherAccuracy$(Google Translate, Translution)$\}\}$
- P_7 = $\{\}$

Next, the line 16 to 24 of AcquireHypotheses checks the consistency of the element in the direct product of $P_1, ..., P_7$. For example, one of the sets of hypotheses that can prove the most reputations is $\{LowerAbility$(Alice, Bob), $LowerAbility$ (Bob, Carol), $HigherAccuracy$ (Google Translate, J-Server), $EasierTask$(Chat, News)$\}$. In the same way, the other set of hypotheses that can prove maximum number of reputations is $\{LowerAbility$ (Bob, Carol), $HigherAccuracy$(Google Translate, J-Server), $EasierTask$(Chat, News)$\}$. This paper doesn't refer to the method to select the set of hypotheses. It will be future work, but this time we are supposed to select the former set of hypotheses. In the next section, the service will be selected by this set.

Algorithm 3 ServiceSelection(u, t)

K /* The set of background knowledge */
CH /* The set of the hypotheses which is consistent and can prove each reputation */
R /* The set of hypotheses */
u /* The information of the user */
t /* The information of the task */
$S = \{s_1, ..., s_n\}$ /* The set of the service */
US /* The set of useful service */
BK /* The set of knowledge for service selection */
$US \leftarrow \emptyset$
$BK \leftarrow K \cup CH \cup R$
for all s_i **in** S **do**
  if $BK \vdash Useful(u, s_i, t)$ **then**
    $US \leftarrow US \cup \{s_i\}$
  end if
end for
return US

C. Service Selection Based on Hypotheses

In Section III-B, we explained our method to acquire consistent hypotheses from reputation information. However, our goal is service selection from reputation information. Therefore, a method of service selection based on a set of consistent hypotheses is needed. Service selection can judge the usefulness of a service based on consistent hypotheses. It proposes useful services given the user and the task. This algorithm is based on deductive inference, and the goal of inference is $Useful(user, service, task)$.

We show a service selection example in Section III-A. With reference to Table I, the example problem is: Alice is looking for a useful service for news article translation. Receiving the inquiry, service selection tries to prove the usefulness of the service candidates: $Useful$(Alice, Google Translate, News), $Useful$(Alice, Translution, News), $Useful$(Alice, J-Server, News). As a result, J-Server is proved to be useful because of $Useful$(Carol, J-Server, News) $\cap$ $LowerAbility$(Alice, Bob) $\cap$ $LowerAbility$(Bob, Carol). This is because inference rule 1 and the transitivity.

First, by transitivity, $LowerAbility$(Alice, Bob) $\cap$ $LowerAbility$(Bob, Carol) convert into $LowerAbility$ (Alice, Carol). Next, inference rule 1, $Useful$(Carol, J-Server, News) $\cap$ $LowerAbility$(Alice, Carol) turns to $Useful$(Alice, J-Server, News). Then, J-Server is useful for Alice to translate news article. Note that, Translution and Google Translate can't be proven useful and so are not chosen.

Algorithm 3 is service selection using the set of consistent hypotheses. ServiceSelection returns the set of useful services according to the user's query, which consists of the information of the user, and the information of the task the user will carry out. We assume that hypothesis acquisition outputs just one set of consistent hypotheses. Also, we assume that the reputation information is correct. The data necessary for service selection are the reputation information, the set of consistent hypotheses, background knowledge, and the set of services. When judging the usefulness of a service, it tries to prove $Useful(user, service, task)$. If proved, that service is useful to the user and the task.

IV. Architecture of Service Selection

In this section, we explain the architecture for service selection based on the algorithm in Section III. It recommends useful services according to user's query using the set of consistent hypotheses output by the hypothesis acquisition system.

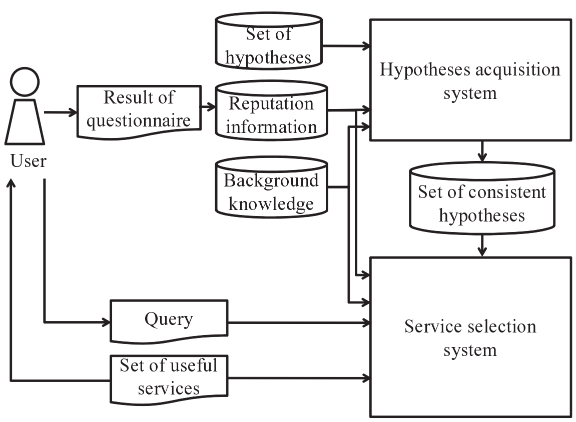


Figure 4. Integrated Architecture

Figure 4 shows the integrated architecture. Here, a user plays two roles. First, he/she inputs a reputation, which is necessary for hypotheses acquisition. Second, he/she asks for a useful service, which triggers the service selection system.

A. Hypotheses Acquisition System

Figure 5 is the system architecture of the hypothesis acquisition system. The system outputs a set of consistent hypotheses using the algorithm explained in Section III-B. Here, the process in the box of the hypothesis acquisition system is the body of algorithm. A user inputs the reputation of services he/she has used already via a questionnaire. Reputation information gathered by the questionnaire is composed of the triplet $(user, service, task)$ and the usefulness for this follows the definition in Section III-A. Whenever a questionnaire entry is sent to the system, the reputation information is added, and the hypothesis acquisition system is executed. The input data for this is all reputation information. The process in the system follows the algorithm detailed in Section III-B. First, the system acquires the sets of hypotheses that can prove each reputation against all reputations. It then checks the sets of hypotheses for consistency. In this way, a consistent set of hypotheses is obtained by hypothetical reasoning.

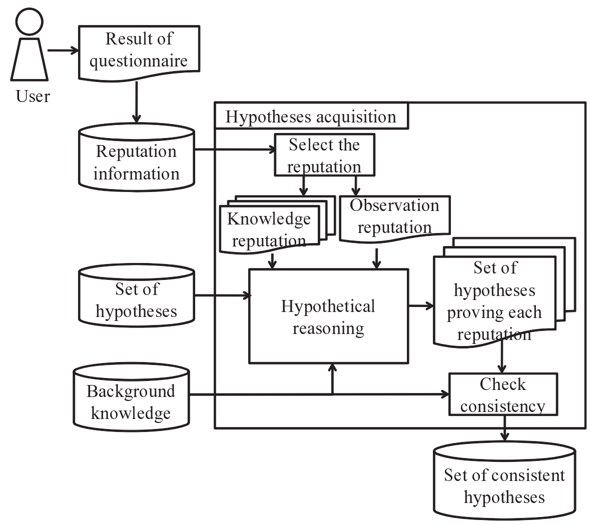


Figure 5. The Architecture of Hypotheses Acquisition System

B. Service Selection System

Figure 6 is the system architecture of service selection. The service selection system is triggered by the user's query, and selects useful services for the user. This system is based on general logic programming. The hypothesis acquisition system outputs one set of consistent hypotheses. The data necessary for service selection are the same as for hypothesis acquisition: background knowledge, inference rules, and reputation information. A user specifies the task, and asks which service is useful to the user and task. According to this query, the service selection system first obtains the user information and task information. Next, for each service in the set of services, it judges the usefulness of the triplet $(user, service, task)$. Service evaluation is based on deductive inference in logic programming. The system judges the usefulness of all services, and returns the services judged to be useful to the user. The user can then invoke and execute these useful services.

V. Application to Language Service Recommendation

We applied our service selection framework to the real world. For this study, we implemented the translation service recommendation system in Language Grid Playground, an interface that allows user customization in a multilingual environment [12]. Playground offers services such as translation services with user dictionary, editing user-sourced bilingual dictionaries, and morphological analyzers. For this

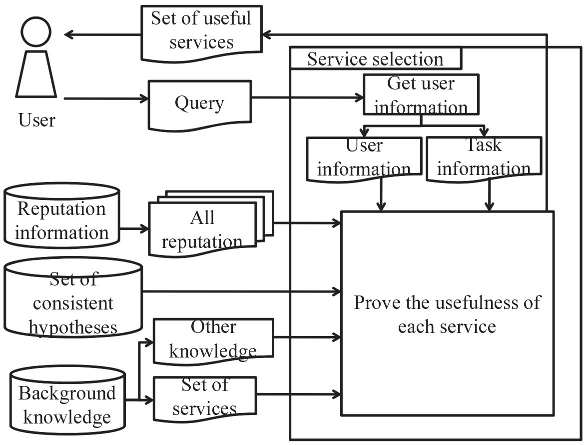


Figure 6. The Architecture of Service Selection System

implementation, we wrote the hypothetical reasoning module in PrologICA. PrologICA is an extension of Prolog, and enables hypothetical reasoning simply by describing knowledge and integrity constraints [13].

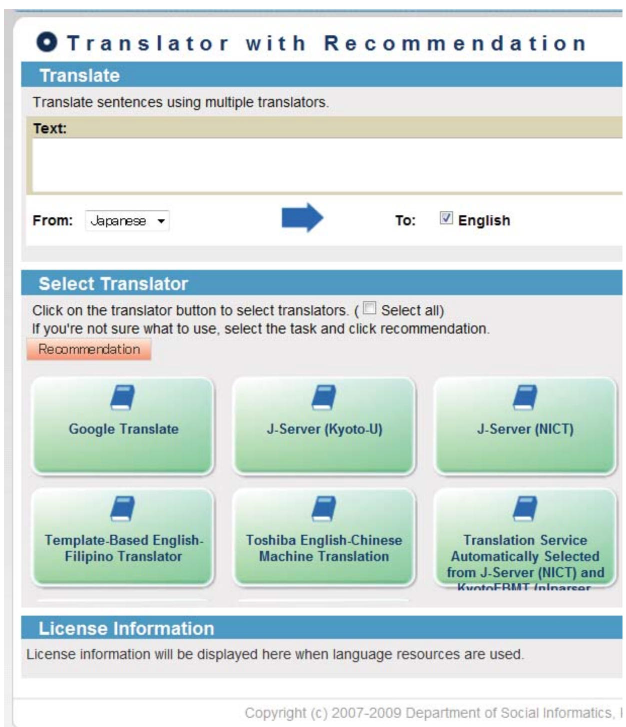


Figure 7. Service Recommendation System: Initial State

Figure 7 and Figure 8 are real Playground screens with service recommendation. First, the bottom of the screen shows a list of machine translators, see Figure 7. A user can invoke and execute a translation service by selecting it

Figure 8. Service Recommendation System: Displaying Recommendation Reason

and specifying the source/target language. However, there are more than ten machine translators, so he/she can't find which service is useful at the first visit. The solution is to select the task and the source/target languages, and push the Recommendation button. After the recommendation process is completed, the useful service is chosen, and he/she can translate sentences. The result of translation can be judged by user in terms of useful/useless. The reason for judging the service as useful is given. Figure 8 shows the reason for the usefulness of J-Server when Alice chooses news as the task. This reason is explained in Section III. J-Server is useful to Alice in translating the news article is equivalent to "J-Server is useful to Carol in translating news articles" and "Alice has lower ability than Carol". Thus, the set of reputations holds "J-Server is useful to Carl in translating news articles" and the consistent hypothesis set has "Alice has lower ability than Carol". Therefore, the system recommends J-Server to Alice.

VI. Conclusion

The contributions of this study are as follows.

Order Relations Acquisition
: We formalized a method to acquire the order relations necessary for service selection using an approach based on hypothetical reasoning. We also proposed an algorithm that uses order relations to select useful services.

Integration for Service Selection Platform
: We proposed an integrated architecture to fuse the hypothesis acquisition engine, based on hypothetical reasoning, and the service selection engine. These engines can be applied to not only language services, but also other services whose evaluation by users varies according to user ability.

In this study, we proposed a service selection framework based on the reputation information instead of a quantitative QoS metric. Note that hypothetical reasoning can resolve two problems: contradiction among users, and data insufficiency.

ACKNOWLEDGMENT

This research was partially supported by a Grant-in-Aid for Scientific Research (A) (21240014, 2009-2011) from Japan Society for the Promotion of Science (JSPS), and also from Global COE Program on Informatics Education and Research Center for Knowledge-Circulating Society.

REFERENCES

[1] T. Ishida, "Language grid: an infrastructure for intercultural collaboration," in *Applications and the Internet, 2006. SAINT 2006. International Symposium on*, jan. 2006, pp. 5 pp. –100.

[2] P. Koehn and C. Monz, "Manual and automatic evaluation of machine translation between european languages," in *Proceedings of the Workshop on Statistical Machine Translation*, ser. StatMT '06. Stroudsburg, PA, USA: Association for Computational Linguistics, 2006, pp. 102–121.

[3] M. Fuji, N. Hatanaka, E. Ito, S. Kamei, H. Kumai, T. Sukehiro, T. Yoshimi, and H. Isahara, "Evaluation method for determining groups of users who find mt useful," in *MT Summit VIII: Machine Translation in the Information Age*, 2001, pp. 103–108.

[4] J. Zhang and H. Cai, *Services computing*. Springer, 2007.

[5] L. Zeng, B. Benatallah, M. Dumas, J. Kalagnanam, and Q. Z. Sheng, "Quality driven web services composition," in *Proceedings of the 12th international conference on World Wide Web*, ser. WWW '03. New York, NY, USA: ACM, 2003, pp. 411–421.

[6] Z. Xu, P. Martin, W. Powley, and F. Zulkernine, "Reputation-enhanced qos-based web services discovery," in *Web Services, 2007. ICWS 2007. IEEE International Conference on*. IEEE, 2007, pp. 249–256.

[7] Y. Liu, A. H. Ngu, and L. Z. Zeng, "Qos computation and policing in dynamic web service selection," in *Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters*, ser. WWW Alt. '04. New York, NY, USA: ACM, 2004, pp. 66–73.

[8] L. Shao, J. Zhang, Y. Wei, J. Zhao, B. Xie, and H. Mei, "Personalized qos prediction forweb services via collaborative filtering," in *Web Services, 2007. ICWS 2007. IEEE International Conference on*, july 2007, pp. 439 –446.

[9] A. Bramantoro and T. Ishida, "User-centered qos in combining web services for interactive domain," in *Semantics, Knowledge and Grid, 2009. SKG 2009. Fifth International Conference on*, oct. 2009, pp. 41 –48.

[10] D. Poole, R. Goebel, and R. Aleliunas, *Theorist: A logical reasoning system for defaults and diagnosis*. Springer-Verlag, 1987, ch. 13, pp. 331–352.

[11] R. Reiter, *On closed world data bases*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1987, pp. 300–310.

[12] S. Sakai, M. Gotou, Y. Murakami, S. Morimoto, D. Morita, M. Tanaka, and T. Ishida, "Language grid playground: light weight building blocks for intercultural collaboration," in *Proceeding of the 2009 international workshop on Intercultural collaboration*, ser. IWIC '09. New York, NY, USA: ACM, 2009, pp. 297–300.

[13] O. Ray, "Prologica: a practical system for abductive logic programming," in *in Proceedings of the 11th International Workshop on Non-monotonic Reasoning*, 2006, pp. 304–312.

Collaborative Translation by Monolinguals with Machine Translators

Daisuke Morita
Department of Social Informatics,
Kyoto University
Yoshida-Honmachi, Sakyo-ku,
Kyoto 606-8501, Japan
Tel: 81-75-753-5396
E-mail: morita@ai.soc.i.kyoto-u.ac.jp

Toru Ishida
Department of Social Informatics,
Kyoto University
Yoshida-Honmachi, Sakyo-ku,
Kyoto 606-8501, Japan
Tel: 81-75-753-4821
E-mail: ishida@i.kyoto-u.ac.jp

ABSTRACT
In this paper, we present the concept for collaborative translation, where two non-bilingual people who use different languages collaborate to perform the task of translation using machine translation (MT) services, whose quality is imperfect in many cases. The key idea of this model is that one person, who handles the source language (source language side) and another person, who handles the target language (target language side), play different roles: the target language side modifies the translated sentence to improve its fluency, and the source language side evaluates its adequacy. We demonstrated the effectiveness and the practicality of this model in a tangible way.

ACM Classification: H5.3 [Information interfaces and presentation]: Group and Organization Interfaces. - Computer-supported cooperative work.

General terms: Design, Human Factors

Keywords: Machine translation, intercultural collaboration, computer-mediated communication

INTRODUCTION
Internationalization and the spread of the Internet are increasing our chances of seeing and hearing many languages. As a result, the number of multilingual groups where the native languages of the members differ is increasing. In the past, communication in such groups typically took place in one language, which was in many cases English. However, members who are required to communicate in a non-native language frequently find communication difficult [2,4,7], thus such collaboration tends to be ineffective[1,8].

Machine translation (MT) is a powerful tool for such groups, because it allows people to communicate in their native language. Actually, many groups in fields of intercultural collaboration use MT in their activities.

MT was useful for realizing some level of communication, because participants could pick up some of the meaning even if some words were badly translated [5]. However, most MT systems make many translation errors. More precisely, many of the machine translated sentences are generally neither adequate nor fluent. In intercultural and multilingual collaboration based on MT, translation errors have caused mutual misconceptions [6]. Moreover, it is difficult to identify translation errors because of the asymmetric nature of MT [9].

In this paper we present the concept of collaborative translation, where two non-bilingual people who use different languages collaborate to perform the task of translation with an MT system. The task of the collaboration is set to translate documents written in one language correctly into another language. In collaborative translation, translation errors decrease the credibility of the translated documents. In the past, only bilingual people could usually detect such translation errors and modify them correctly. This paper presents the model for collaborative translation, where the model does not assume the presence of bilingual people. The collaborative translation is designed to improve imperfect MT quality.

The key idea of this model is to solve the above-described issues about an MT where one person, who handles the source language (source language side) and another person, who handles the target language (target language side), play different roles. The target language side modifies the machine translated sentence to improve its fluency. The source language side evaluates the adequacy between the back-translation of the modified sentence and the source sentence. In addition, we demonstrate the effectiveness of this model with the prototype system of collaborative translation.

HUMAN-ASSISTED MACHINE TRANSLATION
Practice in the Filed of Intercultural Collaboration
In many real fields of intercultural collaboration, MT is used as a tool for communication and information sharing. We will cite internationally active NPO group in Japan as an example of groups working with MTs. This group works with groups

IUI'09, February 8–11, 2009, Sanibel Island, Florida, USA.

in South Korea, Austria, and Kenya. Those groups have a variety of native languages such as Japanese, Korean, German and English.

English is frequently used as a common language for communication in a multilingual community where the native languages of the members differ. However, it is often the case in such community that there are people who are not proficient in English. The problem is that using English or non-native language in communication tends to make it difficult for such people to share the information with others [2,4,7]. In order to foster information sharing and invigorate intergroup discussion by solving this kind of problem, groups noted in the foregoing developed their own web BBS system using MTs. In this system, each person edits an article in his or her native language. The article is translated via this system, and this system enables other people to read contents of the article in their own native language. However, the quality of MT is often imperfect. This can make it difficult to share the information among the members of those groups. Therefore, this system enables people to correct errors of machine translated sentences manually. The illustration of this web BBS system is shown in Figure 1. In this figure, posting a Japanese article is taken for instance. Machine translated sentences can be modified to be natural expression. This makes intragroup information sharing possible.

In addition, the readability as well as the quality of machine translated sentences can be improved by guessing the meaning of translated sentences from the context of text and common knowledge in the community when modifying the translated sentences.

Example 1: Improvement of translation quality by modifying machine translated sentence

The Japanese sentence "All children who looked at the picture were surprised" was translated into English as "Everyone was surprised at the children who saw the picture." This English sentence differed in meaning from the original Japanese sentence. However, a native English speaker guessed the original meaning of the sentence from context and background of his or her community and modified the English sentence as "Children were surprised to look at the picture." This modified English sentence has the same meaning as the original Japanese sentence.

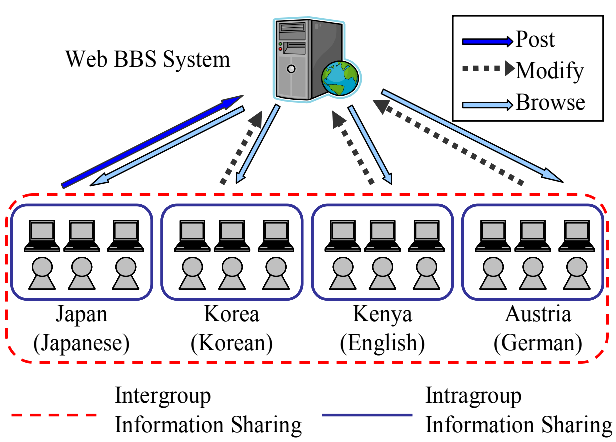


Figure 1: Illustration of web BBS system of the community

Problems in Modifying Machine Translation

Wordy and unnatural machine translated sentences can be expressed naturally by modifying them. This results in making the meaning of translated sentences clearer and intragroup information sharing easier. From this point of view, human-assisted machine translation is useful way for real fields of intercultural collaboration. However, there are two main problems in the naive implementation of human-assisted machine translation. The problems are revealed below.

Example 2: Misinterpretation of a meaning of a machine translated sentences

The Japanese sentence "He needed 1 week to cure a cold" was translated into English as "He was necessary to correct a cold for 1 week." Since there were diction and grammar errors in this English sentence, this sentence was modified to be natural expression by the native English speaker. However, he or she modified this English sentence as "He should recover from a cold within 1 week." This modified English sentence differs in meaning from the original Japanese sentence.

A person who modifies machine translated sentences can never understand original meanings of those sentences. Therefore, he or she might misinterpret meanings of machine translated sentences. Due to this, the modified sentence might differ in meaning from its original sentence.

Example 3: Incomprehension of a meaning of a machine translated sentence

The Japanese sentence "His belly is sticking out" was translated into English as "A stomach has gone out to him." A native English speaker cannot understand the meaning of this machine translated English sentence. Therefore, this sentence remained to be unmodified.

It is almost impossible to modify a phrase of machine translated sentence that he or she cannot make sense of. Such phrase tends to remain to be unmodified. As a result, information about such phrase cannot be shared internationally.

These two problems make it difficult to share information properly. It is true that human-assisted machine translation is helpful as measures for information sharing in real fields of intercultural collaboration. However, these two examples show that the naive implementation of human-assisted machine translation lacks in two procedures; one is a procedure for determining whether a modified version of a machine translated sentence has the same meaning as its original sentence, and the other is a procedure for determining whether the content of a machine translated sentence is understandable.

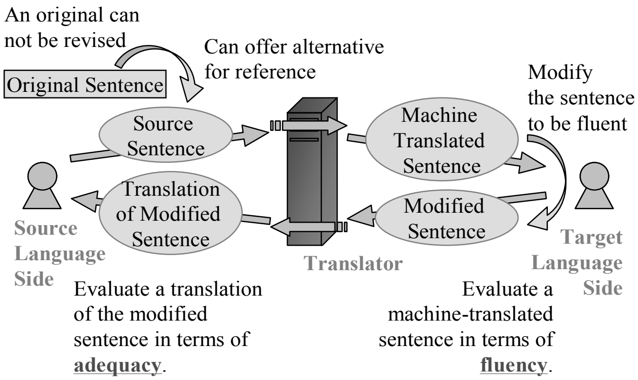


Figure 2: The basic concept of collaborative translation

COLLABORATIVE TRANSLATION

Definition

Participants are two non-bilingual people: one person who handles the source language (source language side), and one person who handles the target language (target language side). Only an MT system performs the task of translation. Participants work at their own computers that are linked over the public network. The goal of collaborative translation is to translate documents correctly. While the original document can not be revised, the source language side can submit alternatives to the original sentences to the MT system to create reference material.

The source language side and the target language side play different roles. The target language side cannot determine whether a machine translated sentence has the same meaning as the original sentence. However, he or she can determine whether the machine translated sentence is fluent. Therefore, he or she can modify the non-fluent sentences more fluent. We assume that the sentences modified by a person are always fluent. Like the target language side, the source language side cannot determine whether the machine translated sentence has the same meaning as the original sentence. However, given machine translation of a sentence modified by the target language side, the source language side can determine whether the back-translation of the modified sentence has the same meaning as the original sentence. By thinking of this, he or she determines whether a machine translated sentence has the same meaning as the original sentence.

The above definitions are illustrated in Figure 2.

Due to the definition, collaborative translation has the procedure for determining whether a modified version of a machine translated sentence has the same meaning as its original sentence. However, a procedure for determining whether the content of a machine translated sentence is understandable is also required as is shown in the previous section. In addition to the basic concept, the procedure for confirming the readability of the machine translated sentence before modifying it is added in collaborative translation. If the target language side cannot understand the content of a machine translated sentence, he or she requests the source language side to modify a source sentence until its machine translated sentence can be understandable. As well, if the source language side cannot

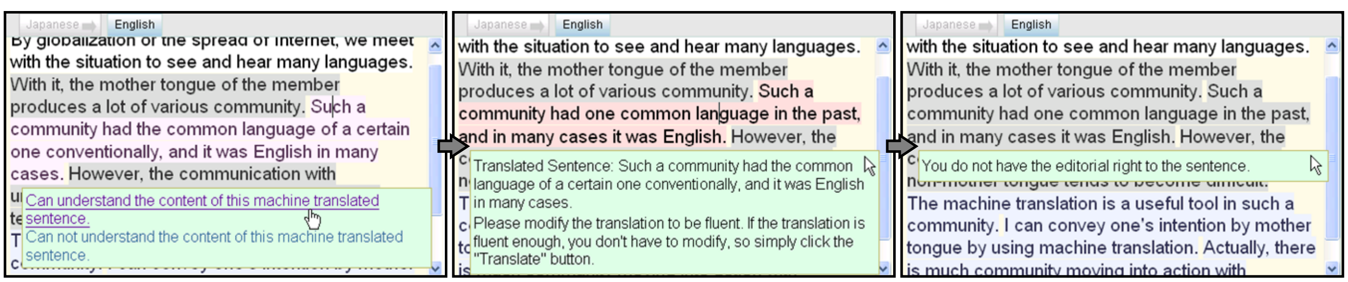


Figure 3: The process flow in the target language side's turn (in Japanese-English translation): (a) he or she evaluates the readability of the machine translated sentence, (b) and if it is human-readable, he or she modifies it to make it fluent. (c) He or she cannot edit the sentence during the source language side's turn.

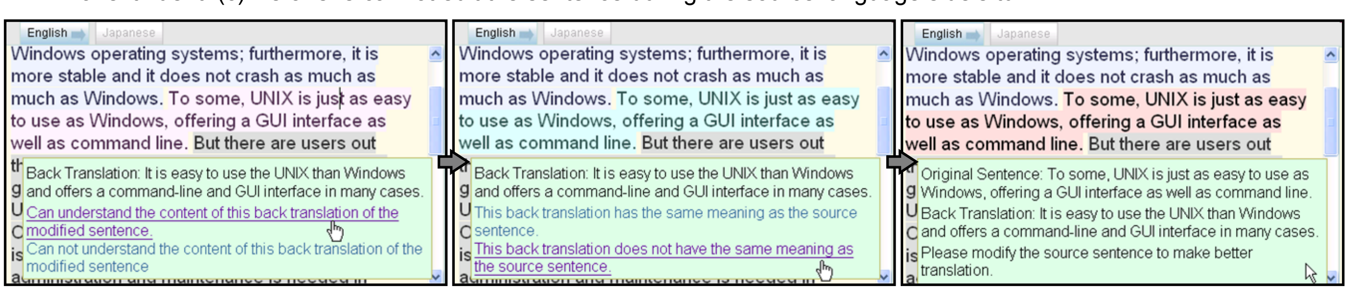


Figure 4: The process flow in the source language side's turn (in English-Japanese translation): (a) he or she evaluates the readability of the back-translation of the modified sentence, (b) and if it is human-readable, he or she also determines whether it has the same meaning as the source sentence. (c) If it does not, he or she modifies the source sentence.

Source Language Side (Japanese)	Target Language Side (English)
He needed a week to cure a cold.	**MT:** *He was necessary to correct a cold for 1 week.*
MT: *He should recover from a cold within a week.*	He should recover from a cold within 1 week.
It took a week for him to cure a cold.	**MT:** *It took 1 week for him to correct a cold.*
MT: *It takes 1 week in order to recover from his cold.*	It needs 1 week to recover from his cold.

Figure 5: The problem of Example 2 is solved in collaborative translation

Source Language Side (Japanese)	Target Language Side (English)
His belly is sticking out.	**MT:** *A stomach has gone out to him.*
	(cannot read the machine-translated sentence)
He is a little fat.	**MT:** *He's a little overweight.*
MT: *He is slightly overweight.*	He's a little bit overweight.

Figure 6: The problem of Example 3 is solved in collaborative translation

understand the back-translation of the modified sentence, he or she cannot determine whether the back-translation has the same meaning as the original sentence. In this case, the target language side is requested to modify the machine translated sentence until the back-translation of its modified version can be understandable.

The Prototype System

The prototype system for collaborative translation was designed to realize its all procedures to test the effectiveness of collaborative translation. This system was developed as a browser-based application. Web services of MTs provided by Language Grid Project [3] were used as MT modules of this system. The prototype divides a document into sentences, and performs the procedures independently in the respective sentences. The user client GUI displays the progress with each sentence, and guides the users on what to do as is shown in Figure 3 and Figure 4. The tasks include modification, readability evaluation, and adequacy evaluation. More concretely speaking, the progress is displayed by highlighting the respective sentences. When the caret is on a sentence, the explanation of what to do or criteria for the evaluation of readability or accuracy are displayed in the pop-up box. Users can conduct the procedures of collaborative translation by following the directions of the user client.

Effectiveness

Figure 5 shows that the problem of the target language side's misinterpretation was solved by applying the collaborative translation system to the Example 2. The source language side is native Japanese speaker, and the target language side is native English speaker. Outputs from MTs are indicated in italics. In the first turn, the target language side modified the machine translated sentence with his or her misinterpretation. However, the source language side could determine that the back-translation of the modified sentence did not have the same meaning as the original sentence. This showed that the target language side may misinterpret the meaning of the translated sentence. The source language side modified the source sentence, and the target language side received its machine translated sentence which was expressed differently from previous one. In the second turn, the target language side modified it with his or her interpretation which was different in the first turn. The source language side determined that the back-translation had the same meaning as the original sentence. In sum, it was confirmed that the translated sentence had the same meaning as the original sentence. The target language side's misinterpretation can be detected and corrected by applying the collaborative translation system.

Figure 6 shows that the problem was solved by applying collaborative translation to the Example 3 which the target language side cannot start to modify a machine translated sentence due to its incomprehension. In the first turn, the target language side could not understand the meaning of the machine translated sentence. Therefore, the system requested the source language side to modify the source sentence. The target language side received its machine translated sentence which was expressed differently from previous one. In the second turn, the target language side modified it because he or she could understand its meaning. The source language side determined that the back-translation had the same meaning as the original sentence. Therefore, it was confirmed that the translated sentence had the same meaning as the original sentence. The collaborative translation system can continue without stopping a series of its processes even if the content of a machine translated sentence is not understandable.

The collaborative translation system provides the procedure for determining whether a modified version of a machine translated sentence has the same meaning as a corresponding original sentence. In addition, if one person cannot understand the content of a machine translated sentence, this system also enables the other person to modify a corresponding source sentence again. Two main problems of the naive implementation of human-assisted machine translation can be solved by collaborative translation. It is revealed that collaborative translation is useful for fields of intercultural collaboration.

CONCLUSION

Although many groups use MT as a collaboration tool, a poor quality of an MT tends to cause many misconceptions. In order to adjust to low quality of MT, people in fields of intercultural collaboration try to share information by modifying machine translated sentences manually. This is very helpful as measures of improving translation quality, but its naive implementation has the disadvantage that it cannot guarantee the quality of a modified version of a machine translated sentence.

To translate documents correctly, a much better translation quality is required. Collaborative translation is the concept that humans adjust machine translated sentences to improve the translation quality. With this system, we can expect a better translation quality.

Our main research contribution is that we have shown the concept of collaborative translation, which is the methodology for improving imperfect machine translation with non-bilingual people's assistance. The key idea of the model of collaborative translation is to solve the above-described issues about an MT where the source language side and the target language side play different roles. The target language side cannot determine whether the machine translated sentence has the same meaning as the original sentence. However, he or she can modify the machine translated sentences to be fluent if he or she can understand the content of those sentences. On the other hand, the source language side can evaluate the translation quality by determining whether the back-translation of the modified sentence has the same meaning as the original sentence. The effectiveness and the practicality of collaborative translation are confirmed by solving examples of real problems in intercultural collaboration with the prototype system.

ACKNOWLEDGMENTS

This research was partially supported by Global COE Program ``Informatics Education and Research Center for Knowledge-Circulating Society". We thank to Language Grid Project for providing us with the web services of MT.

REFERENCES

1. Aiken., M. Multilingual communication in electronic meetings. *ACM SIGGROUP Bulletin*, 23(1):18–19, Apr 2002.
2. Aiken, M., Hwang, C., Paolillo, J., and Lu, L. A group decision support system for the asian pacific rim. *Journal of International Information Management*, 3:1–13, 1994.
3. Ishida, T. Language grid: An infrastructure for intercultural collaboration. *IEEE/IPSJ Symposium on Applications and the Internet(SAINT-06)*, 96–100, 2006.
4. Kim, K. J. and Bonk, C. J. Cross-cultural comparisons of online collaboration. *Journal of Computer Mediated Communication*, 8(1), 2002.
5. Nomura, S., Ishida, T., Yamashita, N., Yasuoka, M., and Funakoshi, K. Open source software development with your mother language: Intercultural collaboration experiment 2002. *International Conference on Human-Computer Interaction (HCI-03)*, 4:1163–1167, 2003.
6. Ogden, B., Warner, J., Jin, W., and Sorge, J. Information sharing across languages using mitre's trim instant messaging. 2003.
7. Takano, Y. and Noda, A. A temporary decline of thinking ability during foreign language processing. *Journal of Corss-Cultural Psychology*, 24(4):445–462, 1993.
8. Tung, L. L. and Quaddus, M. A. Cultural differences explaining the differences in results in gss: implications for the next decade. *Decision Support Systems*, 33(2):177–199, 2002.
9. Yamashita, N. and Ishida, T. Effects of machine translation on collaborative work. *International Conference on Computer Supported Cooperative Work(CSCW-06)*, 512–523, Nov 2006.

Analysis on Multilingual Discussion for Wikipedia Translation

Linsi XIA
Department of Social Informatics
Kyoto University
Kyoto, Japan
xlinsi@ai.soc.i.kyoto-u.ac.jp

Naomi YAMASHITA
Media Information Lab
NTT Communication Science Labs
Kyoto, Japan
naomiy@acm.org

Toru ISHIDA
Department of Social Informatics
Kyoto University
Kyoto, Japan
ishida@i.kyoto-u.ac.jp

Abstract—**In current Wikipedia translation activities, most translation tasks are performed by bilingual speakers who have high language skills and specialized knowledge of the articles. Unfortunately, compared to the large amount of Wikipedia articles, the number of such qualified translators is very small. Thus the success of Wikipedia translation activities hinges on the contributions from non-bilingual speakers. In this paper, we report on a study investigating the effects of introducing a machine translation mediated BBS that enables monolinguals to collaboratively translate Wikipedia articles using their mother tongues. From our experiment using this system, we found out that users made high use of the system and communicated actively across different languages. Furthermore, most of such multilingual discussions seemed to be successful in transferring knowledge between different languages. Such success appeared to be made possible by a distinctive communication pattern which emerged as the users tried to avoid misunderstandings from machine translation errors. These findings suggest that there is a fair chance of non-bilingual speakers being capable of effectively contributing to Wikipedia translation activities with the assistance of machine translation.**

Wikipedia Translation; Multilingual communication; Machine Translation; Multilingual Liquid Threads

I. INTRODUCTION

With the development of Information and Communication Technologies (ICT), knowledge is being shared wider and faster than before [4]. Yet language barriers remain a significant issue when users try to retrieve information written in different languages [6, 9].

Wikipedia provides an excellent example of the situation. For instance, there is a significant difference in the amount of information provided in each language. Due to such uneven distribution of articles among different languages, users have difficulties in cross-language information sharing [7]. Taking Japanese and English for example, it would be hard for Japanese users with low English skills to take advantage of the enormous body of English Wikipedia articles. At the same time, due to the small quantity of Japanese articles, the Japanese Wikipedia cannot provide much information to the Japanese users.

To overcome this problem, and to facilitate cross-language information sharing, Wikipedia contributors are currently carrying out translation activities on a volunteer basis. However, since Wikipedia articles are typically specialized on certain topics fields, such as culture or geography, a Wikipedia translator is basically required to be a bilingual speaker who has knowledge on those specialized topics. The number of such qualified translators is very small, and thus, another approach is desired.

In this paper, we propose an approach that makes use of machine translation technology. This approach is inspired by the fact that two kinds of users are numerous: first, there are many users who have knowledge on a specialized field in the source language. Second, there are also many users who have knowledge of the target language. By bridging these two populations by using machine translation, the former population will be able to transfer their specialized knowledge to the latter population in their native language. The latter population, which has knowledge of the target language, would then be able to paraphrase the source article into target language even if they lack the knowledge of the specialized field and the source language.

However, the difficulty of this approach lies in the simple fact that current machine translations cannot provide a perfect translation result [4]. While translation activities on Wikipedia articles typically require accurate understanding of every term in the source article, this could be quite difficult because the machine translated articles typically include lots of mistranslations and knowledge transfer between the two populations (namely communication between the two populations) could also be hampered by mistranslations. Since the latter population would possibly obtain the ambiguous information of the source article due to mistranslations, translation activities to create an appropriate target article could be quite challenging.

To explore the feasibility of machine translation to support translation activities of Wikipedia articles, we ran an experiment where participants carried out translation activities of Wikipedia articles with the assistance of machine translations. In this paper, we present some findings from analyzing the multilingual communication that took place in the experiment. The findings are important in understanding the communication process and to consider further support for their translation activities.

II. BACKGROUND: MULTILINGUAL LIQUID THREADS

Many tools, such as WikiBhasha, have been developed to support Wikipedia translation activities. However, most of these tools simply provide supports for translating written documents (namely the Wikipedia articles), and do not provide support for communication between contributors using different languages.

Since communication between contributors plays a significant role in current Wikipedia article creation, communication between contributors using different languages should also be well supported [2].

In the current iteration of Wikipedia, a discussion page called "Liquid Threads" is a place for such communication (idea exchanging, knowledge sharing, and debates) between contributors using the same language.

978-0-7695-4546-2/11 $26.00 © 2011 IEEE
DOI 10.1109/Culture-Computing.2011.27

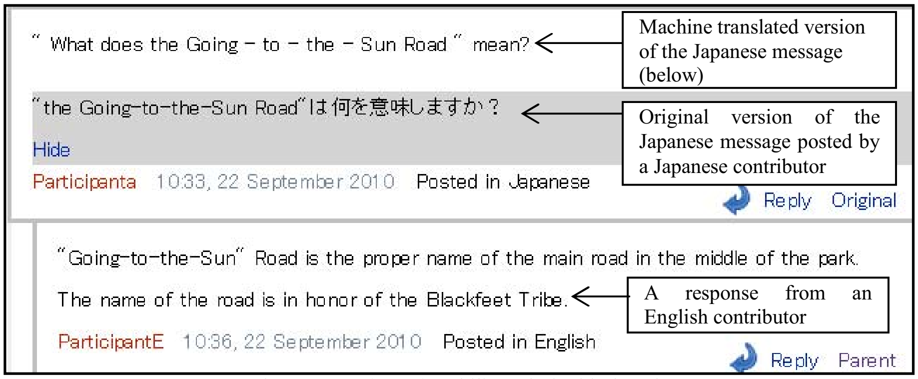


Figure1. Interface of Multilingual Liquid Threads

A multilingual version of the "Liquid Threads" (called "Multilingual Liquid Threads") has recently been released as a MediaWiki Extension. MediaWiki is an open source web-based wiki software application which runs Wikipedia, and was developed by the Wikimedia Foundation. MediaWiki Extensions allow MediaWiki to become more advanced by incorporating many open source projects such as the "Multilingual Liquid Threads".

The language resources in Multilingual Liquid Threads are supported by the multilingual language resource platform called the "Language Grid". The Language Grid is an online multilingual service-oriented platform that enables easy registration and sharing of language services, such as online dictionaries, bilingual corpora, and machine translations [1, 3].

Figure 1 is a screenshot of the Multilingual Liquid Threads. In this example, a Japanese contributor is asking an English contributor for clarification about the meaning of the phrase "the Going-to-the-Sun Road". As we can see from this figure, both the Japanese and English contributors can post messages in their mother tongues. And, since all the messages are automatically translated by machine translations, contributors can view all the messages in their mother tongues regardless of the languages used in the source messages. In the Multilingual Liquid Threads 55 languages are supported in total.

Figure 2 explains how the Multilingual Liquid Threads is situated in Wikipedia translation activities. By enabling multilingual communication with Multilingual Liquid Threads, users who have knowledge on a specialized topic in the source language may be able to help the translators (who have knowledge on the target language) clarify the unclear parts of the articles so as to lead them to successful translation of the articles.

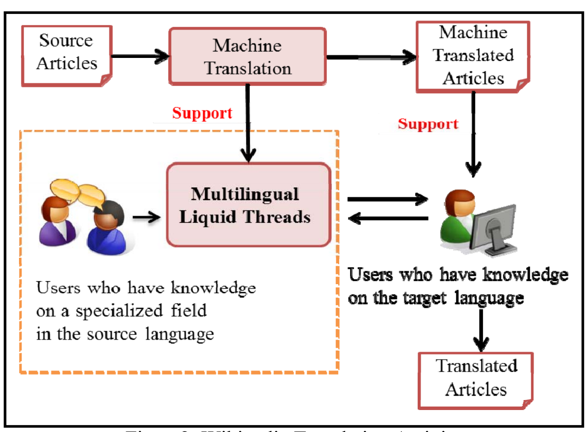


Figure2. Wikipedia Translation Activity with Multilingual Liquid Threads

From next chapter, we will introduce an experiment that shows how Wikipedia contributors work collaboratively with the help of Multilingual Liquid Threads to perform Wikipedia translation activities.

III. CURRENT STUDY: THE WIKIPEDIA TRANSLATION EXPERIMENT

A. Objectives

In order to examine the values of Multilingual Liquid Threads, we decided to evaluate this system from several aspects as follows:

- *System utilization:* First, to evaluate the usefulness of the Multilingual Liquid Threads, we investigated how Multilingual Liquid Threads was used for discussion in Wikipedia translation activities.
- *Ability to transfer knowledge:* Next, to see whether multilingual communication was helpful to their translation activities, we investigated how frequently the users were able to successfully transfer knowledge through the Multilingual Liquid Threads.
- *Influence on communication pattern:* Finally, to see whether and how the system affected the contributors' communication behavior, we observed their multilingual communication pattern throughout their translation activities using Multilingual Liquid Threads.

B. Setting

Task

Three Japanese and two Americans participated in our experiment. The participants were asked to engage in a translation activity using the Multilingual Liquid Threads. Their translation task was to translate the English Wikipedia article "Glacier National Park" into Japanese collaboratively. The Japanese participants were mainly in charge of translating the article into Japanese. The Americans were in charge of helping the Japanese by answering their questions and clarifying the word meanings when requested. All of the communication during the task took place in the Multilingual Liquid Threads. Note that we didn't restrict the language they were able to use.

Participants

Table1. Participants

No.	Nationality	Other Language
A	Japanese	English (High-intermediate)
B	Japanese	English (Intermediate)
C	Japanese	English (Low-intermediate)
D	American	Japanese (Very Little)
E	American	Japanese (Very Little)

Two Americans and three Japanese were recruited for this study. The two Americans were English monolingual speakers with very few Japanese skills. Two Japanese had medium-level English knowledge with a TOEIC score lower than 750, and one Japanese had a TOEIC score higher than 750, but was still not proficient in writing English. Since none of the Japanese had much knowledge about the Glacier National Park, none of the Japanese participants could perform the translation task independently.

Apparatus

In this experiment, the participants were provided with Multilingual Liquid Threads and some additional dictionaries services including the "National Parks Wikipedia Dictionary" and the "Page Dictionary".

We created National Parks Wikipedia Dictionary in advance for this experiment. Titles of English articles that are related to the U.S national parks were extracted and registered into this dictionary. Different language versions of every single article's title were extracted to construct parallel multilingual entries. This specialized dictionary aims to assist translators with better translation result in a specialized topic (namely the U.S National parks). A special dictionary service called Page Dictionary was provided as well. Since multiple contributors worked together on the same article, it was important to assure the consistency of translated terms throughout the article. Page Dictionary is a free-editing dictionary that was implemented in every article so that users can collaboratively create a best-suited dictionary for each article.

To mimic the actual translation activities, we did not restrict the participants from using any language resources on the Web. For example, resources such as Wikipedia and online dictionaries were also available to the participants.

Procedure

The experiment lasted for five days, four hours per day. Prior to their translation activities, the Japanese and American participants were given an instruction on the experiment. (1) All participants were given an introduction about the task. (2) All participants were shown a demonstration to learn apparatus of Multilingual Liquid Threads and Page Dictionary. (3) Every day's working procedure was explained as follows:

Table2. General Working Procedure

Step	Japanese participant	American Participant
1	Task allocation	Read over the original article and get ready to answer questions.
2	Translation	Answer questions when requested
3	Proofreading	Answer questions when requested
4	Interview	Interview

Step1 Since different participants would work on different parts of the article, Japanese participants had to decide the translation task allocation by themselves using Multilingual Liquid Threads before they started to translate article.

Step2 Japanese participants could ask questions at any time during the translation work. Any American or Japanese participant could answer questions. Furthermore, there was no format for an answer and multiple answers were available simultaneously.

Step3 As well as at step 2, both Japanese and American participants could edit the Page Dictionary at any moment and hold discussions on entry creation through Multilingual Liquid Threads.

Step4 At the end of the experiment, every participant was interviewed. Feedback about multilingual communication with Multilingual Liquid Threads was collected.

IV. RESULTS

A. System utilization

First, we investigated how Multilingual Liquid Threads was utilized for discussion in Wikipedia translation activities. All the messages during the experiment were collected and analyzed.

Finally we got 273 messages in total. These messages consisted of 56 threads. A thread is defined as a collection of messages that were discussing the same topic. There were threads which contained only monolingual discussions among Japanese/English participants as well as those which contained multilingual discussion between Japanese and English participants. Messages from American participants were all posted by English, while most of the messages from Japanese participants were posted by Japanese (Only one of them was posted in English by Japanese t A). Note that the content of the English message posted by Japanese A was not directly related to translation activities. A post-interview suggested that the incentive of such behavior from Japanese A was that he thought English messages could express goodwill towards the American participants.

According to the interview, American participants viewed messages in English. Japanese participants basically viewed messages in Japanese, while for messages translated into Japanese, they viewed the original English messages concurrently as assistance for understanding.

To see how the Multilingual Liquid Threads was used during the translation activities, each thread was classified into one of the 4 categories:

- *Translation Task Allocation*
  Threads discussing translation task allocation.
- *Translation Policy*
  Threads discussing policies such as capitalization rules of proper nouns which aimed to build standard translation processes.
- *Article Proofreading*
  Threads clarifying unclear parts of the article and correcting translation errors.
- *Dictionary Checking*
  Threads discussing Page Dictionary creation.
- *Others*
  Threads which do not belong to any of the categories listed above.

Figure 3 shows the categorized result of threads. As shown in Figure 3, the majority of the discussions (73.2%) were devoted to article proofreading.

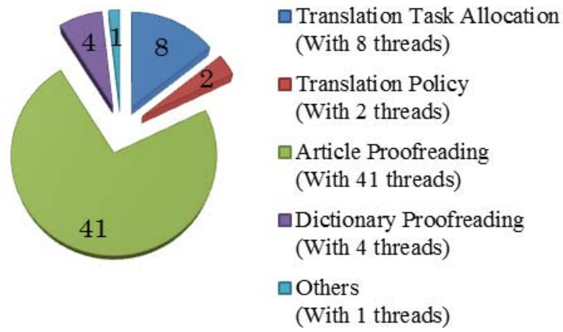


Figure3. Thread Count of Discussion (N=56)

Since discussions on article proofreading were mainly on correcting the mistranslated parts and clarifying the ambiguous terms used in the article, it appears that Multilingual Liquid Threads was mainly used for reducing ambiguity and conveying accurate meaning of the terms used in the article.

B. *Ability to transfer knowledge*

Second, we investigated whether multilingual communication through Multilingual Liquid Threads was actually beneficial to the users in terms of knowledge transfer. In the following, we observed how frequently the users were able to successfully transfer knowledge through the Multilingual Liquid Threads.

All the threads that contained multilingual communications were subject to analysis. As a result, we got 32 threads in total. Table 3 gives a statistics overview as follows.

Table3. Multilingual Thread/Message Count

Multilingual thread count / (All threads)	32 / (56)
Message contained in multilingual threads count / (All messages)	213 / (273)

To see how successful they were in transferring knowledge through the Multilingual Liquid Threads, we used the acknowledgements (such as “I understand”, and “I see”) as a rough indicator of success in knowledge transfer.

Table 4 gives an example of such successful cases. For readability, note that all the Japanese messages were translated into English. In this thread, knowledge about the meaning of the phrase “Going-to-the-Sun” was presented and the knowledge receiver (namely Japanese participants) gave a message of “it was understood” to present successful mutual understanding.

Table4. Example of Successful Knowledge Transfer Cases (Japanese messages were translated into English)

Msg. No.	Original Language	Presenter	Message
1	Japanese	Participant A	What does the “Going-to-the-Sun Road” mean?
2	English	Participant E	“Going-to-the-Sun Road” is the proper name of the main road in the middle of the park. The name of the road is in honor of the Blackfeet Tribe.
3	Japanese	Participant A	It's a proper noun, isn't it? It was understood. Thank you very much.
4	English	Participant E	Correct, it is a proper noun.

We examined all the 32 multilingual communication threads and found that 65.6% (21/32) of all the threads satisfied the requirements for successful knowledge transfer. An observation suggested that each of these 21 threads consisted of a series of questions and answers and began with a Japanese participant issuing a question.

As a result of successful knowledge transfer, a complete and comprehensive Japanese Wikipedia article was created throughout this experiment, which has been uploaded into actual Japanese Wikipedia and is available to access by any Wikipedia viewer.

The result suggests that Multilingual Liquid Threads was basically useful for conveying information between American and Japanese users in our experiment. This result is quite interesting because previous research on machine translation mediated communication has emphasized the difficulties of conveying accurate meaning of the original messages [5].

C. *Influence on communication pattern*

To see how the participants were able to convey accurate meaning of the article, we analyzed their multilingual communication in further details. We focused on those 21 threads which succeed in knowledge transfer.

To see how the information was transferred through a series of questions and answers, we developed a coding scheme that captures the communication style of each thread. The categories used for the analyses are presented in Table 5.

Table5. Message Category

Category	Definition	Example	Freq.
Propositional Question	A question that could be answered with “Yes” or “No”.	[Q] Does “game” have a meaning of Animal?	19.7%
Non- Propositional Question	A question which needs informative answers instead of “Yes” or “No”.	[Q] What does “raid squirrel caches of the pine nuts” mean?	6.0%
Direct Answer	A response which answers to the question directly.	[Q] What is “concession facilities”? Is this one kind of stores? [A] Yes. "Concession facilities" are stores that sell things to tourists.	21.4%

Informative Answer	A response which typically contains more information than requested (in the question).	[Q] Does "game" have a meaning of Animal? [A] Game means wild animals, including birds and fishes, such as are hunted for food or taken for sport or profit. Game is being used as an adjective to describe the fish species found in the lakes and streams.	22.2%
Proposal	A response which contains a proposal to the questioner.	[Q] Thank you very much. Now I understand what Wilder Complex is. But it's a little difficult to choose an appropriate Japan term which corresponds to Complex. [A] My own personal dictionary offers 複合体 or ふくごうたい for this noun "complex". Is this Japanese word too technical?	6.0%
Acknowledgement	Feedback showing that message is understood/accepted.	Thank you very much! It was understood.	17.9%
Other	Uncodable communication.	This is a thread about a question of Wildlife and ecology	6.0%

All the messages were classified into one of the seven categories listed above.

The statistics in Table 5 suggests that the number of propositional questions is three times larger than that of non-propositional questions. Interviews from the Japanese participants revealed that they tried to ask questions in the propositional style to avoid mistranslations by machine translator. However, despite such concerns of the Japanese participants, it appeared that the American participants tended to answer the questions in an informative way; they tended to provide more information than required by the Japanese questioner, even when simple "Yes" or "No" answers were sufficient. Indeed, Table 5 shows that the number of direct answers did not largely surpass the number of informative answers.

The following excerpt is an actual example of a Japanese participant asking a propositional question followed by an informative answer given by an American participant. Note that all the Japanese messages were translated into English for readability.

- **[Question]** *"The one of the The west and northwest are dominated by spruce and FIR and the southwest by redcedar and hemlock; the areas east of the Continental Divide are a combination of mixed pine, spruce, FIR and prairie zones." Is the "redcedar" same as "red cedar"?* Posted by Japanese Participant C
- **[Answer]** *Essentially, yes. Specifically, the mean the Western Redcedar. The Western Redcedar is very different from the Eastern Redcedar which is a type of Juniper and is more bush like.* Posted by Japanese Participant E

In the excerpt above, a simple response as "Yes, it is." should have been enough to answer the question. To see when such an informative response was provided, we further classified the responses of propositional questions into one of the four categories:

Table6. Reponses for Presentations of Proposition

The answer to a propositional question (Yes or No)	Proportion of Direct Answers (Thread Count)	Proportion of Informative Answers (Thread Count)
Yes	14.3% (3/21)	***66.7% (14/21)***
No	0	19.0% (4/21)

Table 6 suggests that the respondents always provided sufficient/additional information when they had to say "no" to the questioner's expectation. More interestingly, the respondents provided additional information even when the questioner's expectation was right.

To figure out the incentives of putting so much effort in providing sufficient information to the questioners, we interviewed the respondents (American participants) for their reasons. American participant D mentioned that:

> *"Sometimes even when I understood the question, I was still worrying about the possibility of Japanese participants raising the questions inappropriately. I mean, they might actually be confused about another part in that sentence? So in case of this situation, I decided to provide useful information as much as I could".*

It seems that the respondents tended to provide more information than requested because of their low confidence in machine translation; they were not sure if they have really understood the questioner's intention because of the potential/possible problems which might have been created due to mistranslation or inadequate English ability of the questioners.

The result reminds us of Yamashita's study [5] where respondents also offered additional information (rather than answering to his/her partner's question) when talking over machine translation. The interesting finding which differs from their study is that the Japanese participants in our study asked questions quite frequently while participants in their study seemed to be reluctant in asking questions. This may in part due to the differences in the tasks used in these studies. Since their task did not require accurate information transferring between the participants, they just ignored the (mistranslated) parts that did not make sense to them. Meanwhile, our task required accurate information transfer, and thus the participants could not ignore the mistranslated parts; they had to ask for clarification when they were not sure if they had understood the meanings correctly.

When a question was issued, it meant that the questioner did not understand a term or wasn't sure if his/her understanding was correct. The respondents thus tried to provide as much information as possible so that the questioner could fully understand the term. Since accurate information transfer was their first priority, providing unnecessary or redundant information was not a big issue for them.

V. CONCLUSION

In this paper we reported on the study of introducing Multilingual Liquid Threads. This system enables monolingual speakers to collaboratively translate Wikipedia articles using their mother tongues. In our experiment using this system, we observed both system performance and human behavior in multilingual communication.

First, a trend of discussions on article proofreading was found. Since article proofreading typically refers to correct the mistranslated parts and clarify the ambiguous terms used in the article, we concluded that Multilingual Liquid Threads was mainly used for reducing ambiguity and conveying accurate meaning of the terms used in the article.

Secondly, statistics revealed that most multilingual discussions seemed to be successful in transferring knowledge between different languages by building mutual understating through multilingual communication. This is quite important since it suggests that Multilingual Liquid Threads was basically useful for conveying information between American and Japanese users in our experiment.

Finally, communication patterns were analyzed to find out how knowledge transfer was achieved successfully. It appears that respondents (namely American participants) typically tried to provide as much information as possible so that the questioner could fully understand the term mentioned in the question, since accurate information transfer was their first priority. Thus providing unnecessary or redundant information was not a big issue for them.

These findings suggest that there is a fair chance of non-bilingual speakers contributing to Wikipedia translation activities with the assistance of Multilingual Liquid Threads. However, currently the system is expecting for further improvement to enable more efficient multilingual communication, because more propositional questions and less informative answers could still be expected to reduce communicative effort for contributors. As one of the reasonable approaches, building up a more usable interface for this system to enable a simple way of asking questions is being considered. For instance, question templates could be helpful to reduce effort of considering the format of asking questions. A fixed format could reduce mistranslations during multilingual communication. This could possibly result in more efficient knowledge transfer and benefit users finally. Furthermore, after completing system upgrading, an evaluation involving actual Wikipedia contributors is going to be carried out in the near future.

ACKNOWLEDGMENT

This research is partially supported by Strategic Information and Communications R&D Promotion Programme and also Scientific Research (A) (21240014, 2009-2011) from Japan Society for the Promotion of Science (JSPS).

REFERENCES

[1] Toru Ishida. “Language Grid: An infrastructure for intercultural collaboration,” IEEE/IPSJ Symposium on Applications and the Internet (SAINT-06), 2006.

[2] Daisuke Morita, Toru Ishida. "Collaborative Translation by Monolinguals with Machine Translators," Proceedings of ACM Conference on Intelligent User Interface (IUI'09), pp. 361-365, 2009.

[3] Masahiro Tanaka, Yohei Murakami, Donghui Lin, Toru Ishida. “Language Grid Toolbox: Open source multi-language community site,” International Universal Communication Symposium (IUCS’10), pp105-111, 2010.

[4] Naomi Yamashita, Rieko Inaba, Hideaki Kuzuoka, Toru Ishida. “Difficulties in Establishing Common Ground in Multiparty Group using Machine Translation,” ACM Conference on Human Factors in Comuting Systems (CHI'09), 2009.

[5] Naomi Yamashita, Toru Ishida. "Effects of Machine Translation on Collaborative Work," Proceedings of ACM Conference on Computer Supported Collaborative Work (CSCW'06), pp. 515-524, 2006.

[6] Andreas Riege. “Three-dozen knowledge-sharing barriers managers must consider,” Journal of Knowledge Management, Volume 9, Number 3, pp. 18-35(18), 2005.

[7] Ari Hautasaari, Masanobu Ishimatsu, Linsi Xia, Toru Ishida. “Supporting Multilingual Discussion of Wikipedia Translation with the Language Grid Toolbox,” The Institute of Electronics, Information and Communication Engineers (IEICE’09), NLC2009-44, pp.67-72, 2009.

[8] Sergio Ferrándeza, Antonio Toralb, Óscar Ferrándeza, Antonio Ferrándeza, Rafael Muñoza. “Exploiting Wikipedia and EuroWordNet to solve Cross-Lingual Question Answering,” Information Sciences: an International Journal archive, Volume 179 Issue 20, September, 2009.

[9] Paul Hendriks. “Why share knowledge? The influence of ICT on the motivation for knowledge sharing,” Knowledge and Process Management, Volume 6, Issue 2, pp.91–100, June 1999.