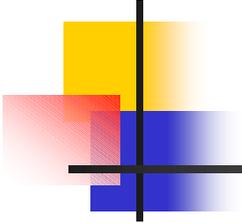


AAMT/Japio特許翻訳研究会 活動報告

2009年11月27日

AAMT/Japio特許翻訳研究会

山形大学 横山晶一
山梨英和大学 江原暉将



目次

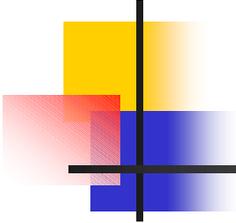
- AAMT/Japio特許翻訳研究会の活動
- 特許翻訳ワークショップ

AAMT/Japio特許翻訳研究会の活動



- 研究会の概要
- 研究成果の紹介

AAMT : アジア太平洋機械翻訳協会
Japio : 日本特許情報機構



研究会の目的

- 特許文書のための機械翻訳技術の調査・研究を行う。
- 研究成果を元に、Japioが技術開発を行う際の有効な指針を与える。
- 特許の機械翻訳に関して国際的な情報交換や連携を行う。



研究会の組織

委員長

辻井潤一 (東大)

副委員長

横山晶一 (山形大)

江原暉将 (山梨英和大)

委員

宮澤信一郎 (秀明大)

梶 博行 (静岡大)

黒橋禎夫 (京大)

宇津呂武仁 (筑波大)

二宮 崇 (東大)

越前谷 博 (北海学園大)

綱川隆司 (静岡大)

安田圭志 (NICT)

熊野 明 (東芝)

下畑さより (沖電気)

潮田 明 (富士通研究所)

三浦 貢 (日本電気)

オブザーバ

中川裕志 (東大)

安藤 進 (元多摩美大)

守屋敏道 (Japio)

渡邊豊英 (Japio)

藤城 享 (Japio)

大塩只明 (Japio)

塙 金治 (Japio)

三橋朋晴 (Japio)

柿田剛史 (Japio)

星山直人 (Japio)

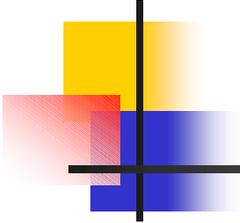
範 暁蓉 (東大 大学院生)

王 向莉 (東大 大学院生)

事務局 村上嘉陽 (ナビックス)

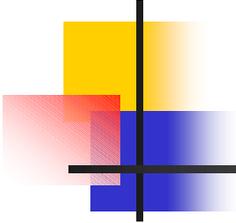
河田容英 (ナビックス)

高田佳代子



研究会の歴史

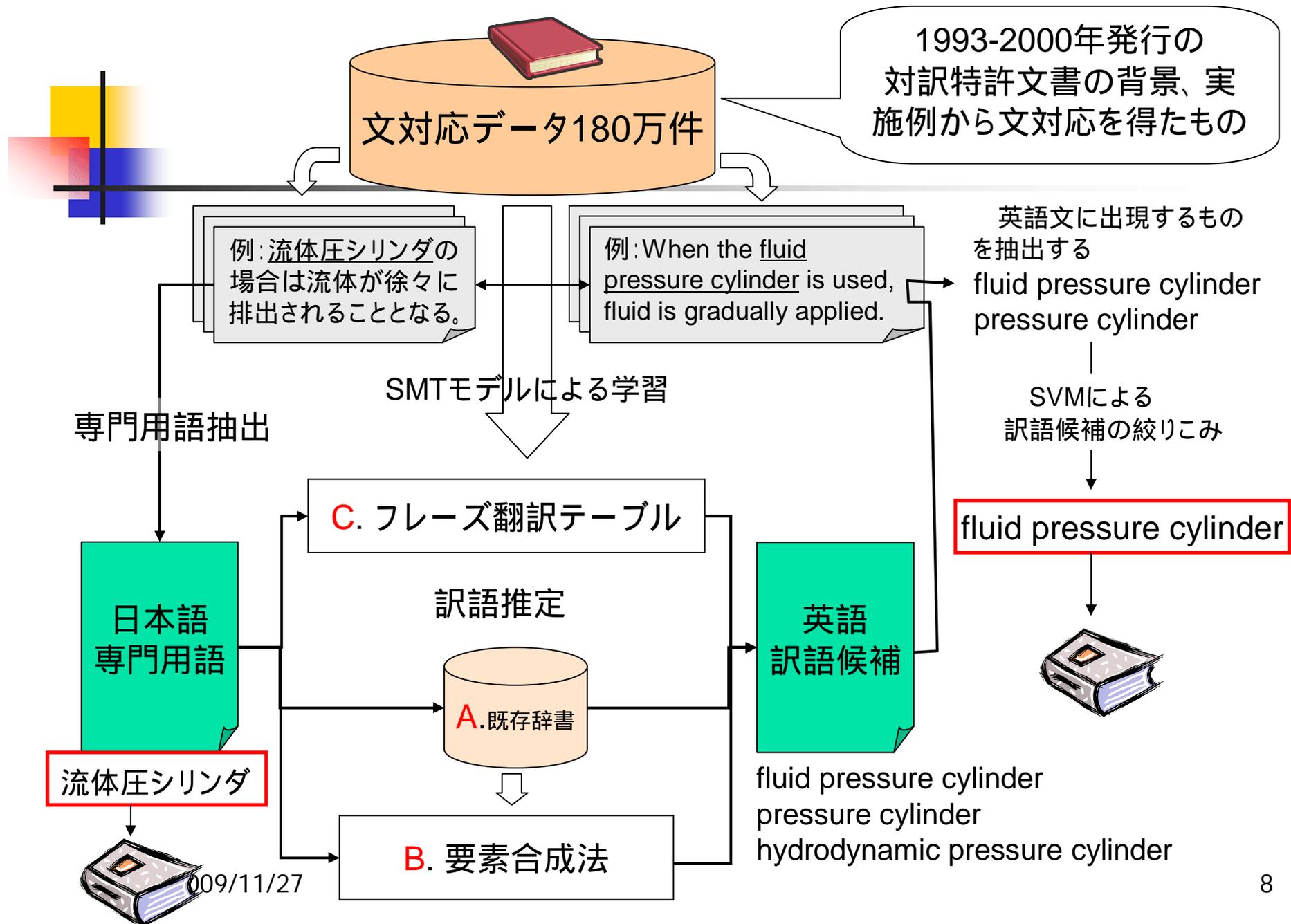
- 平成16年9月発足
- 発足時
委員 9名
オブザーバー 6名



これまでの主な研究内容

- 専門用語辞書の(半)自動構築
- 機械翻訳精度の向上
- 特許文の構文分析
- 翻訳精度の評価法

複数の訳語推定手法を併用した対訳専門用語抽出



既存辞書、要素合成法、フレーズテーブルの 訳語候補が英文中に存在した割合

(a) 全日本語専門用語 (100%)

C. 91%
フレーズテーブル
により訳語候補を出力

B. 43%
要素合成法により
訳語候補を出力

A. 17%
既存辞書により
訳語候補を出力

どの訳語推定手
法でも、訳語を出
力できない
8%

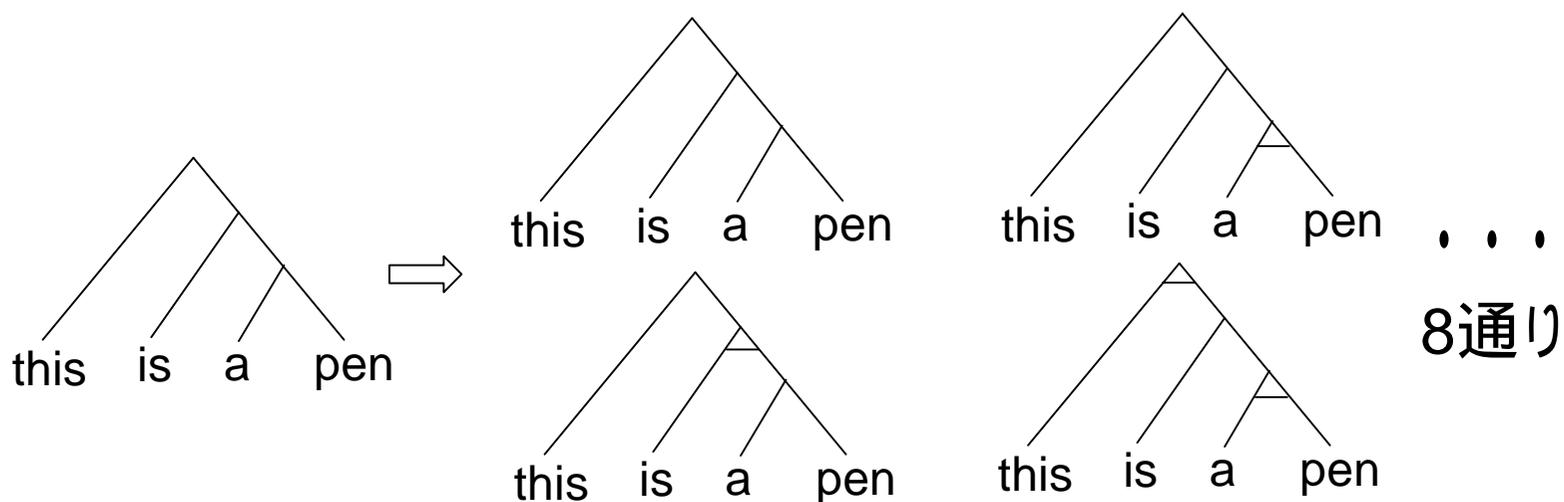
適合率 88.6%
49%
フレーズテ
ーブルのみにより
訳語候補を出力

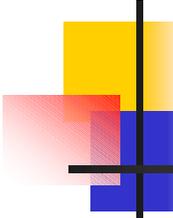
適合率 98.3%
26.2%
要素合成法とフ
レーズテー
ブルのみにより同一の
訳語候補を出力

適合率 98.8%
16.7%
3手法全てによ
り、同一の訳語候
補を出力

構文木による語順制約を導入した 統計翻訳

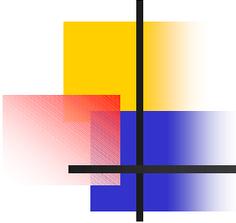
- 句レベル統計翻訳に構文情報を利用
- 語順変更にはIST-ITG制約を適用





英日特許翻訳における性能評価

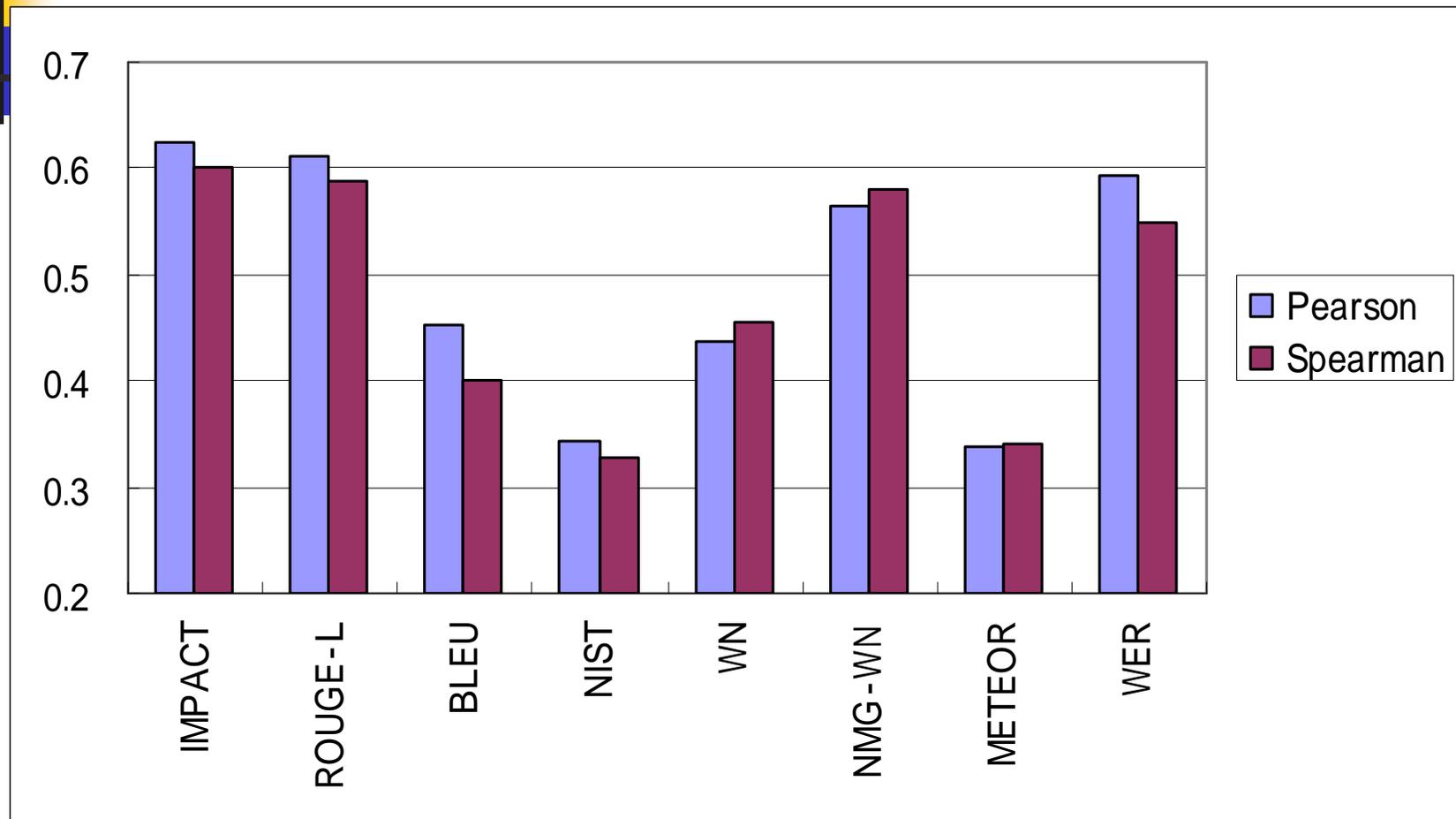
	BLEU	NIST	WER	PER
Monotone	24.91	6.95	79.97	42.02
No constraint	26.83	7.19	81.10	39.52
IBM	28.35	7.29	78.35	39.25
ITG	27.59	7.26	80.29	39.15
IBM+ITG	28.50	7.30	78.01	39.29
IBM+LR	31.17	7.50	76.30	38.61
IST	30.26	7.41	74.90	38.93
IBM+IST	30.07	7.41	73.38	39.05
IBM+LR+IST	32.20	7.61	71.18	38.15



翻訳精度の自動評価法のメタ評価

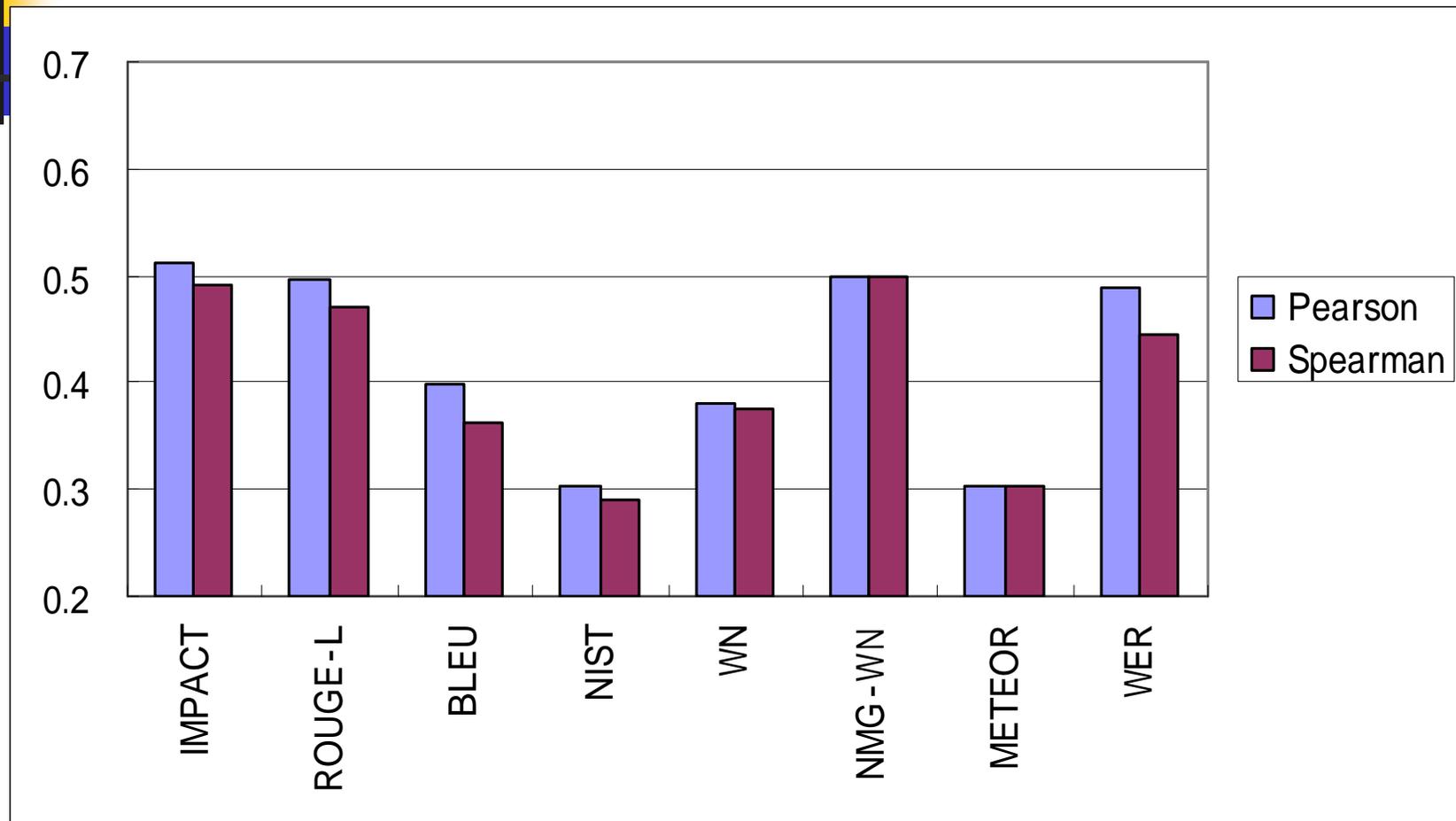
- 人手評価
Adequacy, Fluency
- 自動評価
BLEU, NIST, IMPACT, NMG_WN, WN,
ROUGE-L, METEOR, WER
- メタ評価法
人手評価と自動評価の相関をとる
Pearson's correlation
Spearman's rank correlation

メタ評価結果例1

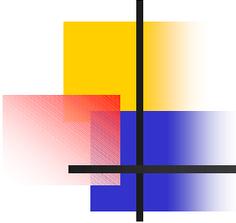


Correlation coefficients to adequacy

メタ評価結果例2



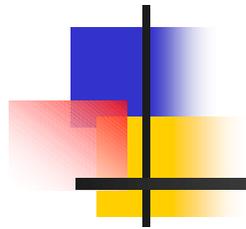
Correlation coefficients to fluency



ここまでのまとめ

- AAMT/Japio特許翻訳研究会の概要
- 専門用語辞書の(半)自動構築
- 構文情報を用いた統計翻訳
- 翻訳精度の評価法

特許翻訳ワークショップ (Workshop on Patent Translation) 概要

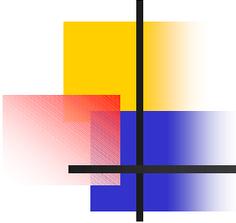


横山 晶一
(山形大学)

特許翻訳ワークショップ

(Workshop on Patent Translation)

- 機械翻訳国際会議(Machine Translation Summit)の付属ワークショップとして開催
- これまでに3回
 - 第1回:2005年ブーケット
 - 第2回:2007年コペンハーゲン
 - 第3回:2009年オタワ



機械翻訳国際会議 (Machine Translation Summit)

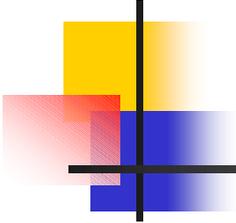
- 機械翻訳の研究、使用、政策に関する唯一の国際会議
- 隔年開催
- IAMT (International Association of Machine Translation)主催
- アジア、欧、米での持ち回り開催

- 3つの機械翻訳協会の連合体

AAMT (Asia-Pacific Association for Machine Translation) アジア太平洋機械翻訳協会

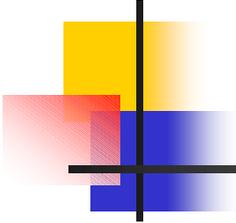
EAMT (European Association for Machine Translation)

AMTA (Association for Machine Translation in Americas)



MT Summit

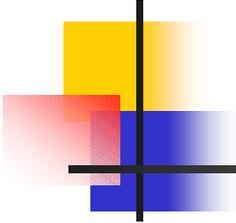
- 研究者、ユーザ、翻訳者、政府機関関係者が一堂に会する国際会議
- 第1回：箱根(1989) ...
- 第10回：プーケット(2005)
- 第11回：コペンハーゲン(2007)
- 第12回：オタワ(2009)
- 第13回：アモイ(2011)(予定)



特許翻訳WSの目的

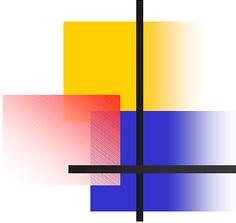
- 特許翻訳に関する研究の現状
- 特許翻訳に関するユーザの問題意識
- 特許翻訳に関する政策的な検討

- AAMT/Japio特許翻訳研究会が中心になって組織



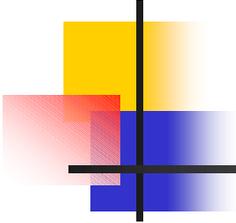
第1回WS

- 日時：2005年9月16日8:30～16:00
- 場所：プーケット(タイ)
- 座長：横山 晶一(山形大学)
- 参加者：約30名
- 参加国：日本、韓国、イスラエル、デンマーク、ドイツなど



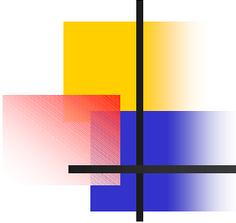
第1回WS招待講演： デンマークにおける特許翻訳

- Prof. Bente Maegaard (Copenhagen University)
- デンマークの特許関連機械翻訳
- 特許の翻訳品質の問題
- EUROTRA
- 翻訳スピード
- 特許テキストの語彙の問題



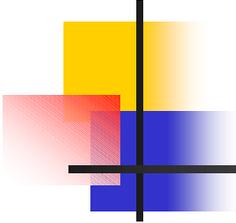
第1回WS:ユーザによる講演

- 各国特許庁による現状の紹介
本WSの特色の一つ
- 河合弘明(特許庁): Patent Information
Web Services using MT at JPO
- Wolfgang Taeger (EPO): European
Machine Translation Programme



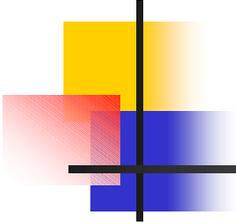
第1回WS:一般講演

- 9件の発表
- 日本: 5件: 沖電気、山形大学、創価大学、静岡大学、Cross Language (在日ドイツ人)
- 韓国: 1件: ETRI
- デンマーク: 2件: LanA Consulting ApS
- イスラエル: 1件: Geoffrey L. Melnick, Ltd.



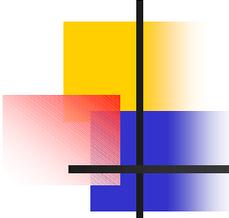
第2回WS

- 日時: 2007年9月11日9:00 ~ 17:00
- 場所: Copenhagen Business School
- 座長: 辻井潤一(東大)、横山晶一(山形大)
- 参加者: 約40名
- 参加国: 日本、デンマーク、アメリカ、カナダ、韓国、スイスなど



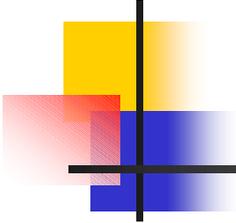
第2回WS招待講演

- Rachel Crem (WIPO (World Intellectual Property Organization)): WIPO's Activities in Patent Translation and Terminology
- WIPO (1967年発足)の概要
- PCT (Patent Cooperation Treaty)の概要
- 特許翻訳サービスの概要



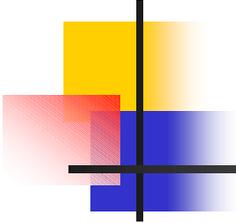
第2回WS: ユーザによる講演

- Wolfgang Taeger (EPO): European Patent Translation Programme
2005年以降の進展、EMTP project概要
- 遠山敬彦(特許庁): Patent Information Web Services using MT at JPO
IPDL (Industrial Property Digital Library),
AIPN (Advanced Industrial Property Network)
概要



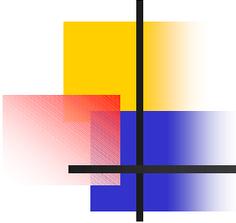
第2回WS:一般講演

- 6件の発表
- 韓国:1件:ETRI
- 日本:3件:富士通、諏訪東京理科大学、山形大学
- デンマーク:2件:University of Copenhagen, LanA Consulting ApS



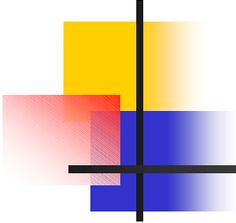
第3回WS

- 日時：2009年8月30日9:00～17:00
- 場所：オタワ
- 座長：江原暉将（山梨英和大）、
横山晶一（山形大）
- 参加者：約40名
- 参加国：日本、中国、韓国、カナダ、アメリカ、フランス、ドイツ、デンマークなど



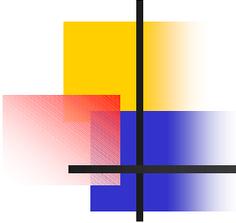
第3回WS招待講演

- Sophie Mangin (EPO): European Machine Translation Programme – Concept, Status and Future Plans
- 王丹(SIPO): SIPO's Efforts on Improving Quality of Chinese-English Patent Machine Translation Service



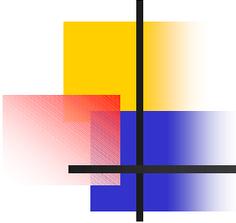
第3回WS: ユーザによる講演

- Arti Shah (USPTO) : Translation of Patent Documents at the United States Patent and Trademark Office
- Young Pyo Kim: KIPO's MT Activities and IP5 Mutual MT Project
- 番井進 (特許庁) : Current Status of MT Application in JPO



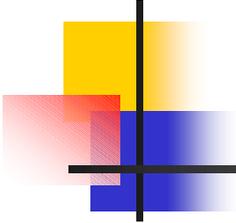
第3回WS: パネル討論

- Moderator: 潮田明 (富士通)
- パネリスト
Georg Artelsmair (EPO), Tao Wang (SIPO),
番井進 (特許庁), Arti Shah (USPTO), Philipp
Koehn (Univ. Edinburgh)
- 内容: 現状のサービス、機械翻訳の利点と将来
計画



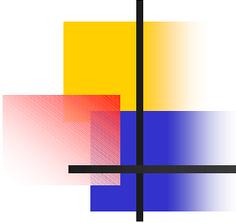
第3回WS:一般講演

- 5件の発表
- 日本: 3件: NICT、山梨英和大、山形大
- 香港: 1件: City Univ. Hong Kong
- デンマーク: 1件: LanA Consulting ApS



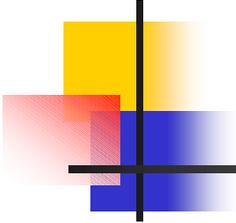
Workshopの意義

- 各国特許庁の最新情報
- 各国特許庁間の情報交換
- ユーザとしての特許庁と研究者との交流
- 特許翻訳に対する研究の現状



機械翻訳への政府の関与

- 本会議：元来は政府の意見交換の場
- カナダの会議では本来の立場が復活
- アメリカなどに比べて日本の立ち遅れ



まとめ

- 機械翻訳の現状と課題
- AAMT/Japio特許翻訳研究会の活動
- 特許翻訳ワークショップ

参考サイト

- ・Machine Translation Archive
- ・Japio Year Book